Computational Methods in Social Sciences

Vol. III, Issue 1/2015

http://cmss.univnt.ro/

contact: cmss@univnt.ro

ISSN 2344 – 1232
ISSN–L 2344 – 1232

“Nicolae Titulescu” University Publishing House

Phone: 004.021-330.90.32, Fax: 004.021-330.86.06

Calea Văcăreşti, nr. 185, Sector 4, Bucureşti, România

E-mail: editura@univnt.ro

Computational Methods in Social Sciences

The problem of allowing correlated errors in structural equation modeling: concerns and considerations

Richard HERMIDA[*]
George Mason University, 3575 Owasso Street, Shoreview, MN, USA, 55126

Abstract

Results of structural equation models may be negated by inappropriate methodological procedures. Model fit is known to be improved by the addition of pathways. Some pathways are added due to modification indices. These a-theoretical pathways will improve model fit at the expense of theory and reduction in parameter value replication. Furthermore, some additions to the model like correlating measurement errors are usually theoretically unjustifiable. This quantitative review examines the frequency of correlating measurement errors and examines the reasons, if any, for having these pathways in the model. Additionally, this quantitative review examines the consequences of correlating measurement errors in structural equation modeling.

Keywords: *Structural Equation Modeling, Confirmatory Factor Analysis, Measurement, Research Methods, Statistics.*

Introduction

The use of structural equation modeling (SEM) to test theoretical models has increased dramatically over the past 25 years in fields such as psychology, sociology, economics, marketing, and even behavior genetics (Markon & Kruger, 2004). While advances in statistical software packages have made it easier than ever for researchers to employ structural equation models (MacCallum & Austin, 2000), dissemination of best practices in SEM has arguably not kept pace. For example, there seems to confusion regarding the practice of allowing measurement errors to correlate in order to improve model fit. Many authors have cautioned against this practice (see Brannick, 1995; Cliff, 1983; Cortina, 2002; Gerbing & Anderson, 1984; Kaplan, 1989; Kaplan, 1990; MacCallum, 1986; MacCallum, Roznowski, & Waller, 1992; Shah & Goldstein, 2006; Steiger, 1990; Tomarken & Waller, 2003) for a variety of methodological reasons.

Other authors have attempted to identify the situations in which it is appropriate. Landis, Edwards, and Cortina (2009) argue that estimation of measurement errors in SEM is only appropriate when correlations amongst measurement errors are unavoidable. Such situations include when multiple measures of the same construct are used in longitudinal research, or when indicator variables share components. However, when measurement errors are allowed to correlate based on post hoc specification searches, a theoretical justification for engaging in this practice is lacking and the corresponding research moves from a confirmatory analysis to an exploratory analysis. Indeed, researchers have argued that such practices may only increase model fit indices because of capitalization on chance. Freeing the paths between measurement errors is particularly problematic in post hoc specification searches given that the correlations between errors may indicate that the model is misspecified; that correlating measurement errors based on post hoc modification may actually mask the underlying structure of the data; and that there is no theoretically defensible reason for allowing measurement errors to correlate based on post hoc modifications (Anderson & Gerbing, 1984; Landis et al., 2009).

Given the lack of consistent guidelines, researchers continue to engage in this practice without justification. Consequently, it is important to review the state of this practice as there is a high potential for misguided conclusions from structural equation models which could possibly have acceptable fit statistics due to estimation of correlations among measurement errors. Additionally, formal guidelines can only be established once it is understood how extensive this problem is and the reasons for which researchers allow measurement errors to correlate. As a result, the purpose of this article is to conduct a methodological review of the current practices of correlating SEM measurement errors at both the measurement and structural levels. Specifically, this review will examine the extent to which this practice is conducted within psychology and other related disciplines. It will also determine if this practice differs by journal impact rating; will explore the degree to which researchers

[*] Corresponding author: rhermida3@gmail.com

provide citations or explanations for allowing measurement errors to correlate; will detail what explanations have been provided; and will assess whether explanations have been provided a priori or post hoc in manuscripts. Finally, this review will uncover the amount, on average, that model fit is increased by allowing measurement errors to correlate as well as determine if this practice is influenced by model complexity.

Overview

This paper is organized as follows: First, a description of fit indices and specification searches in SEM is provided. Second, a presentation of the problems associated with allowing indicator residuals to correlate when testing both measurement and structural models is provided. Third, a brief review of past studies assessing the practice of allowing measurements errors to correlate is provided. Last, Iwill describe two factors that Ihypothesize to be related to the practice of error correlation: model complexity and journal impact. Iconclude the introduction with a summary of the unique contributions of this study. Current practices of allowing measurement errors to correlate are subsequently examined and meta-analyzed. Results of the examination are then discussed, along with the implications the current studies' findings have on the interpretation of results put forth from SEM researchers.

SEM and Model Fit Indices

In SEM, global model fit statistics are computed based on the pathways not estimated in the hypothesized model (Jöreskog, 1969). The addition of model paths therefore, will decrease the discrepancy between the observed and reproduced variance-covariance matrices. Commonly reported model fit indices include the chi-square statistic, root mean square error of approximation (RMSEA), comparative fit index (CFI), and the normed fit index (NFI), although many other global fit indices are available in SEM data analysis programs. Additionally, each individual path in the model can be examined for magnitude and statistical significance.

When evaluating the measurement model portion of a theoretical model in SEM, two distinct parameter estimates are generated for each indicator variable (Landis et al, 2009). The first, true score common variance, is the shared variability an indicator has with other indicators of the same latent variable (Maruyama, 1998).Thus, the term true score denotes that this estimate represents the underlying latent variable that the indicator variables were hypothesized to measure. The term common variance implies that the variance derived is shared by all indicators of the same latent trait. The factor loading of a given indicator on an underlying latent variable should be large and statistically significant to the extent that the indicator has a large degree of true score common variance.

The second estimate is a residual term into which all other unique sources of variance go. Residual terms can be broken down into true score unique variance, defined as the systematic variance of an indicator that is uncorrelated with the variance of other indicators, and error variance, or the unsystematic variability of an indicator (Maruyama, 1998). When estimating the parameters of measurement models, it is typical for the covariances among measurement errors to be fixed at zero given that residual variances are defined by their uniqueness to each indicator (Landis et al, 2009).

Specification Searches

A common procedure in SEM when a model has inadequate fit is to perform a specification search. Specification searchers are post hoc explorations which provide users with information on how to modify models in order to improve fit values (MacCallum, 1986; MacCallum et al., 1992; Sörbom, 1989). These modifications are based on statistical criteria and once conducted, shifts the research focus from confirmatory analysis to exploratory and data-driven analysis. Modification indices often show that model fit would improve if one or more residuals among indicator variables were allowed to correlate. At least some researchers will at this point, correlate the residuals among indicator variables. This practice is problematic for a variety of reasons.

Problems with Allowing Correlated Errors

The first problem with allowing measurement errors to correlate in structural equation models based on post hoc modifications is that it allows researchers to achieve good fit statistics in spite of omitting relevant variables from their models (Cortina, 2002). As explained by Landis et al. (2009), "to the degree that two residuals correlate, there is evidence that there exists a cause of both of the variables to which the residuals are attached but that is not specified in the model" (p. 17). When indicator variables are systematically influenced by the same extraneous variable in addition to the specified latent variables they represent, the influence of the extraneous variable may be estimated through measurement error correlations without a specification of what the influence is (Fornell, 1983). As a result of the estimation of such correlations, the fit of the model improves, but our understanding of the phenomenon in question does not.

The second issue with allowing measurement errors to correlate in a post hoc fashion is that significant correlations are likely to be due to sampling error. Given that the number of off-diagonal elements of error covariance matrices can be very large, the probability of one or more such covariances being large simply because of sampling error is substantial. Several studies have shown that modified models often capitalize on the idiosyncratic features of sample data, and are likely to produce a departure from the true population model (Chou & Bentler, 1990; Green, Thompson, & Babyak, 1998; Green, Thompson, & Poirer, 1999; Lance, Conway, & Mulaik, 1988; MacCallum, 1986; MacCallum et al., 1992). To that end, Grant (1996) found that changing a hypothesized model to allow measurement errors to correlate based on specification search recommendations improved model fit in an initial sample, but failed to hold in cross-validation samples.

A third problem with allowing measurement errors to correlate is that this practice may bias parameter estimates of both the measurement and structural model (Tomarken & Waller, 2003). For example, Gerbing and Anderson (1984) argue that even when correlated measurement errors do not significantly alter parameter estimates of a measurement or structural model, they can still mask the underlying structure of modeled relationships.

Previous Reviews

I will now turn our attention to previous reviews that have addressed the issue of error correlation in SEM. Three reviews have attempted to provide estimates of the extent to which published studies using SEM permit the errors among measurement errors to correlate. Unfortunately, these reviews report somewhat contradictory findings and sampled only a limited number of journals and studies.

The first review was conducted by Shah and Goldstein (2006), with a focus on this practice in management and operations journals. The authors examined the use of allowing correlated error and reviewed studies from *Management Science*, *Journal of Operations Management*, *Decision Science*, and *Journal of Production and Operations Management Society*, looking at all studies conducted from 1984 until 2003. While the practice of correlated errors was not the main focus of the study, it was included as an area of interest for the main topic, which was structural equation modeling practices. This review estimates that around 29% of articles testing CFA models freed the parameters among measurement error terms, while only 11% of articles testing strictly structural models engaged in this practice. Additionally, Shah and Goldstein state that only fifty percent of structural equation models that they reviewed provided justification for allowing measurement errors to correlate.

A second review was carried out by Cole, Ciesla, and Steiger (2007). The authors examined studies found in *Psychology Assessment*, *Journal of Counseling Psychology*, *Journal of Applied Psychology*, *Health Psychology* and *Journal of Personality and Social Psychology* in the year 2005. Across 75 studies, 21% of the articles explicitly allowed for error correlation, while an additional 5% almost certainly correlated errors, which was inferred from the calculation of degrees of freedom, while 5% of the articles were too vague to ascertain the presence of correlated errors. Therefore, according to the authors, anywhere from approximately 27% to 32% of published studies allowed measurement errors to correlate, an estimate quite different than that given by Shah and Goldstein (2006).

In addition to providing an estimate of allowing measurement errors to correlate when using SEM, Cole et al. ascertained what justifications were provided in the studies they reviewed for allowing errors to correlate. According to the authors, 71% of the justifications were driven by theory. That is, residuals were allowed to correlate when the measures were administered to the same informant. Twenty-nine percent of the justifications were driven empirically. That is, when residual correlations were allowed in order to generate a model that provided better fit to the data.

The authors also found that very few of the articles that allowed errors to correlate cross-validated the revised model with new data. This is undesirable, given that the practice of correlating errors can be seen as a capitalization on chance. Additionally, the focus of this study was an examination of the affects of not including design-driven correlated residuals in latent-variable covariance structure analysis. The stated conclusion derived by the authors was that failure to include certain correlated residuals can change the meaning of the extracted latent variables and generate potentially misleading results. Consequently, Cole et al. (2007) argue that after allowing measurement errors to correlate, authors should revise the description of the latent variables under question.

A final examination of the practice was examined by Landis et al., (2009). The authors examined 58 empirical articles derived from the *Journal of Applied Psychology*, *Journal of Management*, and *Personnel Psychology* that used structural equation modeling. According to the authors, 9 to 12% of published studies in these journals from 2002 to 2007 allowed measurement errors to correlate and approximately 70% of the researchers who engaged in this practice did so as a post hoc modification. Because the authors did not count articles as having correlated errors unless they were explicitly stated, it is likely that their estimates are an underestimation of the practice.

One reason for such contradictory findings is that these prior reviews is that studies covered came from a limited number of journals over brief spans of time. For instance, Landis et al. (2009) narrowed their review to studies published in *Personnel Psychology*, *Journal of Applied Psychology*, and *Journal of Management* from 2002 to 2007. Cole (2007) examined studies found in *Psychology Assessment*, *Journal of Counseling Psychology*, *Journal of Applied Psychology*, *Health Psychology* and *Journal of Personality and Social Psychology* from 2005. Finally, Shah and Goldstein (2006) reviewed studies from *Management Science*, *Journal of Operations Management*, *Decision Science, and Journal of Production and Operations Management Society*, although they did look at all studies conducted from 1984 until 2003. Clearly further investigation is needed in order to clarify the discrepancies found in these previous reviews by examining a broader range of journals over a larger span of time. Inow turn the discussion to factors that may influence the practice of error correlation.

Model Complexity

Model complexity can be defined as the ability of a model to fit a diverse array of data patterns well by some established criterion of fit (Dunn, 2000; Myung, 2000; Pitt, Myung, & Zhang, 2002; Preacher, 2006). Previous research has suggested model complexity can be thought of as the average fit of a model to regions of data space, or the space containing all empirically obtainable data patterns relevant to a particular modeling domain (Preacher, 2006). Model complexity can be seen as the antithesis of parsimony. That is, all other things equal, as model complexity increases, parsimony is likely to decrease.

There are a number of factors that contribute to model complexity. Perhaps the most obvious is the effective number of free parameters in the model, which is defined as the number of freed parameters minus the number of functional constraints placed on otherwise free elements of the model. This difference is defined in SEM terminology as q. In structural equation modeling, all other things being equal, models with a higher q will be better able to fit data (Forster & Sober, 1994; Jeffreys, 1957; Wrinch & Jeffreys, 1921). This is because freeing model parameters reduces the number of dimensions in which observed data can differ from the data that would be implied by the hypothesized model (Mulaik, 2001; Mulaik, 2004). Q is inversely related (all else being equal) to degrees of freedom. That is, given a certain model, degrees of freedom represent the number of dimensions in which observed data can differ from the data that would be implied by the hypothesized model.

Another factor that contributes to model complexity is sample size. Sample size, like degrees of freedom is inherently related to the number of dimensions in which observed data can differ from the data that would be implied by the hypothesized model. Sample size is in turn connected with Chi-Square such that for virtually all models that are not perfectly clean fitting, Chi-Square increases as a function of sample size. Because virtually all of the major fit indices used today are derivatives of the Chi-Square in some way, sample size is inherently connected to fit index values in nearly all cases. It is therefore expected that the practice of error correlation will be related to model complexity via degrees of freedom and the sample size of the model in question.

One possible explanation for the allowance of correlated errors is that because extensive time, money, and effort goes into data collection and interpretation, researchers ultimately do not want to disregard data with poor fit and will instead attempt to save the data by improving fit through correlated errors (Hermida et al., 2010; Landis, Edwards, & Cortina, 2009). It has been conjectured that rather than abandon poor fitting data straight away, it might make more sense to modify the model so as to fit the data better (Sörbom, 1989). If correlated errors occur because of a researcher's desire to see good model fit for their data, then it stands to reason if a model already has acceptable fit, the likelihood of correlated errors is lessened. With respect to falsifiability, if power is associated with fit indices in covariance models such that more falsifiable studies are more likely to be rejected, and if researchers are motivated to correlate errors because of poor model fit, it stands to reason that falsifiability will be associated with the practice of error correlation such that the practice of error correlation will be positively related to the falsifiability of exact and close fit tests.

Journal Impact and Influence

Another contribution that this study will offer is examination of how the practice of error correlation is influenced by the particular impact and influence of the journal in which the study was reported. Although the practice of correlated errors is widespread in psychology, many studies and reports have indeed exposed the problems associated with the practice (Brannick, 1995; Cliff, 1983; Gerbing & Anderson, 1984; Kaplan, 1989; Kaplan, 1990; Landis et al., 2009; MacCallum, 1986; MacCallum, Roznowski, & Waller, 1992; Shah & Goldstein, 2006; Steiger, 1990; Tomarken & Waller, 2003, Cortina, 2002). If the practice of correlated errors is indeed an undesirable practice, and journals of higher impact and influence are indeed of higher quality, it follows that the practice of error correlation should be less prevalent in higher quality journals. Therefore, it was hypothesized that the practice of error correlation will be related to journal impact and influence such that error correlation prevalence will be negatively correlated with journal impact and influence.

Unique Contributions and Aims

While there have been attempts to capture the practice of correlating errors in structural equations modeling, this quantitative review seeks to provide additional contributions over and above those of the three previous studies. First, the three previous studies differ in their estimations of prevalence for the practice of allowing correlated errors in SEM. This is most likely because each study used a small sample of journals across a narrow range of disciplines, and on most occasions, analyzed studies over different time periods. The current study seeks to remedy this problem by examining a much wider array of journals over a consistent time period. More specifically, this review will examine a wide range of journals over a ten year time period from 1997 to 2007. By casting a wider net of journals and disciplines over a longer and consistent time period, this quantitative review will provide a more unifying and comprehensive examination of the degree to which correlated errors are practiced. A question of interest for this study is the degree and form in which the practice occurs in specific subdisciplines in psychology and disciplines other than psychology, such as management and educational research. Examination of this question allows us to determine if some fields of study allow researchers to engage in this practice to a greater extent than others, and thus discuss the implications this has on conclusions garnered from SEM research in specific sub domains.

A second major contribution of this study is a fuller and more in-depth examination of the justifications researchers provide for allowing measurement errors to correlate. Certainly, improved model fit is a common justification, but it is possible that other explanations exist. For example, it is not uncommon for studies to allow for correlated errors in initial model testing when the research design is longitudinal and the errors that are allowed to covary are the same indicators at different time periods. It is also possible for researchers to hypothesize a priori that errors will be correlated, based on the nature of study variables. For example, many studies allow errors to correlate when the variables have shared components. Kenney and Judd (1984) suggested using all possible cross-products of latent variable indicators as indicators of a latent product to be used for testing multiplicative structural equation models. Some of these cross-products will share components, so it is almost certain that their errors will correlate. Ultimately, these two reasons for allowing errors to correlate are part of the design, and are not necessarily related to sampling error or omitted variables issues. Thus, this study will help determine if the majority of studies which allow measurement errors to correlate are doing so for theoretically justifiable reasons, such as longitudinal research, or for unjustifiable reasons, such as improvement of model fit.

In this study I also attempted to examine the less appropriate justifications for allowing errors to correlate, specifically when they are made ex-post facto. I determined the rate of these justifications and if any articles in particular are being citied as support for this being a valid practice. An attempt will be made to match what was ascertained and stated by the researchers in the original article and what was ascertained and stated by the researches in the correlated errors study to examine if there are any disconnects between the two articles. It is my expectation that for some of the correlated errors studies, where the correlation occurred ex-post facto, there will be some miscommunication between cited articles and application of the ideas in those articles. By uncovering a source of miscommunication, I can begin to unlock the antecedents to this practice and will be in a better position to offer recommendations for practice.

Finally, no study to date has examined the idea of model complexity being related to the practice of correlating errors. One of the most parsimonious explanations for why researchers allow correlate errors in structural equation modeling is that they do not wish to abandon data that has poor fit; therefore they will correlate errors unjustly for the purpose of having a better fitting model. If this is true, it stands to reason that antecedents to poorer fitting data might be relevant to the practice of correlating errors. One such antecedent is model complexity. All other things being equal, a more complex model will generate better overall model fit. Therefore, it seems more complex models will be less likely to engender the practice of correlating errors.

Method: Literature Search

The goal of this study was to review the practice of allowing correlated measurement errors in a variety of disciplines. Thus, studies from various fields of study including Psychology, Education, Business, and Sociology were included. Consequently, PSYCINFO, ProQuest, ERIC, AB-INFORM, were used to collect empirical articles. Article searches were limited to years 1997 through 2007 in order to represent modern methodological practices. In all databases, keywords covariance models, structural equation modeling, and SEM were utilized in order to identify all articles that used structural equation modeling.

Method: Summary of Meta-Analytic Dataset

From searches of the aforementioned databases, 315 useable articles that that allowed measurement errors among indicator variables to correlate were identified. I excluded literature reviews, methodological papers, and papers that created models with simulated data. The 315 articles were coded by two different coding pairs, with each all coders possessing advanced degrees related to Psychology. Table 1 illustrates interrater agreement statistics.

Table 1. Interrater Agreement Statistics

Meta-Analytic Category	Type of Agreement	Agreement Value
Sample Size	Intraclass Correlation Coefficient	0.98
Degrees of Freedom	Intraclass Correlation Coefficient	0.82
Journal Quality	Intraclass Correlation Coefficient	0.99
Fit Index Values	Intraclass Correlation Coefficient	0.98
Correlated Errors (yes/no)	Cohen's Kappa	0.97
Model Type (measurement/structural)	Cohen's Kappa	0.95
Rationale for Correlating Errors	Cohen's Kappa	0.88

Method: Inclusion Criteria

In order to begin coding articles, it first had to be determined which studies allowed measurement errors to correlate. The articles were searched for any combination (or variant) of the following terms: correlated, freed, errors, and residuals. The models themselves were also visually inspected for measurement error pathways. It is possible and probable that some published manuscripts neither mentioned nor depicted the measurement error pathway even if the author did free the path. Unfortunately, in the absence of those explicit text or diagram indicators there is no feasible method for including those manuscripts that are consistent across all articles.

Method: Treatment of Multiple Reported Models

When multiple tested models were reported in a single manuscript additional decisions were made in coding. The inclusion of correlated measurement error pathways is usually done based upon the modification indices from the results of a tested theoretical model. The model with the included correlated measurement error pathways is referred to here as the modified model. When the coders were able to determine which reported model contained correlated measurement errors the model tested just previous was included as the theoretical model. The two models were included in the coding as the theoretical model and the modified model.

Method: Coding Procedures

Once all studies allowing the measurement error pathways were identified, relevant moderators were also coded. Moderator selection was based upon there potential influence on model statistics, model complexity, authors' decisions, and reporting standards. Those four classifications encompass both the impact of correlated measurement errors on model parameters and the conditions under which researchers' practice SEM. Model statistics serve as the measure of impact from the correlated measurement error pathway inclusion. Model complexity is a potential reason for the correlation of measurement errors, and also has the potential to directly affect model statistics. Each of the classifications is further detailed below.

Method: Model Statistics

When theoretical and modified models were reported, both sets of fit statistics, degrees of freedom, and statistical significance were coded. Model fit statistics were differentially reported but every fit statistic value reported was coded. If the pathway that correlated errors was reported, that value was also recorded in terms of the correlational relationship. When more than one such pathway was present an average correlation was recorded. In addition, the model degrees of freedom and statistical significance were coded. The three codes for model significance included if both the theoretical and modified models were non-significant, if statistical significance changed to non-significance, and if the model remained significant after allowing correlated measurement errors.

Method: Model Complexity

Model complexity moderators include the model type, the number of manifest items, and the number of latent constructs. The presence or absence of item parcels and the range of items per parcel were coded to ensure accurate reflection of model complexity. The models where coded as either a measurement model or structural model.

Method: Author Decisions

The researchers' motives and justifications for allowing measurement errors to correlate were coded. A distinction was made between a priori and longitudinal justifications. Models that coded the measurement errors post-hoc were categorized into 3 groups: as related to model content if stated by the authors; methodological if the authors stated that parameters between measurement errors were freed because of the same measure(s) being used over time or if they were specified due to estimating a model with a moderator and allowing the moderator indicator residuals to correlate with the component indicators residuals (but not if the component indicators were allowed to correlate); and data driven if the researcher specified that measurement errors were allowed to correlate because they improved model fit. Any citations given as a justification were also coded. These cited references were subsequently reviewed by the coders and a judgment was made if the cited reference matches the reported justification.

Method: Reporting Standards

Authors present and report model parameters in conformity to their research field. As the field and journal characteristics are potential influences on the inclusion or exclusion of measurement error pathways, they were coded. It was hypothesized that higher impact journals would be stricter in allowing the practice of correlating measurement errors. The journal impact and value are based upon the year the article was published (Bergstrom, 2007).

Results: Point Estimation

One aim of this research was to establish a point estimate for the percentage of studies that correlated errors in the psychological literature. To examine this issue, a random selection of 985 studies that used some form of structural equation modeling was obtained. Next, I calculated the number of studies that correlated errors. The results indicated that within these studies, 315 studies correlated errors out of a possible 985 studies. The percentage of studies that correlate errors in the psychological literature is thus best estimated at 32%. The average correlation between correlated errors was 0.31.

Results: Rationales Given

Of critical interest to this study was the rationale researchers gave for error correlation. Rationales were classified under five categories: a-priori, longitudinal designs, post-hoc construct theory, post-hoc method theory, and modification indices. Listing of the percentages by rationale appears in Table 2. The most common rationale was correlating errors as a result of modification indices specified after running the initial model. It should be noted that only 25% of the models were correlated for a-priori reasons, or because the data in question was longitudinal. Additionally of interest were the particular citations given by researchers. Unfortunately, most researchers did not cite a specific article in support of error correlation. Specifically, only 15% of studies gave a direct citation in support of the choice to correlate errors. Out of the 15% of articles that gave a specific citation, the vast majority of those citations were related to post-hoc construct theories, with the rationale equating to the correlation of measurement errors based on the fact that a different study had found the variables to be significantly associated with one another.

Table 2. Rationale for Error Correlation

Rationale	Percentage of Sample
Modification Indices	37%
Post-Hoc Construct Theory	24%
Longitudinal Data	18%
Post-Hoc Method Theory	14%
A-Priori Theory	7%

Results: Model Complexity

Another aim of this research is to examine the model complexity of covariance models that correlated errors vs. the model complexity of covariance models that did not correlate errors. Indeed, there was a statistically significant difference in the degrees of freedom of models in studies that correlated errors (M = 73.25, SD = 8.23), and those that did not (M = 69.02, SD = 6.34), $t(920) = 12.36$, $p < .05$, d =.58. Similarly, there was a statistically significant difference in the sample size of models in studies that correlated errors (M = 314.18, SD = 64.23), and those that did not (M = 279.55, SD = 74.21), $t(920) = 10.71$, $p < .05$, d =.50. As expected, models that correlated errors were significantly more complex, and were significantly more falsifiable than models that did not correlate errors.

Results: Journal Impact

Another aim of this research was to establish the relationship between journal impact and the practice of error correlation. To examine this issue, the journal impact for all studies was collected. Next, each study was assigned a binary code to assess if they correlated errors (0 = no 1 = yes). Logistic regression analysis was conducted with the dichotomous variable of error correlation as the dependent variable, and journal impact as the independent variable. The results indicated that indeed, journal impact was associated with the practice of error correlation such that as journal impact increased, the probability of error correlation decreased, (odds ratio [OR] = 0.61, 95 % confidence interval [CI] = 0.55-0.68, p<.01).

Results: Fit Index Improvement

Another important aim of this research is to establish the degree to which fit indices are impacted by the practice of error correlation. To examine this issue, two sets of fit indices were evaluated: those calculated by the researcher before error correlation and those calculated by the researcher after error correlation. I examined this issue across studies where error correlation was conducted for invalid reasons (i.e. not a-priori or longitudinal). Table 3 displays the change in fit across the examined fit statistics. In general, the change in fit as a result of error correlation resulted in a .02-.03 betterment in fit across indices. This difference amounts to approximately half of the difference between qualitatively different assessments of fit, such as a value of.08 on the RMSEA fit statistic corresponding to “moderate” model fit, and a value of.05 (difference of.03), or such as a value of.90 on the CFI fit statistic corresponding to “reasonable” fit, and a value of.95 (difference of.05) corresponding to “close” fit (Hu & Bentler, 1999).

Because of the possibility of researchers correlating models to move through the unofficial rules of thumb with respect to fit, the percentage of studies that correlated errors to pass through “good” and “excellent” thresholds of model fit was examined. Specifically, I examined the percentage of models that did not meet a fit index value associated with the “rules of thumb”, and the percentage of models that passed through that threshold after the correlated or errors. I examined the RMSEA and CFI fit indices, with the rules of thumb equating to RMSEA values of .05 and .08 for good and moderate fit, respectively (Brown & Cudeck, 1992), and CFI values of .90 and .95 for reasonable and close fit, respectively (Hu & Bentler, 1999).

Out of the all the models that a) presented both pre and post correlation RMSEA fit values, and b) failed to pass the .08 threshold for RMSEA prior to the correlation of errors, 79 percent of models passed the .08 threshold for model fit *after* the correlation of errors. Similarly, out of the all the models that a) presented both pre and post correlation RMSEA fit values and b) failed to pass the .05 threshold for RMSEA prior to the correlation of errors, 34 percent of models passed the.05 threshold for model fit *after* the correlation of errors.

Out of the all the models that a) presented both pre and post correlation CFI fit values, and b) failed to pass the .90 threshold for CFI prior to the correlation of errors, 78 percent of models passed the .90 threshold for model fit *after* the correlation of errors. Similarly, out of the all the models that a) presented both pre and post correlation CFI fit values and b) failed to pass the .95 threshold for RMSEA prior to the correlation of errors, 48 percent of models passed the .05 threshold for model fit *after* the correlation of errors.

Table 3. Fit Statistic Change as a Result of Error Correlation

Model	DF	χ^2	RMSEA	CFI	GFI	NNFI
Before Error Correlation	75	395.03	.09	.92	.92	.90
After Error Correlation	73	328.73	.07	.95	.94	.93
Difference	2	66.30	.02	.03	.02	.03

Discussion

With few exceptions, there is no theoretical defensible reason for the practice of error correlation. As found in the current study, nearly one-third of all studies that use structural equation modeling engage in the practice of measurement error correlation, and of those that do, most researchers engage in this practice for invalid reasons. Consequently, the current quantitative review has shown the severity of this problem across various fields of study, and highlights the need for SEM users, journal reviewers, and researchers to be wary of misspecified models found in the literature as well as the need for more stringent guidelines for reviewing and accepting SEM manuscripts. If the errors that are correlated are random in nature, then correlating errors is taking advantage of random chance, thereby moving the researcher from confirmatory model testing to exploratory model testing. Alternatively, if the errors that are correlated are the result of omitted variables, it is imperative to identify the missing variables, collect data from a second sample, and test the hypotheses that the omitted variable accounted for the correlated error (Cortina, 2002). For both of those reasons, if error correlation is seen in a poor-fitting model's modification indices and is not a result of the study's design, then the best possible course of action is most likely to form a hypothesis about the reason for the errors being correlated, such as hypothesizing what variable was omitted in the model, and then test the new model with new and independent data (Hayduk & Glaser, 2000).

Summary of Findings

Researchers most frequently engaged in the practice of allowing measurement errors to correlate because of the increases to be gained in model fit shown by the modification indices. Apart from modification indices, researchers most frequently correlated errors because the two variables in question were cited as being associated with one another in a different study. This justification for allowing measurement errors to correlate post hoc is invalid and atheoretical for many reasons.

First, if a researcher can provide a strong justification for allowing correlated errors as a result of a known omitted variable, it is reasonable to wonder why the parameter was not represented in the original model (MacCallum et al., 1992). Second, performing such modifications is tantamount to rewarding poor scale construction and/or model development. Third, there remains a potential disconnect between the omitted variable claimed to have caused the error correlation in the model, and the omitted variable that actually caused the error correlation in nature. In psychology, it has been said that "everything correlates with everything else", given rise to what David Lykken (1968) called "the crud factor" (Meehl, 1997). If indeed variables have associations with numerous other variables, it seems unlikely that the researcher in question could accurately ascertain the exact nature of what caused the error correlation in that particular model without conducting a follow up study and cross-validating the model.

The current study also demonstrated how error correlation improves fit in a non-trivial way. The improvement in model fit was on average approximately half of the distance between models that are judged to be adequate according to the rules of thumb, and models that are judged to be good. Related to that point, approximately three-quarters of models that prior to error correlation did not reach the minimum threshold for fit, according to the rules of thumb, reached minimum fit after error correlation. This is problematic, and suggests that researchers correlate errors to pass through the unofficial rules of thumb SEM practitioners have set regarding model fit and model testing in general. This could be because models that do not pass through these rules of thumb are less likely to be published after going through the review process and researchers are thus rewarded for using modification indices to obtain value on the sunk costs they put into data collection. Therefore, it seems that researchers are incentivized to correlate errors, when faced with a model that is within range of passing the thresholds associated with the rules of thumb.

Recommendations and Future Research

Now that the extent of the problem of allowing measurement errors to correlate in structural equation models and the reasons for this problem are known, some potential remedies can be offered.

One remedy that is offered is for reviewers and referees to abandon qualitative rules of thumb descriptions to model fit based on unofficial rules of thumb. I make this recommendation for several reasons. First, attaching qualitative labels to quantitative factors (such as model fit) causes loss of information in much the same way as dichotomization of continuous variables will cause a loss of information. To give an extreme example, RMSEA values of.079 and.081 fit an entire covariance model to virtually the same degree (all else being equal), but would have different qualitative labels of model fit, if the rules of thumb are followed. It is also known that optimal cutoff criteria are heavily dependent on specific aspects of the model apart from fit (Marsh et. al, 2004; Nye & Drasgow, 2011). Specifically, simulation research has shown that the optimal cutoff criteria are actually dependent on numerous features of the model, including the estimation method used, the sample size used, the number of free parameters, and the degree to which assumptions of multivariate normality are met or not met (Hu & Bentler, 1999, Marsh et. al, 2004; Tomarken & Waller, 2005). Because of this, traditional fit indices, models evaluation of model fit should be considered with respect to the model's specific characteristics. Cohen (1988) cautioned against strict interpretations and uses of rules of thumb with respect to effect size, arguing for a more nuanced and thoughtful approach. This recommendation is also applicable to interpretation of fit indices in SEM and would hopefully eliminate a possible incentive researchers have for correlating errors inappropriately, without harming proper interpretation of fit indices.

An additional consideration regarding fit indices relates to a second issue: the use of approximate model fit indices to evaluate model fit. Some researchers have argued that Chi-Square is the only acceptable fit index to use in evaluating structural equation models (Barrett, 2007; Hayduk, Cummings, Boadu, Pazderka-Robinson, & Boulianne, 2007) The argument for the sole use of the Chi-Square in evaluating models is centered on the following points: 1) there are no single thresholds for GOF indices that can be applied under all possible measurement and data conditions, 2) GOF indices allows for researchers to avoid careful model specification and examination, 3) GOF indices can allow mediocre models to make it through the peer-review process, 4) the potential for the seriousness of casual misspecification to be uncorrelated with GOF indices values, and 5) Chi-Square does a better job at detecting model misspecification than does any other fit index.

Readers who are interested in examining this issue at length can examine the above issues and counterarguments in the special issue of the *Personality and Individual Differences* journal that summarizes this debate. I would simply point out that to the degree approximate model fit indices are inappropriate; is the degree that use of GOF indices provides little *benefit* for scientific advancement, considering the *cost* of prompting researchers into inappropriate error correlation.

The second recommendation is for researchers to engage more seriously in cross-validation and replication of models. If a researcher believes they understand the cause for modification indices turning up significant pathways between measurement errors, the researcher should collect data from a new sample, with all variables included, thus confirming the model uncovered by exploratory analysis.

The third and final recommendation is for psychology to engage in more serious enforcement of quantitative sections of studies that use SEM and also to engage in more serious quantitative education with respect to SEM. A main theme from this quantitative review is that inappropriate statistical practices most likely stem from either misunderstanding of statistical issues, or through the mechanism of individual self-interest (or both). If reviewers and editors insist on researchers not improperly correlating errors, then the self-interest component of the researcher will be oriented towards not correlating errors in inappropriate situations. Additionally, if more serious quantitative education with respect to SEM is undertaken, it seems likely that researchers who are concerned with quantitative methods will not correlate errors improperly.

Limitations

This study contains some limitations. First, this quantitative review only contained articles that made it through the publication process. While this was intentional, as the focus was specifically on published works, it could be the case that the practice of inappropriate error correlation differs from published to non-published studies. For example, it could be the case that researchers who refuse to inappropriately correlate errors are underrepresented in published articles, thus altering the percentage of studies that correlate errors. Therefore, it is important to not generalize the findings in this quantitative review to non-published studies. Second, there are cases where refusing to correlate measurement errors might make solid model fit an especially conservative undertaking. For example, residual correlation has been recommended in multiple mediation models because covariances among the mediators are unlikely to be completely explained by their mutual dependence on the independent variable (Preacher & Hayes, 2008). While I take the position that residual and measurement error correlation is to be avoided in this type of case, it should also be recognized that by doing so, model fit will possibly be slanted in a conservative manner to some degree, and there remains a possibility for model misspecification as a result of not permitting residual covariances to vary, especially if the those covariances are significant in magnitude.

Conclusions

This study provides strong evidence for the importance of understanding the practice of inappropriate error correlation in structural equation modeling. Methodological issues have been emphasized in structural equation modeling in scientific research and education in quantitative methods. An additional issue to emphasize in these venues should be the abolishment of inappropriate error correlation practices.

References

[1] Markon, K.E., & Kruger, R.F. "*An empirical comparison of information-theoretic selection criteria for multivariate behavior genetic models*", Behavior Genetics, 34. 593-610. 2004.

[2] MacCallum, R.C., & Austin, J.T. "*Applications of structural equation modeling in psychological research*", Annual Review of Psychology, 51.201-226.2000.

[3] Brannick, M.T. "*Critical comments on applying covariance structure modeling*", Journal of Organizational Behavior, 16. 201-213.1995.

[4] Cliff, N. "*Some cautions concerning the application of causal modeling methods*", Multivariate Behavioral Research, 18.115-126.1983.

[5] Cortina, J. "*Big things have small beginnings: An assortment of 'minor' methodological misunderstandings*", Journal of Management, 28. 339-262. 2002.

[6] Gerbing, D.W., & Anderson, J.C. "*On the meaning of within-factor correlated measurement errors*", Journal of Consumer Research, 11. 572-580. 1984.

[7] Kaplan, D. "*Model modification in covariance structure analysis: Application of the expected parameter change statistic*", Multivariate Behavioral Research, 24. 41-57. 1989.

[8] Kaplan, D. "*Evaluating and modifying covariance structure models: A review and recommendation*", Multivariate Behavioral Research, 24. 137-155. 1990.

[9] MacCallum, R.C. "*Specification searches in covariance structure modeling*", Psychological Bulletin, 100. 107-120. 1986.

[10] MacCallum, R.C., Roznowski, M., & Necowitz, L.B. "*Model modifications in covariance structure analysis: The problem of capitalization on chance*", Psychological Bulletin, 111. 490-504. 1992.

[11] Shah, R., & Goldstein, S.M. "*Use of structural equation modeling in operations management research: Looking back and forward*", Journal of Operations Management, 24. 148-169. 2006.

[12] Steiger, J.H. "*Structural model evaluation and modification: An interval estimation approach*", Multivariate Behavioral Research, 25. 173-180. 1990.

[13] Tomarken, A.J., & Waller, N.G. "*Potential problems with "well fitting" models*", Journal of Abnormal Psychology, 112. 578-598. 2003.

[14] Landis, R., Edwards, B. D., & Cortina, J. "*Correlated residuals among items in the estimation of measurement models*". In C. E. Lance & R. J. Vandenberg (Eds.). Statistical and methodological myths and urban legends: Doctrine, verity, and fable in the organizational and social sciences (pp. 195-214). New York: Routledge. 2009.

[15] Anderson, J.C., & Gerbing, D.W. "*The effect of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis*", Psychometrika, 49. 155-173. 1984.

[16] Jöreskog, K.G. "*A general approach to confirmatory maximum likelihood factor analysis*", Psychometrika, 34. 183-202. 1969.

[17] Maruyama, G. "*Basics of Structural Equation Modeling*". Thousand Oaks. Sage. 1998.

[18] Sörbom, D. "*Model modification*", Psychometrika, 54. 371-384. 1989.

[19] Fornell, C. "*Issues in the application of covariance structure analysis: A comment*", Journal of Consumer Research, 9. 443-448. 1983.

[20] Chou, C. P., & Bentler, P. M. "*Model modification in covariance structure modeling: A comparison among the likelihood ratio, Lagrange Multiplier, and Wald tests*", Multivariate Behavioral Research, 25. 115–136. 1990.

[21] Green, S. B., Thompson, M. S., & Babyak, M. A. "*A Monte Carlo investigation of methods for controlling Type I errors with specification searches in structural equation modeling*", Multivariate Behavioral Research, 33. 365–383. 1998.

[22] Green, S. B., Thompson, M. S., & Poirer, J. "*Exploratory analyses to improve model fit: Errors due to misspecification and a strategy to reduce their occurrence*", Structural Equation Modeling, 6. 113–126. 1999.

[23] Lance, C.E., Cornwell, J.M., & Mulaik, S.A. "*Limited information parameter estimates for latent or mixed manifest and latent variable models*", Multivariate Behavioral Research, 23. 171-187. 1988.

[24] Grant, L. D. "*A comprehensive examination of the latent structure of job performance*", Ph.D. dissertation, North Carolina State University. 1996.

[25] Cole, D., Ciesla, J., & Steiger, J. "*The insidious effects of failing to include design-driven correlated residuals in latent-variable covariance structure analysis*", Psychological Methods, 12, 381-398. 2007.

[26] Dunn, J.C. "*Model Complexity: The fit to random data reconsidered*", Psychological Research, 63. 174-182. 2000.

[27] Myung, I. J. "*The importance of complexity in model selection*", Journal of Mathematical Psychology, 44. 190–204. 2000.

[28] Pitt, M.A., Myung, I. J., & Zhang, S. "*Toward a method of selecting among computational models of cognition*", Psychological Review, 109. 472-491. 2002.

[29] Preacher, K.J. "*Quantifying Parsimony in Structural Equation Modeling*", Multivariate Behavioral Research, 41. 227-259. 2006.

[30] Forster, M.R., & Sober, E. "*How to tell when simpler, more unified, or less ad hoc theories will provide more accurate predictions*", British Journal for the Philosophy of Science, 45. 1-35. 1994.

[31] Jeffreys, H. "*Scientific inference*", Cambridge University Press, Cambridge. 1957.

[32] Wrinch, D., & Jeffreys, H. "*On certain fundamental principles of scientific inquiry*", The London Edinburgh and Dublin Philosophical Magazine and Journal of Science, 42. 339-369. 1921.

[33] Mulaik, S.A. "*The Curve-fitting problem: An objectivist view*", Philosophy of Science, 68. pp 218-241. 2001.

[34] Mulaik, S.A. "*Objectivity in science and structural equation modeling*". In David Kaplan (Ed.), The Sage Handbook of Quantitative Methodology for the Social Sciences. Thousand Oaks. Sage. 2004.

[35] Hermida, R., Conjar, E.A., Najab, J.A., Kaplan, S.A., & Cortina, J.M. "*On the Practice of Allowing Correlated Residuals in Structural Equation Models*". Unpublished Manuscript, Department of Psychology, George Mason University, Fairfax, Virginia, United States. 2010.

[36] Kenney, D., & Judd, C.M. "*Estimating the Non-linear and Interactive Effects of Latent Variables*", Psychological Bulletin, 96. 201-210. 1984.

[37] Bergstrom, C. Eigenfactor.org: *Ranking and Mapping Scientific Knowledge*, from http://www.eigenfactor.org/. 2007.

[38] Browne, M.W., & Cudeck, R. "*Alternative ways of assessing model fit*", Sociological Methods and Research, 21. pp. 230-258. 1992.

[39] Hu, L., & Bentler, P. "*Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives*", Structural Equation Modeling, 6. 1-55. 1999.

[40] Hayduk, L.,&Glaser, D. N. "*Jiving the four-step, waltzing around factor analysis, and other serious fun*". Structural Equation Modeling, 7. pp. 1-35. 2000.

[41] Lykken, D.T. "*Statistical significance in psychological research*", Psychological Bulletin, 70. pp 151-159. 1968.

[42] Meehl, P. E. "*The problem is epistemology, not statistics: Replace significance tests by confidence intervals and quantify accuracy of risky numerical predictions*". In L. L. Harlow, S. A. Mulaik, & J.H. Steiger (Eds.), What if there were no significance tests?, Lawrence Erlbaum Associated, Mahwah, pp. 393-425. 1997.

[43] Marsh, H., Hau, K., & Wen, Z. "*In Search of Golden Rules: Comment on Hypothesis-Testing Approaches to Setting Cutoff Values for Fit Indexes and Dangers in Overgeneralizing Hu and Bentler's (1999) Findings*", Structural Equation Modeling, 11. 320-341. 2004.

[44] Nye, C. D., & Drasgow, F. "*Assessing goodness of fit: Simple rules of thumb simply do not work*", Organizational Research Methods, 14. pp. 548-570. 2011.

[45] Tomarken, A., & Waller, N. "*Structural Equation Modeling: Strengths, Limitations, and Misconceptions*", Annual Review of Clinical Psychology, 1. 31-65. 2005.

[46] Cohen, J. "*Statistical power analysis for the behavioral sciences*", Lawerence Erlbaum Associates, Hillsdale, New Jersey. 1988.

[47] Barrett, P. "*Structural equation modeling: Adjudging model fit*", Personality and Individual Differences, 42. pp. 815-824. 2007.

[48] Hayduk, L., Cummings, G., Boadu, K., Pazderka-Robinson, H., & Boulianne, S. "*Testing! testing! One, two, three-Testing the theory in structural equation models!*", Personality and Individual Differences, 42. pp. 841–850. 2007.

[49] Preacher, K.J., & Hayes, A.F. "*Asymptotic and resampling strategies for assessing and comparing indirect effect in multiple mediator models*", Behavior Research Methods, 40. pp. 879-891. 2008.

Analyzing the health status of the population using ordinal data

Maria Livia ŞTEFĂNESCU[*]
Research Institute for Quality of Life Romanian Academy

Abstract

We intend to estimate the health status of the people using a Gini kind index GO for measuring the inequality and a polarization indicator PO too. The both indices were applied for ordinal health data which were selected from three national representative samplings designed in the period 2003-2010 in Romania. The results evaluate the evolution level of the polarization and inequality phenomena in the health domain.

Keywords: *ordinal data, polarization and inequality indices, health status, SAH data*

JEL classification : *I14, I19, C43, C19*

1. Introduction

The data in health domain are accessible under different forms. We mention here the nominal, ordinal and cardinal (for interval scale) health data. Obviously, the indices which are used to evaluate health status of the individuals or of the nations must be highly associated with the specific kind of data.

Our research is based on there national representative Romanian sampling surveys designed in the years 2003, 2006 and 2010. In the following we intend to analyze the response at the question *Q1*: "*How do you evaluate your status of health*". The question *Q1* was addressed to all the persons belonging to the selected samplings. At this question we specified five possibilities of answer, that is: *very bad* (code 1), *bad* (code 2), *satisfactory* (code 3), *good* (code 4), *very good* (code 5).

In this context the people answers concerning the subjective self-assessment of the Romanian health status are ordinal data. More precisely, we identify practically five categories C_1 - C_5 characterized by the answer codes 1-5. The individuals from a group C_k have often more difficulties concerning their health status in contrast with the persons belonging to the superior groups as C_{k+1}, C_{k+2}, etc.

Effectively, the primary data are characterized by the frequencies f_k, $1 \le k \le m$, where the natural number f_k, $f_k \in N$, represents the frequency to have individuals in the category C_k from the specified sampling. Shortly, we will designate by $\underline{f}$ the vector which includes all the frequencies f_k, $1 \le k \le m$, that is $\underline{f} = (f_1, f_2, f_3, \dots, f_m)$.

To simplify our presentation, we will use the notation $f_{k,+}$, $1 \le k \le m$, to designate the following expression

$$f_{k,+} = f_1 + f_2 + f_3 + \dots + f_k\ , \quad 1 \le k \le m \tag{1.1}$$

The size n of the sampling is just $f_{m,+}$.

More, by $\Delta_{m,n}$ we understand the subdomain of N^m having the form

$$\Delta_{m,n} = \{\ \underline{f}\ \mid\ \underline{f} = (f_1, f_2, f_3, \dots, f_m)\ ,\ f \in N^m\ ,\ f_1 + f_2 + f_3 + \dots + f_m = n\ \} \tag{1.2}$$

In the subsequent we intend to measure two distinct social phenomena, that is the polarization and also the inequality aspects in health. In practice, the both phenomena are closely related, but we have not a strictly dependence relation between them. More precisely, one of these phenomena can affect in part the other. But for effective real situations this multifaceted relation could have very distinct intensity degrees. A concrete evaluation concerning the inequality and the polarization health levels associated to the years 2003-2010 was given for Romania living in the rural environment.

[*] Corresponding author: livia.stefanescu@gmail.com

2. The polarization index *PO* for SAH data

In the literature, for nominal, ordinal or cardinal types of data, are known a lot of indicators to measure the degree of the polarization phenomenon. These measures depend effectively on the particular type of data. So, for the class of cardinal data we point out the papers: Esteban & Ray (1994), (2012), Duclos & Esteban & Ray (2004), Chakravarty & Majumder (2001), Chakravarty (2009), Rodriguez & Salas (2003), Zhang & Kanbur (2001), Wang & Tsui (2000), Foster & Wolfson (2010), Bossert & Schworm (2008), Deutsch & Silber & Yalonetzky (2013). Some polarization indices were also proposed for categorical data. We remark here the work of Permanyer & D'Ambrosio (2013). Also we mention especially the following references focused on ordinal data: Apouey (2007), (2010), Montalvo & Reynal-Querol (2005), Apouey & Silber (2013), Kobus (2014), Chakravarty & Maharaj (2012), Makdissi & Yazbecky (2014).

Strong related with a polarization measure are different indices of variation with could now be applied for ordinal data too. See, for example Berry & Mielke (1992), Blair & Lacy (1996), (2000).

In many recent papers are studied diverse techniques to transform a kind of data in a new type of data. In this context we underline the review of Van Doorslaer & Jones (2003). Therefore, an index used initially for a specific data can be modified to use for another data types.

Apouey (2007) proposed an one parameter class of polarization indices applied to ordinal SAH (self-assessed health) data. These types of indices depend on the probabilities of the ordinal categories. In the present paper we will express one of these Apouey polarization indicators in function of the frequency associated to every ordinal class C_1- C_m . More precisely, for any $\underline{f} \in \Delta_{m,n}$ the polarization index *PO* has the following expression :

$$PO(\underline{f}) = 1 - \frac{1}{m-1}\sum_{k=1}^{m-1} |\, 2\, f_{k,+} / n - 1| \tag{2.1}$$

Apouey (2007) established two main axioms which must be fulfilled compulsory by any polarization index applied to ordinal data. All his proposed indicators satisfy the both axioms and more these indices have good properties to measure the polarization phenomenon in health domain, Apouey (2007), (2010).

It is easy to show by a direct calculus that

Proposition 2.1. For any $\underline{f} \in \Delta_{m,n}$ we have always the inequalities

$$0 \le PO(\underline{f}) \le 1 \tag{2.2}$$

3. An inequality coefficient *GO* for ordinal data

For an arbitrary ordinal variable X characterized by the frequencies $\underline{f}$ is not able to operate correctly with the mean $\mu(X)$. Indeed, a score k attached to the ordinal category C_k of X is subjective. These scores k establish only the hierarchy of the classes C_k , $1 \le k \le m$. More, the value k is not often relevant when is used to characterize all the individuals belonging to the same class C_k . Any other set of real values $v_1 < v_2 < v_3 < \ldots < v_m$ could define the weights of the ordered groups C_k , $1 \le k \le m$.

The methodology to measure the inequality phenomena was intensively developed in the last 50 years. We mention only a bit from this multitude of references : Atkinson (1970), Chakravarty (2009), Duclos & Araar (2006), Haughton & Khandker (2009), Betti & Lemmi (2008), Foster & Seth & Lokshin & Sajaia (2013).

Gini coefficient $G(X)$ is the most popular index to evaluate the degree of inequality for a distribution of cardinal data X. This very known indicator was proposed by the famous Italian economist and statistician Corrado Gini at the beginning of the twenty century (Gini (1909a),(1909b)). We remind that Gini index $G(X)$ is based on the Lorenz curve where the mean of the cardinal variable X plays an essential role.

Since the mean $\mu(X)$ of an ordinal variable X has not a clear interpretation we can't apply correctly the classical Gini coefficient $G(X)$ to measure the inequality from X.

More indicators were proposed to evaluate inequality aspects in the case of ordinal data. We mention here some references regarding different approaches: Allison R. A., Foster J. E. (2004), Abul & Yalcin (2008), Madden (2010), Giudici & Raffinetti (2011).

Giudici & Raffinetti (2011) adapted the classical Gini coefficient $G(X)$ to any ordinal variable X.

More exactly, for all individuals belonging to the class C_k we associate the same rank r_k, $1 \le k \le m$. But, the rank r_k is modified in function of the ordinal distribution $\underline{f}$ which is analyzed (Giudici & Raffinetti (2011)). So

$$r_1 = 1 \quad , \quad r_k = r_{k-1} + f_{k-1} \text{ for any } 2 \le k \le m \tag{3.1}$$

The ordinal Gini index $GO(\underline{f})$ is based on the Lorenz curve defined by the points having the cartesian coordinates $(f_{k,+} / f_{m,+} \, , q_k / q_m)$, $0 \le k \le m$, where

$$q_k = \sum_{j=1}^{k} r_j f_j \quad , \quad 1 \le k \le m \tag{3.2}$$

with the convention $f_{0,+} = q_0 = 0$.

With these notations we define (Giudici & Raffinetti (2011))

$$GO(\underline{f}) = 1 - \sum_{k=1}^{m} (q_k / q_m + q_{k-1} / q_m)(f_{k,+} / f_{m,+} - f_{k-1,+} / f_{m,+}) \tag{3.3}$$

After a straightforward computations we deduce too

Proposition 3.1. For any $\underline{f} \in \Delta_{m,n}$ the following inequalities are true

$$0 \le GO(\underline{f}) \le 1 \tag{3.4}$$

4. Some differences between *PO* and *GO* indicators

Therefore, the indicators $PO(\underline{f})$ and $GO(\underline{f})$ can be used successfully to measure the polarization degree, respectively the inequality level for an arbitrary distribution of frequencies $\underline{f}$ which characterize an ordinal variable X. We established that the both coefficients vary in the interval $[0 , 1]$.

But, in practice, between the polarization and the inequality phenomena there is a complex relation of dependence. For this reason, from very closed values of the polarization index *PO* is possible to often obtain very different values of the inequality coefficient *GO*.

For the subsequent we will prove this assertion taking into consideration the frequency distributions $\underline{f}$, $\underline{f} \in \Delta_{3,500}$, of sixteen ordinal variable X precised in *Table 4.1* .

In *Graphic 4.2* are represented the points j , $1 \le j \le 16$, having the cartesian coordinates ($PO(X_j)$, $GO(X_j)$), where X_j are the ordinal variables with the frequency distributions $\underline{f}$ from *Table 4.1*. The scatter of the points j from *Graphic 4.2* suggests us that there is not a simple dependence relation between the polarization and inequality aspects of the variables X_j . More, none of the following inequalities $PO(X) < GO(X)$ or $PO(X) > GO(X)$ are always true for any ordinal variable X (see *Graphic 4.2*).

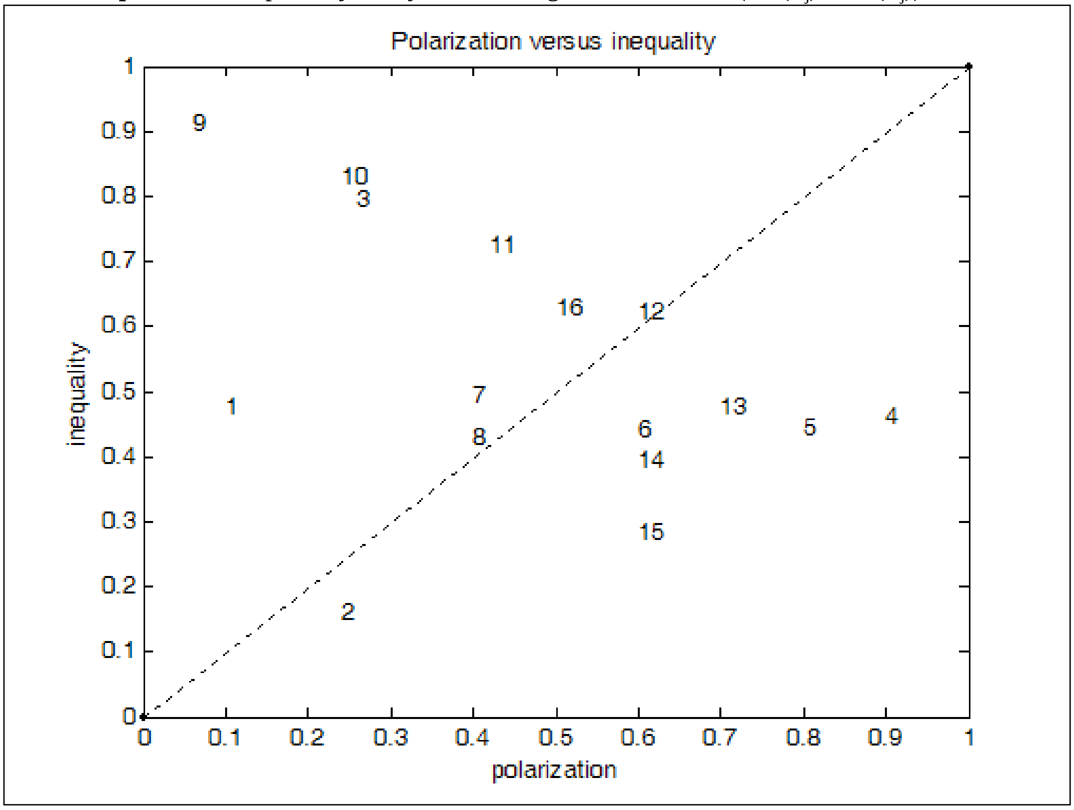
Having in mind this conclusion, for an accurate interpretation of the health status of a given population, we recommend to use together the two indicators *PO* and *GO*.

Table 4.1. The frequency distributions f of the ordinal variables X_j

(m=3 , n=500).

j	f_1	f_2	f_3	$PO(X_j)$	$GO(X_j)$
1	25	450	25	0.10	0.4775
2	20	80	400	0.24	0.1611
3	400	70	30	0.26	0.7975
4	225	50	225	0.90	0.4622
5	200	100	200	0.80	0.4472
6	150	200	150	0.60	0.4428
7	50	300	150	0.40	0.4963
8	150	300	50	0.40	0.4301
9	480	10	10	0.06	0.9131
10	420	40	40	0.24	0.8318
11	360	70	70	0.42	0.7258
12	300	100	100	0.60	0.6241
13	200	150	150	0.70	0.4789
14	100	200	200	0.60	0.3975
15	100	100	300	0.60	0.2837
16	300	150	50	0.50	0.6287

Graphic 4.2. The points j , $1 \le j \le 16$, having the coordinates $(PO(X_j)\,,\,GO(X_j))$.



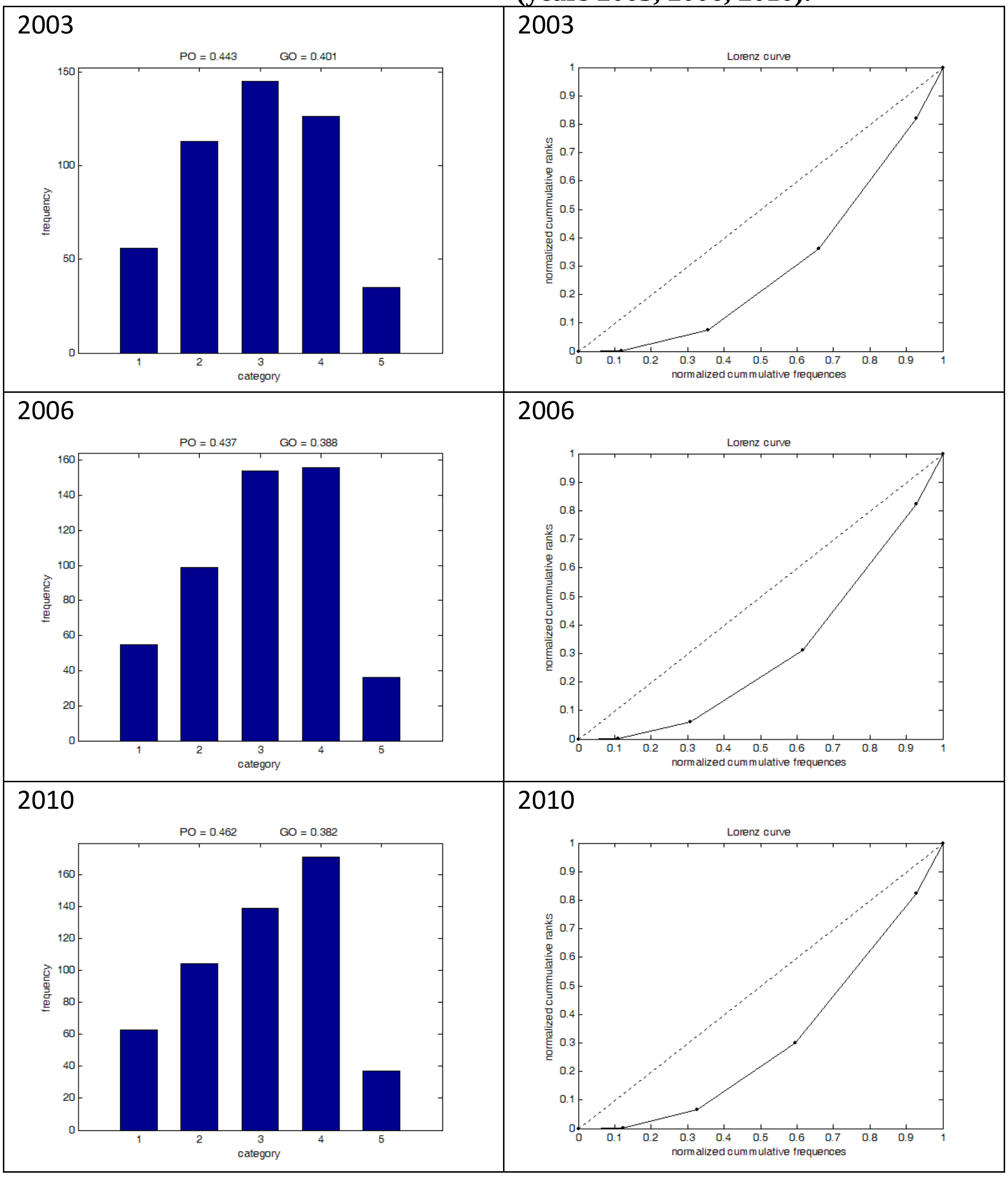
5. An application

We will analyze the evolution of the health status for Romanian living in rural, during the years 2003-2010. In this context we used three representative samplings concerning the quality of life of the Romanian people. The sampling surveys were designed at Research Institute for Quality of Life, Romanian Academy, in the years 2003, 2006 and 2010.

The frequency distributions f at the question *Q1* where illustrated in *Graphic 5.1* together with the associated Lorenz curves obtained after the Giudici & Raffinetti (2011) methodology.

From *Graphic 5.1* we remark that the studied distributions and their Lorenz curves are very similar. For this reason is very difficult to evaluate the progress of the Romanian health status in the period 2003-2010.

Graphic 5.1. The frequency distribution and the Lorenz curve at the question *Q1* (years 2003, 2006, 2010).

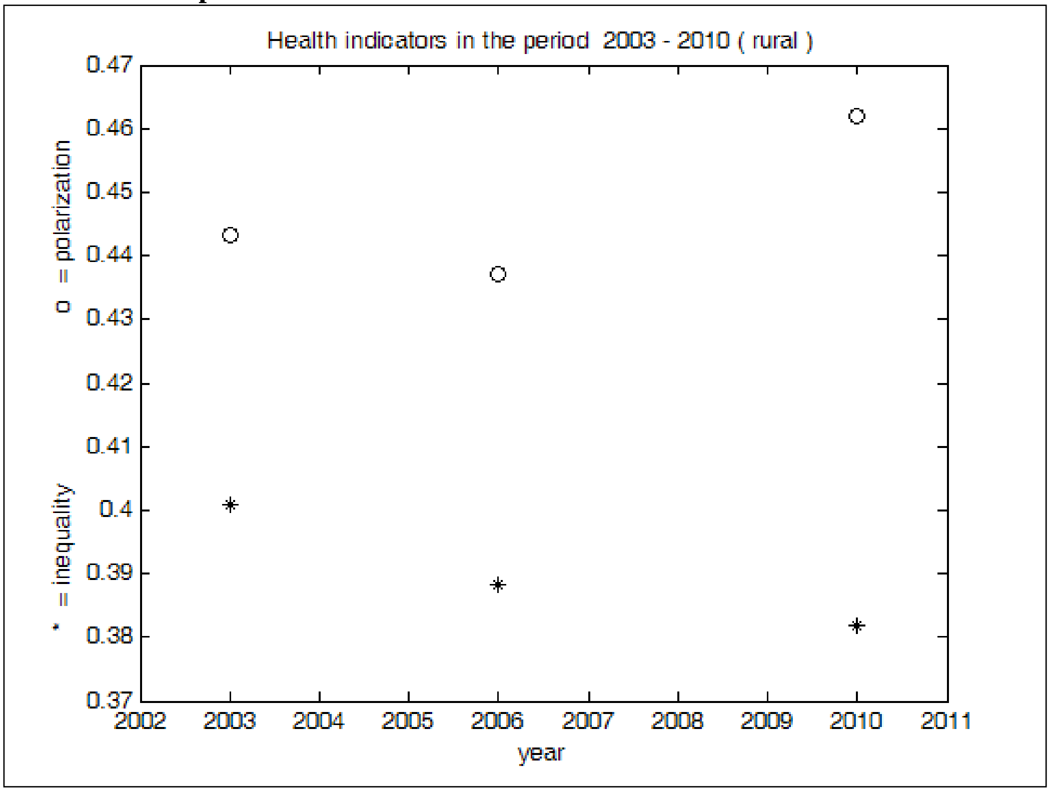


By applying the polarization PO and inequality GO indices to the selected samplings having the volume n we obtained the values mentioned in *Table 5.2*. *Graphic 5.3* suggests the evolution of the rural Romanian health status in the period 2003-2010. We remark an easy decreasing of the inequality and an enough consistent increasing of the polarization aspects.

Table 5.2. Synthesis results regarding the selected samplings.

year	2003	2006	2010
n	475	500	514
PO	0.443	0.437	0.462
GO	0.401	0.388	0.382

Graphic 5.3. The evolution of the rural Romanian health status.



6. Partial conclusions

This proposed methodology, based on the polarization and inequality indices for ordinal data, was applied to evaluate health status for rural Romanian people in the period 2003-2010.

The indices *PO* and *GO* measure two distinct aspects of the reality, that is the polarization and the inequality phenomena. The two coefficients *PO* and *GO* vary inside the interval [0, 1]. Considering sixteen possible answer distributions we proved that the polarization and inequality situations can be often close related but not identical. In reality, the increase of the polarization level into a community do not compulsory involve the grow of the inequality degree inside that population (see Graphic 4.2). For a precise interpretation of the evolution for the population health status we recommend to use together the both indicators *PO* and *GO*.

For the rural Romanian communities we have a stable decrease of the *GO* inequality coefficient in the period 2003-2010. But the behavior of the *PO* polarization indicator is different. So, after a light decreasing of the *PO* values it results finally an enough consistent increase of the polarization (Graphic 5.3).

To apply correctly our proposed approach is necessary to study the properties of the indices *PO* and *GO* and in addition, to precise clearly the concrete cases when the both indicators can act in the same direction. It is essential to use more indices to measure distinct aspects of a complex reality. Primary, our option regards an index to identify a positive evolution of a concrete situation. From complementary studies we must also establish some reference distributions considered as equilibrium circumstances for the society.

In the future, using the same kind of processing, we intend to compare the self- assessed health answers at the question *Q1* of the people which is divided in more groups. So, the individual health data must be analysed in contrast for different age categories, taking also into consideration the gender of the persons and their domicile, the household income, families with more children, the unemployed people, individuals with disabilities or other deprivated groups.

References

[1] Abul Naga R., Yalcin T. (2008). Inequality measurement for ordered response health data. *Journal of Health Economics* 27, 1614-1625.
[2] Allison R. A., Foster J. E. (2004). Measuring health inequality using qualitative data. *Journal of Health Economics* 23, 505-524.
[3] Apouey Bénédicte, (2007). Measuring health polarization with self-assessed health data. *Health Economics*, 16 (9), 875–894.
[4] Apouey Bénédicte. (2010). On measuring and explaining socioeconomic polarization in health with an application to French data. *Review of Income and Wealth*, *56*(1), 141-170.
[5] Apouey Bénédicte, Silber Jacques, (2013). Inequality and bi-polarization in socioeconomic status and health: Ordinal approaches. PSE Working Papers, no. 2013-30. https://halshs.archives-ouvertes.fr/halshs-00850014.
[6] Atkinson A. B., (1970). On the measurement of inequality. *Journal of Economic Theory* 2, 244-263.
[7] Berry K.J, Mielke P.W., (1992). Indices of ordinal variation, *Perceptual and Motor Skills*, 74, 576-578.
[8] Betti Gianni, Lemmi Achille (eds.), (2008). Advances on income inequality and concentration measures. Routledge – Taylor & Francis Group, London.
[9] Blair, J., Lacy M.G., (1996). Measures of variation for ordinal data as functions of the cumulative distribution. *Perceptual & Motor Skills*, *82*(2), 411-418.
[10] Blair J, Lacy MG., (2000). Statistics of ordinal variation. *Sociological Methods and Research*, 28(251), 251-280.
[11] Bossert W., Schworm W., (2008). A Class of Two-Group Polarization Measures. *Journal of Public Economic Theory* **10**, 1169-1187.
[12] Chakravarty S.R., (2009). Inequality, polarization and poverty. *Advances in Distributional Analysis*. Springer, New York.
[13] Chakravarty S.R, Maharaj B., (2012), Ethnic polarization orderings and indices. *Journal of Economic Interaction and Coordination*. 7(1), 99-123.
[14] Chakravarty S.R, Majumder A., (2001). Inequality, polarization and welfare: theory and applications. *Australian Economic Papers* **40**, 1-13.
[15] Deutsch J., Silber J., Yalonetzky G., (2013). On relative bi-polarization and the middle class in Latin America. A look at the first decade of the 21st century. Paper presented at the XX Encuentro de Economía Pública, Sevilla, Spain.
[16] Duclos J-Y., Esteban J., Ray D. , (2004). Polarization: Concepts, measurement, estimation. *Econometrica* **72**, 1737-1772.
[17] Duclos Jean-Yves, Araar Abdelkrim, (2006). Poverty and equity: measurement, policy and estimation with DAD. Kluwer Academic Publishers, Boston.
[18] Esteban J., Ray D., (1994). On the measurement of polarization" *Econometrica* **62**, 819-852.
[19] Esteban J., Ray D., (2012). Comparing polarization measures. In Garfinkel M R, Skaperdas S (Eds), The Oxford Handbook of the Economics of Peace and Conflict, Chapter 7, Oxford University Press, New York.
[20] Foster J., Wolfson M.C., (2010). Polarization and the decline of the middle class: Canada and the US, *Journal of Economic Inequality*, 8, 247-273.
[21] Foster James, Seth Suman, Lokshin Michael, Sajaia Zurab, (2013). A unified approach to measuring poverty and inequality - Theory and Practice. Streamlined analysis with ADePT software. The World Bank, Washington D.C.
[22] Gini C., (1909a). Il diverso accrescimento delle classi sociali e la concentrazione della ricchezza. Giornale degli Economisti, 38, 27–83.
[23] Gini C., (1909b). Concentration and dependency ratios (in Italian). English translation in Rivista di Politica Economica, 87 (1997), 769–789.
[24] Giudici Paolo, Raffinetti Emanuela, (2011). A Gini concentration quality measure for ordinal variables. *Quaderni del Dipartimento di Economia, Statistica e Diritto*, Uuniversità di Pavia, Serie Statistica, no. 1/2011 (April 7, 2011), 18 pag.
[25] Haughton Jonathan, Khandker Shahidur R., (2009). Handbook on Poverty and Inequality. The World Bank, Washington D.C.
[26] Kobus Martyna, (2014). On the measurement of polarization for ordinal data. ECINEQ 2014 – 325, March 2014.
[27] Madden D., (2010). Ordinal and cardinal measures of health inequality: an empirical comparison. *Health Economics Letters* 19, 243-250.
[28] Makdissi Paul, Yazbecky Myra, (2014). Robust Wagstaff orderings of distributions of self-reported health status. University of Ottawa (Canada), University of Queensland (Australia), Report September 2014, 39 pages.
[29] Montalvo J., Reynal-Querol M., (2005). Ethnic polarization, potential conflict and civil wars. *American Economic Review*, **95**, 796-815.
[30] Permanyer Inaki, D'Ambrosio Conchita, (2013). Measuring social polarization with ordinal and categorical data. *Journal of Public Economic Theory*, 11 pag.
[31] Rodriguez J., Salas R., (2003). Extended bi-polarization and inequality measures. *Research on Economic Inequality*, **9**, 69–83.
[32] Zhang X., Kanbur R., (2001). What difference do polarization measures make? An application to China. *Journal of Development Studies*, **37**, 85-98.
[33] Van Doorslaer E., Jones A.M., (2003). Inequalities in self-reported health: validation of a new approach to measurement. *Journal of Health Economics* 22, 61-87.
[34] Wang Y.Q, Tsui K.Y., (2000). Polarization orderings and new classes of polarization indices. Journal of Public Economic Theory, 2, 349-363.
[35] ***** Data bases "Diagnosis for Quality of Life", Research Institute for Quality of Life, Romanian Academy, (2003, 2006, 2010).

The issue of statistical power for overall model fit in evaluating structural equation models

Richard HERMIDA[1], Joseph N. LUCHMAN[2], Vias NICOLAIDES[3], Cristina WILCOX[4 *]
[1]George Mason University, rhermida3@gmail.com, 3575 Owasso Street, Shoreview, MN, USA, 55126
[2]George Mason University, jluchman@gmu.edu, 4400 University Drive, Fairfax, VA, USA, 22030
[3]George Mason University, vnicolai@gmu.edu, 4400 University Drive, Fairfax, VA, USA, 22030
[4]George Mason University, cfwilcox9@gmu.edu, 4400 University Drive, Fairfax, VA, USA, 22030

Abstract

Statistical power is an important concept for psychological research. However, examining the power of a structural equation model (SEM) is rare in practice. This article provides an accessible review of the concept of statistical power for the Root Mean Square Error of Approximation (RMSEA) index of overall model fit in structural equation modeling. By way of example, we examine the current state of power in the literature by reviewing studies in top Industrial-Organizational (I/O) Psychology journals using SEMs. Results indicate that in many studies, power is very low, which implies acceptance of invalid models. Additionally, we examined methodological situations which may have an influence on statistical power of SEMs. Results showed that power varies significantly as a function of model type and whether or not the model is the main model for the study. Finally, results indicated that power is significantly related to model fit statistics used in evaluating SEMs. The results from this quantitative review imply that researchers should be more vigilant with respect to power in structural equation modeling. We therefore conclude by offering methodological best practices to increase confidence in the interpretation of structural equation modeling results with respect to statistical power issues.

Keywords*: Statistical Power, Structural Equation Modeling, Measurement, Statistics, Research Methods.*

Introduction

Structural equation modeling (SEM) is an increasingly popular analysis framework in many areas of scientific inquiry, including psychology, management, and sociology (MacCallum & Austin, 2000). However popular, SEM is highly complex and its statistical mechanics are often not well understood by users. As a result, SEM has the potential to be misapplied, affecting the interpretation of scientific findings. For example, researchers often allow data to dictate which measurement errors should correlate, as opposed to appealing to *a priori theory*, which can make poorly fitting models appear "passable" (Hermida, Conjar, Najab, Kaplan, & Cortina, 2010; Landis, Edwards, & Cortina, 2009), engage in empirical redundancy of constructs (Le et. al., 2010), or give inappropriate statements regarding causality without the backing of theoretical assumptions within SEM (Pearl, 2009; 2012; 2014).

While there are a number of methodological subtleties to SEM (Bagozzi & Yi, 2012; Bentler & Bonnett, 1980; MacCallum & Austin, 2000), one issue that has largely escaped the attention of SEM users is consideration of statistical power in SEM with respect to overall model fit. Understanding power in model fit is important because power reflects the probability that a model will differentiate between good and bad theory-implied constraints or specifications (Cohen, 1988; 1992). Since overall model fit is one of the main standards by which empirical evidence provided by structural equation models (SEMs) are judged, understanding issues related to power of SEM fit indices can provide the understanding necessary to effectively conduct SEM-based data analysis, improve inferences regarding SEMs, and increase the rate of scientific progress enjoyed by users of SEM.

The primary purposes of the present study are fivefold. First, we wish to inform researchers how sampling variability and power can potentially harm our inferences regarding our judgments of SEMs. Second, we wish to give sufficient background as to how power is calculated in a structural equation modeling context for overall model fit. Third, we wish to benefit researchers by explaining the main influencers of power so as to aid researchers in study design efforts. Fourth, we wish to examine certain methodological situations that could signal a need for the researcher to pay particular attention to statistical power. Fifth, we wish to conduct a

[*] Corresponding author: rhermida3@gmail.com.

quantitative review regarding power in order to a) gain understanding of the distribution of power as it exists in published journal articles, and b) test the degree to which power is associated with certain methodological situations.

The results of our review speak to the level of uncertainty related to the *decisions* scholars make about model fit. Thus our survey provides best practices to researchers in terms of "powering" their study to detect non-trivial problems related to model misfit. Our research is primarily based on MacCallum, Browne, and Sugawara (1996), who conducted the pioneering work in developing the concept of statistical power in overall model fit indexes. It is our hope that after reading this review, researchers will understand the statistics of power as it relates to overall model fit, and moreover be able to identify how some methodological issues might provide signals to the researcher regarding power of their tested models.

Overview of Power

The primary strength of SEM, and the root of its popularity, is in integrating the measurement model focus of factor analysis with a structural or theoretical model that has been the focus of path analytic or regression modeling. An important issue to both measurement and structural models is examining how well the model implied by theory fits to the data collected (Kaplan, 1995; Specht, 1975). To the extent that a theoretical model fits empirical data, the theoretical model is confirmed, as it is a plausible explanation for the covariance structure amongst the variables (Mulaik et al., 1989). The issue of how to evaluate model fit is complicated, and opinions have yet to converge on the most appropriate method. As a result of different opinions related to how fit should be assessed (Barrett, 2007; Hayduk et al., 2007), a number of model fit indices have been developed for SEMs (Bagozzi & Yi, 1998; Bentler, 1990), each with different properties across a number of dimensions, such as absolute vs. relative fit (McDonald & Ho, 2002).

Of the many structural equation model fit indices available in the literature, the RMSEA is a popular index of absolute fit (i.e., it is not *relative* to the null model as are indices such as the confirmatory fit index or CFI) and is noted for its insensitivity to estimator by comparison to relative fit indices (Steiger, 1990, Steiger & Lind, 1980; Sugawara & MacCallum, 1993). The RMSEA is, most fundamentally, a function of the model chi-square value, but also includes the model degrees of freedom, and sample size as seen in Equation 1.

$$\mathrm{RMSEA} = \frac{\sqrt{(\chi^2 - \mathrm{df})}}{\sqrt{(\mathrm{df})(N-1)}} \tag{1}$$

Important to note is that model degrees of freedom, for traditional maximum likelihood SEM with continuous factor indicators, are computed by obtaining the total number of elements in the variance-covariance matrix that can be analyzed minus the number of estimated parameters. Readers who wish to review these concepts should consult a more in-depth explanation of degrees of freedom by Rigdon (1994).

The functional form of the RMSEA in Equation 1 can be explained by noting first that nested within the RMSEA index is the assumption that the model being estimated is misspecified to some extent and, consequently, the model chi-square statistic follows what is known as a *non-central chi-square distribution*. The non-central chi-square distribution can be thought of as the chi-square distribution when chi-square possesses any non-zero value (Patnaik, 1949). Relevant to this review, the non-central chi-square *distribution* comes into play when one wants to know the chance that chi-square exceeds a particular chi-square value when the true population value of chi-square is non-zero (Cox & Reid, 1987). Of chief concern is the noncentrality parameter, which is simply the parameter that occurs in a distribution that is a transformation of the normal distribution (like the non-central chi-square distribution), and how this parameter relates to power.

Let us walk through a brief statistical sample to illustrate these interactions. At this point, the reader is encouraged to walk through these steps in order to become more intimate with the procedure. It would also be helpful to have the seminal work (MacCallum et al., 1996) on hand for easy access to referenced graphs and figures, as well as tools to easily calculate power on hand (Preacher & Coffman, 2006).

Suppose a researcher tested a model with 20 degrees of freedom, a sample size of 200. Next suppose the researcher wanted to obtain the probability of rejecting the null hypothesis that the obtained chi-square would be equal to or less than RMSEA = .05 (i.e. – χ^2 = 29.95 in this context), if the true value was RMSEA = .08 (i.e. – χ^2 = 45.47 in this case), with alpha equal to the traditional .05 level.

Calculation of the noncentrality parameters associated with the null and alternative hypothesis is easily accomplished with the following formula:

$$\lambda = (N-1)(df)(RMSEA^2) \tag{2}$$

Therefore, in this example the noncentrality parameter associated with the null hypothesis (ncp0) is 9.95, and the noncentrality parameter associated with the alterative hypothesis (ncpa) is 25.47. It is important to note at this point that the extent to which the model is correctly specified, the model is better approximated by the "central" chi-square (which has a mean or expected value equal to the model degrees of freedom) and λ approaches 0. The more the model is misspecified is the degree to which the noncentrality parameter and central chi-square diverge (see figure 1, pg. 136; MacCallum et al., 1996). We will now turn to specific hypothesis tests taken from MacCallum et al. (1996) that will be relevant for our quantitative review.

MacCallum et al. (1996) use the non-central chi-square distribution to propose three hypothesis tests which evaluate different aspects of model fit by assessing the degree of overlap between a pair of non-central chi-square distributions (i.e., the null and alternative distributions). The first test proposed by MacCallum et al. is a test of *exact fit*. Exact fit is analogous to the central chi-square test of model fit in that it evaluates whether a model's fit to the data is sufficiently good to be "exactly" as the specified model dictates. The null value used for the exact fit test is not, however, 0 (i.e., no model discrepancies from the data), but rather some very small RMSEA value, such as .01. In the instance that the estimated RMSEA is sufficiently large—that is, large enough to be significantly different from a small value such as .01 (MacCallum et al., 1996)—then we can infer that the fit of the model is not likely to be exact.

A departure from the exact fit idea is proposed through the second test of *close fit*. Close fit differs from exact fit in that it evaluates whether the confidence interval about RMSEA centers around, but does not exceed, 0.05. Hence, the purpose of the close fit test is to evaluate a model in which the null hypothesis is .05, with an alternative hypothesis that is *larger* than .05—suggesting that the RMSEA in the population is likely to be ≤ .05. Models that are not significantly larger than .05 are inferred to have a close fit, and although close fit is not exact, MacCallum et al. argue that the data approximates the model "closely" or well enough to be of use scientifically.

The final test proposed is *not close fit*. Not close fit supplements conceptual deficiencies in the previous tests by evaluating whether the estimated RMSEA ≥.05 therefore indicating that the model *is* likely to be a poor fit to the data. Similar to the test of close fit, the null distribution centers on RMSEA of .05, however the alternative distribution for not close fit is fixed at a value less than .05. Thus, as MacCallum et al. show (p. 136-8), the not-close fit test adds to the information provided by the close fit test by distinguishing between situations where the RMSEA's confidence interval falls relatively close to a value of .05. Specifically, based on the pattern of tests accepted and rejected the researcher can triangulate on the likely "true" RMSEA of the model. For example, when the test of not close fit is rejected and the test of close fit is accepted, a researcher can infer that the true RMSEA falls somewhere below .05. Alternatively, when close fit is rejected but not close fit is accepted, a researcher can infer that the true RMSEA falls above .05. A final possibility is that both tests are accepted, which suggests that RMSEA's confidence interval centers around .05. In combination, all three tests permit a researcher to evaluate the degree of model fit more flexibly than using only "rule of thumb" cut off values for fit indices or a single chi-square test, as the "three test" procedure admits to the idea that degrees of freedom and sample size play an important role in the precision of the estimates obtained using SEM and the level of uncertainty we have about their true values—which is thereby reflected in the confidence interval associated with the estimated model's fit (McQuitty, 2004).

The hypothesis testing procedure for tests of overall model fit in SEM differ slightly from the way hypothesis tests are structured for traditional statistics such as ANOVA F-tests or t-tests. To be precise, accepting the null is a sought after result in the case of close and exact fit. Moreover, accepting the null of a not close fit test does not imply *unacceptably poor* fit, but only *not good* fit. Whereas the interpretation of the tests differs, the logic of the hypothesis testing procedure does not differ from the usual procedure as outlined in introductory statistics texts.

Hypothesis testing for RMSEA proceeds by evaluating the distributions of two values of RMSEA—which for the sake of consistency with MacCallum, et al., we will, for the remainder of the present section, refer to as ϵ—

the null value: ϵ_o and the alternative value: ϵ_a. Using ϵ_o and ϵ_a we then can compare the non-central chi-square distribution associated with ϵ_o to the non-central chi-square distribution associated with ϵ_a. The overlap between ϵ_o and ϵ_a is overall covariance model power—based on ϵ. The null hypothesis value for each of the tests (exact, close, or not close) we describe above (e.g., .01 for exact fit) and similar to differences between means in a t-test, the differences between ϵ_o and ϵ_a values can be conceptualized as the "effect size" component that factors into the power calculations for t-tests. When $\epsilon_o > \epsilon_a$, power is estimated as:

$$\pi = P(\chi_d^2, \lambda_a < \chi_c^2) \tag{3}$$

Whereas when $\epsilon_o < \epsilon_a$, power is estimated as:

$$\pi = P(\chi_d^2, \lambda_a > \chi_c^2) \tag{4}$$

In both cases, χ_c^2, χ_d^2,, and λ_a represent the non-central chi-square distributions associated with ϵ_o and ϵ_a and π represents statistical power . When $\epsilon_o > \epsilon_a$, the power is measured as the portion of χ_d^2, or λ_a, that lies to the left of alpha (α) in the left tail of χ_c^2. When $\epsilon_o < \epsilon_a$, power is measured as the portion of χ_d^2, or λ_a, that lies to the right of the critical value α in the right tail of χ_c^2. When $\epsilon = 0$, the non-centrality parameter λ is also 0 (see Equation 3). In this case, ϵ is distributed as a regular, central chi-square and perfect fit is implied. We will now move to a discussion of statistical factors that influence power.

Factors Affecting Power

One major area that influences power (all else equal), is the closeness of the RMSEA values associated with the null and alternative hypotheses. Closeness of null and alternative RMSEA values is *negatively* associated with power. That is, the closer the RMSEA values, the more power decreases.

This pattern occurs because to the degree that null and alternative RMSEA values are similar is the degree to which the non-central chi-square distributions overlap. To the degree the distributions overlap, is the degree to which there is a lack of ability to find area under the alternative non-central chi-square distribution that is beyond the critical value associated with the null hypothesis *and* not overlapping with the non-central chi-square distribution associated with the null. To use our previous examples, the ncp0 and ncpa associated with RMSEA = .05 and RMSEA = .08 are 9.95 and 25.47 (difference of 15.52) and generate a power coefficient of .45. If the null was moved to RMSEA = .07, the ncp0 would shift to 19.50 (difference of 5.97). This would cause the null and alternative distributions to move closer together, creating more overlap and less power (in this case .13). However, if the null was moved to RMSEA = .01, the ncp0 would shift to 0.40 (difference of 25.07). This would cause the null and alternative distributions to move further apart, creating less overlap and more power (in this case .88). The bottom line here is that the more differentiation there is between the null RMSEA and the alternative RMSEA, the more power will increase, all else being equal.

Sample size has a positive association with power. That is, as sample size increases, power increases. This is because as sample size increases, the ability for the null and alternative noncentrality parameters to separate themselves from one another increases as well. For example, in our running example with a sample size of 200, the difference between the ncp0 and ncpa was 15.52 (derived from 25.47-9.95), which equates to a power coefficient of 0.45. If the sample size is increased from 200 to 500, this difference increases to 38.92 (derived from 63.87-24.95), and drives power to 0.86. If the sample size decreased to 100, the difference between the parameters would drop to 7.72 (derived from 12.67-4.95), causing power to 0.24. This is because the individual noncentrality parameter is calculated via a multiplicative term involving degrees of freedom, sample size, and the square of the RMSEA null or alternative hypothesis in question (see equation 2). Ultimately, as sample size increases, power will increase as well, all else being equal.

A more subtle influence on power is the degree of model misspecification. All else being equal, it is easier to obtain power as model misfit increases. For example, with df = 20, N = 200, RMSEA null = .00, and RMSEA alt = .05, power is only 0.40. However, if the null and alt RMSEA were .05 and .10, power would increase to 0.84. While these examples are contrived to illustrate the general principle, the general theme here is that the *more precise your models, and the more precise of a comparison you wish to make, the more difficult it is to obtain power*. Just as before, this occurs because the noncentrality parameters are connected to degrees of freedom, sample sizes, and RMSEA. For high RMSEA values, there is more *potential* for the ncp0 and ncpa to differ than for low values of RMSEA, holding all else equal. The bottom line here is that as the degree of model misfit increases, power will increase as well, all else being equal.

Finally, the relationship between degrees of freedom and power is positive. That is, as degrees of freedom increase, power increases as well. This relationship takes place on two different fronts. First, degrees of freedom impact obtained noncentrality parameters. As the degrees of freedom increase, so too does the noncentrality parameter. Applied to the current context, as degrees of freedom increase, the noncentrality parameters associated with the null and alternative hypotheses increase, but the *degree of difference* between the noncentrality parameters also increases. This is because the individual noncentrality parameter is calculated via a multiplicative term involving degrees of freedom, sample size, and the square of the RMSEA null or alternative hypothesis in question. For example, with df = 5, N = 200, RMSEA null = .05, and RMSEA alt = .08, power is only 20, with the ncpa and ncp0 possessing a difference of 3.88 (6.37-2.49). However, if the degrees of freedom were increased to 50, power would increase to 0.73, with the ncpa and ncp0 possessing a difference of 34.92 (57.31-22.39).

A more obscure way that degrees of freedom impact power is through manipulation of the shape of noncentral chi-square distributions. This is because the shape (variance) associated with the noncentral chi-square distribution is dependent on both degrees of freedom and the noncentrality parameter, as represented in the formula below:

$$2(\mathrm{df} + 2\lambda) \tag{5}$$

This can be seen further by examining the formulas associated with the skewness and kurtosis of the noncentral chi-square distribution, as shown in formulas 6 and 7, respectively:

$$\frac{2^{1.5} * (\mathrm{df} + 3\lambda)}{(\mathrm{df} + 2\lambda)^{1.5}} \tag{6}$$

$$\frac{12(\mathrm{df} + 4\lambda)}{(\mathrm{df} + 2\lambda)^{1.5}} \tag{7}$$

To the degree the noncentral chi-square distributions change, power will change as well, all else being equal. The bottom line here is that as the number of degrees of freedom increases, power will increase as well, all else being equal.

To summarize the described effects on power to detect model misfit, power increases as sample size, degrees of freedom, difference between null and alternative RMSEA values, and degree of model misfit increase. It is important to keep in mind that all of the aforementioned effects were described in the context of all other influences being held constant. It is critical to note that in reality, all of these elements interact with one another to produce statistical power, and consequently, it is possible to have several of the elements oriented towards low power, but to have a single element that is so strong as to compensate for the weakness of the other elements in producing power, or vice versa.

To give an extreme example in order to illustrate the principle, suppose a researcher sought to find power of a model that had extremely few degrees of freedom (5) and null and alternative RMSEA values that were both small *and* close together (.00 and .01). While these elements would generate low power in most situations, a researcher that obtained a sample size of 50,000 would obtain a power coefficient of .98. Conversely, suppose a researcher sought to find power about a model that had 70 degrees of freedom, a sample size of 500, and an alternative RMSEA value that had a fairly higher degree of misfit (.08). While these elements would generate high power in most situations, if the null RMSEA was set at .07, power would only be 0.53.

Ultimately, it is our desire for researchers to understand the main influencers of power, so that researchers can more easily ascertain power in different scenarios, appreciate what might need to be done to obtain more power in a particular research setting, and ultimately engage in meaningful power analysis at the planning stages of the research process.

Rationale for Quantitative Review

Hypothesis testing and power estimation for overall model fit is conceptually identical to hypothesis testing for less technically-complicated statistical analyses such as bivariate correlation or analysis of variance (ANOVA). Unfortunately, until relatively recently, no computational algorithm or computer program has been available from the literature to allow practicing researchers to *easily* compute a priori power values for the RMSEA index (see Preacher & Coffman, 2006). Moreover, the technical documentation of the RMSEA hypothesis testing procedure was presented in a very technical way by MacCallum et al. (1996). We believe that the combined influence of both of these factors have contributed to the relative neglect of power vis-à-vis overall model fit in structural equation modeling.

As opposed to the use of hypothesis testing, SEMs in the literature are usually evaluated on the basis of commonly accepted point estimate "cut-offs" such as .05 for "excellent" or .08 for "adequate" fitting models (Chen, Curran, Bollen, Kirby, & Paxton, 2008). The use of cut-off values are necessary, however, using only cut-off values and omitting hypothesis tests altogether does not allow a researcher to account for sampling variability, as we note above. For example, our confidence in a model with a RMSEA of .06 and a standard error of .01 is very different from the same RMSEA value with a standard error of .05. In particular, a RMSEA with a smaller standard error will, in the long run, be more similar to the result just obtained than an *identical* RMSEA value with a larger standard error. Hence, our certainty about the true value of the RMSEA, and thus the true fit of the model to the data, is necessarily different.

An implication of the relative neglect of power in SEM is the potential for models to have acceptable RMSEA values, yet high levels of sampling variability—suggesting the possibility that the value a study's fit index obtained is merely due to chance. The issue of chance values of fit is an important one, as obtaining a RMSEA value that is deemed "adequate" in magnitude but not "adequate at beyond chance levels" in terms of its confidence intervals is much like a large correlation that is not sufficiently larger than 0 to be statistically significant. In our view, the current lack of attention to overall model fit sampling variability casts doubt on the fidelity of the results obtained in our literature for SEMs. Stated differently, owing to the neglect of power-related issues in SEM, it is possible that published research using SEMs does not have acceptable levels of power to differentiate failing from adequate models and, thus, do not have adequate power to make an accurate decision about model fit based on the data. Although it is possible that our SEMs in the organizational sciences do not have acceptable levels of power, and thus are not particularly informative about model fit, the extent of the problem is an empirical question. Therefore, in order to evaluate the possibility that SEMs in the organizational sciences do not offer adequate information about model fit, we conduct an extensive survey of the organizational science literature to ascertain the state of the field regarding power of SEMs in influential research from top-tier journals. Specifically, the current survey will consider many important aspects to the topic of power in SEM, such as the distribution of power across all studies included, differences in power between models that are ultimately deemed to be the best model in a study vs. those models that are not deemed to be the best model, differences in power across different types of model configurations (i.e., measurement vs. structural models), and we also incorporate information regarding sample size issues—as sample size is directly linked to statistical power. Finally, our study seeks to contribute to the discourse regarding good practice in SEM, and therefore will conclude by make recommendations related to "best practices" for power in SEM.

Variables to be Reviewed

We have reason to believe that there are several critical issues to examine vis-à-vis power in structural equation modeling. We list these factors, with emphasis on why we think these factors are important to examine in the quantitative review. These are the exploratory variables we will review, as we do not have a specific hypothesis associated with these variables.

The first and most important attribute that we wish to review is the distribution of power across models in Industrial-Organizational Psychology journals. Of particular interest in the percentage of studies that exhibit low levels of power, which in this case would correspond to inadequately falsifiable models. This is important because models that lack falsifiability have the potential to lead researchers down incorrect paths with respect model fit and the understanding of psychological relationships.

A second and related issue to be analyzed is the sample size associated with models to be tested in structural equation modeling. This is important because sample size is mostly controllable in model testing, and is directly related to a model's ability to be falsified with respect to overall model fit. It is our suspicion that in some cases, models possess a sample size that is grossly inadequate to meet a reasonable level of falsifiability and power. If this is true, it could signal the needs for researchers to increase the sample size involved in testing their models, in order to achieve power that signifies a reasonable level of falsifiability.

Moderators

We have reason to believe that several methodological artifacts will have an influence on the overall power of tested models in this quantitative review. We will list these variables and associated hypotheses now, with emphasis on why we think these artifacts will influence power.

We have reason to believe that the type of model researchers conducting structural equation modeling on will have an impact on the power of the tested model. Specifically, we believe that measurement models will possess higher levels of power as compared to models that are purely structural in nature. This is because measurement models tend to have greater degrees of freedom than structural models, owning primarily to the use of multiple indicators in modeling a single latent variable. It seems unlikely that on average, structural models will have enough variables to counteract the degrees of freedom obtained through the use of multiple indicators in measurement of latent variables. All else being equal, lower degrees of freedom for tested models equates to lower overall power. Therefore, our first hypothesis is:

> *Hypothesis 1: Power will be associated with model type such that measurement models will have significantly more power than structural models.*

We also have reason to believe that the model selection process in and of itself could lead to differing levels of power. Often, researchers will choose the best fitted model as their model of choice after adding pathways suggested by modification indices (MacCallum & Austin, 2000). Additionally, many researchers will judge models on the basis of overall model fit, and chose the best fitted model as their final model. If this is true, degrees of freedom in competing models will be lower than in the final model, which would in turn cause lower power, all else being equal. Additionally, if researchers are choosing models that have the best fit as their best models, it is possible that because power to detect misfit is negatively associated with fit (all else being equal), that main models will have best fit, but only because of lower power. Our hypothesis is therefore:

> *Hypothesis 2: Power will be associated with main models such that main models will have significantly less power than competing models.*

The previous two hypotheses dealt with methodological artifacts that could reduce power via reduction in the number of degrees of freedom of the tested model. We believe that power is also susceptible to being lower in models that deal with team and group topics than other types of models through sample size reduction. It is common in teams/groups studies to reduce the tested sample size in terms of N, because team/group variables usually required the original sample size to be divided by some factor in order to produce teams or groups to study. For example, a sample size of 600 *individuals* might be reduced to 200 *teams* composed of three individuals each. This process means that it is much harder for these studies to obtain sample sizes on which the final tested model will have an appreciably high N, than non-group/teams studies where the division of sample size does not take place, all else being equal. Because sample size is a factor that influences power, we hypothesize that:

> *Hypothesis 3: Power will be associated with teams/groups models such that teams/groups models will have significantly less power than non-teams/groups models.*

A third issue to be analyzed is the relationship between fit indices and power in models published in I/O Psychology journals. This is important because model fit index values are one of the main standards on which the value of a model is judged. To the degree that model fit is associated with power is the degree to which there is potential for fit indices to be artificially high and not reflective of true population values. Logically, because fit indices are all based on the degree of model misfit in some way, it follows that the degree to which the model is powered to detect misfit is the degree to which the model fit index will be worsened. It could be the case that for some models, fit is only seen as acceptable because the power to detect misfit is low. This review seeks to quantify these issues. We therefore hypothesize:

> *Hypothesis 4: Power will be associated with fit index quality such that as power increases, fit index quality decreases.*

Finally, we believe that journal quality may be related to overall power in that higher quality journals may be associated with more highly powered models. If higher powered models are more scientifically sound, and higher quality journals publish more scientifically sound researcher than lower quality journals, then there seems to be reason to believe that all else being equal, higher quality journals would present findings related to more falsifiable models. Our specific hypothesis is:

> *Hypothesis 5: Power will be associated with journal quality such that as journal quality increases, model power increases.*

Method: Sample of Studies

As the goal of the present work was to review trends of power in Organizational Psychology we conducted a comprehensive literature search of studies using SEM in journals frequently referenced in organizational psychology. Literature searches were conducted using the PSYCINFO, ProQuest, ERIC, AB-INFORM databases. Journals included were *Journal of Applied Psychology*, *Personnel Psychology*, *Academy of Management Journal*, *Human Performance*, *International Journal of Selection and Assessment*, *Journal of Management*, *Journal of Organizational Behavior*, *Organizational Behavior and Human Decision Processes*, and *Journal of Vocational Behavior.* As such, we only included studies using SEM from each of these 9 journals. We limited our search from 1996 to 2012, as 1996 is the year in which MacCallum et al.'s (1996) article was disseminated to the research community. In all databases, we used the keywords "covariance models," "confirmatory factor analysis," "structural equation modeling," and "SEM" to identify articles that used SEM.

Method: Selection Criteria

In order to be included in the present quantitative review, each study was required to report information needed to ascertain power of at least one model. Specifically, degrees of freedom and sample size for a structural equation model estimated was necessary. We also collected information on overall fit indexes such as the chi-square, RMSEA, CFI, and NFI. We identified a total of 365 studies across the 9 journals that met initial inclusion criteria. However, after reviewing each of the articles we excluded 25 for a grand total of 340 usable studies. In general, the articles that were excluded were measurement equivalence studies. We elected to exclude measurement equivalence studies due to the fact that these studies did not include information necessary for model comparison in that they rarely had a "main model" for coding. Additionally, measurement equivalence studies are generally less focused on absolute fit index values, and more focused on parameter equivalence between groups. Since the focus of the present study was on statistical power of overall model fit, we elected to discard measurement equivalence studies. Within the valid articles, 1,692 individual SEMs were included in the present study.

Method: Article Coding

Once we had identified a set of usable studies, we coded each article for relevant variables. First, we coded features of the articles such as the year, authorship, and journal where the article appeared.

Second, each SEM within each article was coded for its reported degrees of freedom (df) and sample size (n). Using information on df and n, we calculated the power coefficient for the tests of exact, close, and not-close fit, using software available from Preacher and Coffman (2006). In addition, we calculated the sample size that would be required to obtain a power coefficient of .80 and recorded the difference between this sample size and the actual, reported sample size.

Third, we recorded aspects of each SEM. The first coding task was to evaluate whether the model was a measurement only, structural only (i.e., with no estimated measurement-related parameters), or combination of measurement and structural model. The second coding task focused on whether the model in question was a model that evaluated phenomena about team and/or group functioning by dispersing the original sample size across teams or groups. This variable was included because we hypothesized that the necessary reduction in sample size to study variables at a team-level as opposed to individual level would decrease n and thus decrease power, all else being equal.

Fourth, we evaluated each reported SEM within an article to arrive at the "main" model(s) of the article. We considered the main models of the article to be models that were either most justified by theory presented earlier in the article and the focus of the study's hypotheses or the model that the study's authors deemed the "final" model either explicitly or indirectly in the language included in their discussion and results sections. The final model was always deemed to be the model that the authors declared their "final" model, even when the deemed "final" model differed from the model hypothesized in the introduction section of the article in question.

Finally, we included information about each of the overall fit indices of the SEMs. We coded for all possible fit indices, including a category of "other", for fit indices that are not traditionally reported in I/O Psychology, such as the Akaike Information Criterion (AIC).

Each of 1,692 structural equation models was double coded for accuracy. Descriptive statistics are presented in Table 1. The most problematic area with respect to coding was identification of main models, with an interrater agreement statistic of .76. Full interrater agreement statistics are displayed in Table 2. While no individual category possessed unacceptable, or even mediocre degrees of agreement, the difficulties in reliability centered almost exclusively on identification of the main model of the article. Specifically, there were instances in studies where the language used by authors made it unclear what the final model was meant to be. Often, these difficulties arose in studies where it appeared the authors engaged in post-hoc modeling while not explicitly stating they were doing so. In these cases, we carefully examined the introduction sections of the studies in order to ascertain the likelihood of authors truly supporting a particular post-hoc model vs. simply arriving at a post-hoc model through model modification.

A second area of importance with respect to judgment calls in coding came from identification of whether the model in question was a measurement model, structural model, or combination measurement/structural model. While this may seem surprising, we encountered studies that depicted misleading model figures (i.e., graphics) along with misleading text in indicating what model actually went into a particular statistical program. Often, this was indicated by degree of freedom figures that were dramatically misaligned with information presented in pictures and text. The most common feature of this idea was when authors only presented information in graphics relating to structural models, but in fact simultaneously tested a measurement/structural model, while neglecting to include this information in a footnote or text. In the instances where the degrees of freedom were dramatically misaligned with what would be a structural or measurement model in isolation and otherwise had no strong textual evidence to indicate what type of modeling was actually conducted, we elected to code the model as a combined measurement/structural model.

Table 1. Descriptive Statistics for Coded Variables

Variable	Mean	Standard Deviation
Eigenfactor Score	0.0125	0.0092
Degrees of Freedom	71.51	111.28
Sample Size	421.64	448.52
Model Power	0.79	0.29
Chi-Square	841.72	1927.14
RMSEA	.08	.06
CFI	.90	.10
NFI	.90	.10
Sample Size Required	428.77	124.57
Sample Size Difference	-12.95	167.86
Team/Groups Model vs. Other Model	1.92	2.78
Measurement vs. Structural Model	1.27	0.44
Main Model vs. Competing Model	1.66	0.94

Table 2. Interrater Agreement Statistics

Quantitative Review Category	Type of Agreement	Agreement Value
Sample Size	Intraclass Correlation Coefficient	0.98
Degrees of Freedom	Intraclass Correlation Coefficient	0.98
Statistical Power Coefficient	Intraclass Correlation Coefficient	0.98
Team Aggregation (yes/no)	Cohen's Kappa	0.98
Model Type (measurement, structural, combined)	Cohen's Kappa	0.85
Main Model (yes/no)	Cohen's Kappa	0.76

Method: Power Calculations

This quantitative review followed suggestions from MacCallum et al. (1996) by using values of 0.00 and 0.05 for testing exact fit, 0.05 and 0.08 for testing close fit, and 0.05 and 0.01 for testing not close fit; each value corresponding to ϵ_o and ϵ_a, respectively. However, because the general trend of power associated with the types of hypothesis tests were strongly correlated ($r = .99$), we elected to report the results of the close fit test, following recommendations from MacCallum et al. (1996).

Results: Distribution of Estimated Power

A primary goal of this quantitative review was to examine the *distribution* of estimated power coefficients in I/O Psychology. The distribution of estimated power coefficients can be seen in Table 3. Across all studies and models, approximately 22 % of models had a power coefficient less than .50. Therefore, 22% of all SEMs tested in Organizational Psychology have less than a 50 % chance of correctly rejecting an invalid model again where the null RMSEA of .05 and the alternative RMSEA of .08. Also, the majority of the power coefficients fell in the range above .90. Therefore, although most models have high levels of power, a non-trivial percentage of models have unacceptably liberal levels of power.

Table 3. Distribution of Power for Test-of-Close-Fit

Power	N	Proportion	Cumulative Proportion
0.00-0.09	21	0.01	0.01
0.10 – 0.19	88	0.05	0.06
0.20 – 0.29	67	0.04	0.10
0.30 – 0.39	124	0.07	0.17
0.40 – 0.49	75	0.05	0.22
0.50 – 0.59	85	0.05	0.27
0.60 – 0.69	47	0.03	0.30
0.70 – 0.79	66	0.04	0.34
0.80 – 0.89	104	0.06	0.40
0.90 – 1.00	1015	0.60	1.00

Results: Sample Size Issues

A goal of this quantitative review was to evaluate the degree to which models were judged to have too small a sample size, relative to the number of degrees of freedom of the model, in order to obtain a certain level of power. For this study, we used a power coefficient of .80 as the definition of a properly "powered" study. Interestingly, approximately 27% of models needed at least 100 more participants to reach a power coefficient of .80, while approximately 11% of models needed at least 500 more participants to reach a power coefficient of .80. In general, our results suggest that a nontrivial amount of models have sample sizes that are grossly inadequate to test their theoretical models. The results of the sample size analysis can be seen in Figure 1.

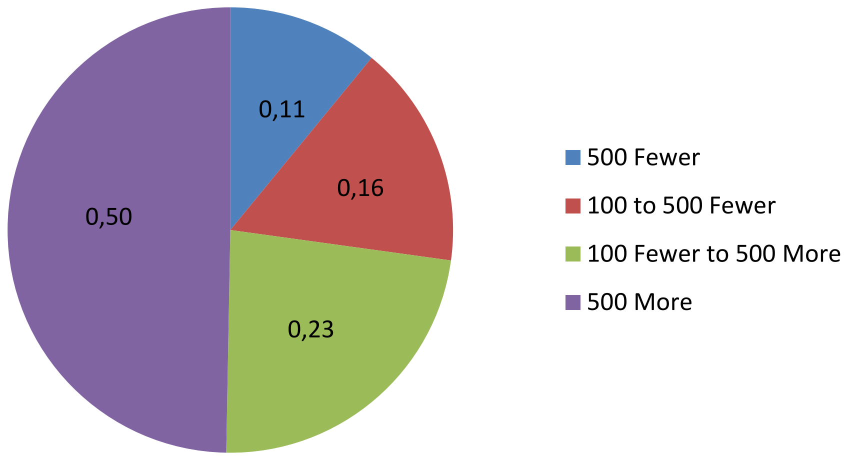


Fig. 1. Sample Size Required to Obtain a Power Coefficient of .80

Results: Fit Index Values

Because statistical power is likely to predict the fit of SEMs, we used obtained power coefficients as a predictor of model fit for the RMSEA, CFI, NNFI, and Chi-Square. Because we were concerned with SEMs in published articles, we limited our analysis to values of fit indices that were in the range of values likely to be published, and within the 95 percent confidence interval for models included in this review. Therefore, we examined the relationship between power and fit for values of RMSEA that were between 0.00 and 0.08, values of CFI between 0.90 and 1.00, and values of NFI between 0.90 and 1.00 (additionally, 95 percent of observations fell between these values for each fit index). We also examined Chi-Square via the probability value associated with the Chi-Square test statistic.

Results indicated that for all fit indices, statistical power was associated with worsened fit, as judged by the fit index in question. This finding held for RMSEA $r(784) = .25$, $p < .05$, CFI $r(889) = -.27$, $p < .05$, and NFI $r(278) = -.20$, $p < .05$. The correlation coefficient associated with the Chi-Square probability value was significantly associated with power, $r(1547) = .29$, $p < .05$, which indicates that as power increases, the likelihood of finding *nonsignificant misfit* decreases. These correlations were all statistically significant at the .05 level. In terms of interpretation, this means that within published studies and fit index values commonly seen in the literature, power is negatively related to fit index quality—liberally powered SEMs are more likely to obtain better fitted models as compared to conservatively powered SEMs, as judged by common fit indices.

Moderators: Differences in Model Power across Model Type

A goal of this quantitative review was to examine power across SEM types, specifically measurement vs. structural models. For measurement and structural models, the average estimated power coefficients for the test of close fit were .84 for measurement and .65 for structural models. The distributions of estimated power are indicated in Tables 4 and 5 for both model types. Across all studies, approximately 17% of measurement and 39% of structural models had a power coefficient less than .50 under the test of close fit. Stated differently, 16% of measurement and 39% of structural models tested have less than a 50% chance of correctly rejecting close fit. Further analysis indicated that the power of structural models was indeed significantly less than the power of measurement models, $t(1028) = 9.36$, $p < .05$, d = .64. The difference between measurement and structural models was most driven by the difference in degrees of freedom between measurement models and structural models, as the median value for degrees of freedom was 33 for structural models and 104 for measurement models. Thus, on average, structural models had a 19% *less* chance of correctly rejecting invalid models, compared to measurement models—owing to fewer degrees of freedom observed in studies focusing on evaluation of structural models. As such, structural models close to RMSEA of .08 in reality may appear—simply owing to chance—to be sufficiently close to a true RMSEA of .05 to accept the model as "good fitting".

Table 4. Distribution of Power for Test-of-Close-Fit for Measurement Models

Power	N	Proportion	Cumulative Proportion
0.00 – 0.09	00	0.00	0.00
0.10 – 0.19	33	0.04	0.04
0.20 – 0.29	30	0.04	0.08
0.30 – 0.39	33	0.04	0.12
0.40 – 0.49	30	0.04	0.16
0.50 – 0.59	31	0.04	0.20
0.60 – 0.69	16	0.02	0.22
0.70 – 0.79	30	0.04	0.26
0.80 – 0.89	42	0.06	0.32
0.90 – 0.99	511	0.68	1.00

Table 5. Distribution of Power for Test-of-Close-Fit for Structural Models

Power	N	Proportion	Cumulative Proportion
0.00 – 0.09	05	0.02	0.02
0.10 – 0.19	32	0.12	0.14
0.20 – 0.29	18	0.06	0.20
0.30 – 0.39	29	0.11	0.31
0.40 – 0.49	24	0.09	0.40
0.50 – 0.59	16	0.06	0.46
0.60 – 0.69	17	0.06	0.52
0.70 – 0.79	09	0.03	0.55
0.80 – 0.89	09	0.03	0.58
0.90 – 0.99	115	0.42	1.00

Moderators: Difference in Models Team/Groups vs. Others

An additional goal of this quantitative review was to determine the degree of difference in power between models that aggregated participants to team or group levels using composition models, and models that did not. It was our expectation that aggregation would lower the final sample size, thus lowering power compared to models that did not aggregate. The average difference in estimated power coefficients between team/groups models ($M = 0.65$, $SD = 0.33$) and other models ($M = 0.80$, $SD = 0.29$) was statistically significant, $t(1672) = 5.63$, $p < .05$, d = .48. The difference between aggregated and non-aggregated models was driven by the difference in sample size between team/groups models (median = 155) and other models (median = 288). Thus, on average, team/group models that aggregated responses at the team or group level had a 12% *less* chance of correctly rejecting invalid models, compared to models that did not aggregate—owing to smaller sample sizes observed in studies focusing on the team-level of aggregation. Therefore, these types of models tend to be *less falsifiable* than other types of models in organizational psychology, and may warrant special attention in both model construction and experimentation before the statistical testing of the model, as well as evaluation of the model after testing.

Moderators: Journal Quality

The final hypothesized variable to impact power is journal quality. We specifically hypothesized that journal quality would be positively associated with power such that as quality increased, power increased. Journal quality was indeed statistically significantly related to power, although the effect size was extremely modest $r(1420) = .06$, $p < .05$.

Discussion

The primary purposes of the present study were fourfold. First, we attempted to inform researchers how sampling variability and power can potentially harm our inferences regarding our judgments of SEMs. Second, we attempt to provide guidance as to how power is calculated in a structural equation modeling context. Third, we illuminated the main influencers of power so as to aid researchers in study design efforts. Fourth, we examined certain methodological situations that could signal a need for the researcher to pay special attention to power. We conducted a quantitative review of the literature to help address these issues. The summary of our findings confirms the need to explicitly consider power in structural equation modeling, as recommended by several methodologists (e.g., Kim, 2005; Kaplan, 1995; MacCallum, Browne, & Cai, 2006; MacCallum et al., 1996, MacCallum & Hong, 1997).

Summary of Findings

We find our results disconcerting in that nearly one-quarter of SEMs have less than a 50% chance of correctly invalidating a bad model. As such, we can conclude that it is likely that the results obtained from at least some of the models included that fall into the less than 50% power category have model fit that is suspect in nature, especially in cases where the model fit in the particular study was adequate (i.e., near . 08) and not excellent (i.e., <.05). Related to this point, we also discovered that a significant number of studies possessed sample sizes that were far removed from what the sample size ought to have been to have a more appropriately powered test of the SEM in question. These findings are particularly disillusioning as these models all appeared in influential journals for Industrial-Organizational Psychology. Therefore, we speculate that invalid models have likely been accepted into top journals, and consequently accepted by researchers in the scientific community. As was previously discussed, neglect of power can slow the advance of scientific progress by leading researchers toward theory that departs from reality as low model fit power reduces the likelihood of rejecting incorrect models. Owing to our findings, these errors are possible as essentially none of the sampled studies conducted a power analysis on their SEM–in spite of the availability of web-based, power analysis tools from Preacher and Coffman (2006).

Finally, the results from this quantitative review pinpoint particular situations where overly liberal statistical power is more likely to occur in research. Specifically, research that involves the evaluation of structural models or models that evaluate team or group level phenomena are more likely to be susceptible with respect to overly liberal statistical power. For structural models, this occurs because of reduced amounts of degrees of freedom, as compared to measurement models. For team/groups models, this occurs because of a reduced sample size, as compared to most individual-level models. Additionally, we found that models that were seen as the correct model for the study in question had significantly more liberal levels of power than competing models. It is our intention that bringing light to these situations, and their implications for science, will alert researchers and practioners when they need to pay particular attention to statistical power.

Deriving from our discussion of SEM power as well as our quantitative review of current practice in organizational psychology, in the coming sections we outline what we believe are important recommendations for researchers as a whole, and introduce issues and recommendations related to power that will most likely require the combined attention and consideration from researchers, consumers, and editorial gatekeepers to better advance scientific advancement in organizational psychology via improved research methodology.

User Recommendations

A desired end in the present study is to prompt researchers to conduct a-priori statistical power analysis for SEMs. Given the availability and user-friendly nature of statistical power tools (see Preacher & Coffman, 2006); we believe very little stands in the way of researchers conducting SEM overall fit index power analysis. As we suggest throughout the present work, understanding the level of power of a SEM is instrumental in the interpretation of overall model fit. To that end, we attempt to present an ordered list of recommendations for the common user of SEMs in research.

First, the researcher should construct a theoretical model of interest and study design. Ideally, the model should maximize the tradeoffs amongst explanatory power, parsimony (Williams & Holahan, 1994), potential for true model fit, and analysis of model misfit. These types of tradeoffs can most likely be estimated by reviewing similar styled models in the particular research domain of interest.

Second, the researcher should determine the sample size they are likely to acquire in testing the theoretical model of interest under their current study design. As with all study designs, it is important to make allowances for methodological artifacts that will decrease sample size over the course of a study.

Third, the user should determine the other elements of the hypothesis test. For this quantitative review, we focused our review around what MacCallum and his associates (1996) dubbed the "test of close fit", whereby the null hypothesis was a RMSEA value of .05 and the alternative RMSEA value was a value of .08. However, it is important to note that any values can be used for the null and alternative RMSEA, even values outside the three tests discussed in the MacCallum et al. (1996) paper. Similarly, any value of alpha can be used, theoretically. It might not always be the case that the researcher is interested in testing null and alternative RMSEA values in line with the tests described by MacCallum et al. (1996). Readers interested in this line of thought can consult work on *isopower* by MacCallum, Lee, & Browne (2010).

With all of these elements, the researcher should conduct a power analysis *before* starting the study. Provided the researcher knows the model degrees of freedom, sample size, alpha, and null/alternative hypothesis tests, power can be calculated for any of the aforementioned power tests using tools provided by Preacher and Coffman (2006) or direct R syntax (available from the first author upon request). These tools require no programming knowledge and can be done via graphical interface (i.e. – "point and click" or "copy and paste"). The user will then have an obtained power coefficient. At this point, some recommendations are warranted, depending on the level of the power coefficient.

After a-priori power analysis has been conducted, the first recommendation is that the user should immediately revise their research plan if the power coefficient is extremely low. In such situations researchers are particularly susceptible to *accepting invalid models* as a result of the lack of falsifiability of the model and lack of ability to establish the *verisimilitude* (i.e. – likelihood of truth) of the theoretical model in a meaningful way. This is particularly important in a field that is wed to the use of approximate fit indices and rule of thumb interpretations of such fit indices that contain a relative lack of nuance and appreciation for how dependent approximate fit indices are to statistical artifacts that are usually not even explored, let alone reported (Marsh, Hau, & Wen, 2004; Nye & Drasgow, 2010; Williams & O'Boyle, 2011). This recommendation might beg the question of what power coefficients are considered unacceptably low. Since we do not wish to establish a mechanistic "rule of thumb" regarding this topic, we would simply encourage researchers to "think continuously" as phrased by Cortina & Landis (2011) instead of "thinking discretely" with respect to power coefficients. In the case of extremely low powered models, researchers can remedy low power by obtaining a larger sample size or finding some way to gain degrees of freedom (removing added paths or adding variables of interest to the model are two examples), or both. We suspect in most cases it would be more methodologically sound to increase sample size than degrees of freedom for reasons we will explain later. If a researcher is interested in how large a sample size they need to reach a certain level of power, they can refer to the aforementioned tool from Preacher and Coffman (2006) for guidance.

In instances where models have sample sizes large enough to generate high power (i.e. – near 1) in two or more models, cross-validation recommendations are warranted. For example, a model with 30 degrees of freedom and a sample size of 2,000 would generate a power coefficient of 1 for a single model, but could also generate power coefficients of 1 for an original model and cross-validation model (or even near 1 if two cross-validated models were tested), if the sample size was split into two equal groups. While, a larger sample size is always better from a statistical standpoint all else being equal, there are issues of diminishing returns with respect to sample size and power to detect misfit. Consequently, there are instances where the benefits derived from cross-validation vastly exceed the very minor benefits made to statistical power. In these instances we recommend cross-validation.

However, not all situations will allow for cross-validation by splitting the original sample into multiple groups. One situation that does not allow for cross-validation in such a way is when splitting sample size into two groups would compromise power, reducing power from one high powered model to two modestly powered models. For example, a model with 17 degrees of freedom and a sample size of 500 would have a power coefficient of .80. If that sample size was split into two equal groups, the power coefficient for those groups would reduce to .50, which means that likelihood of rejecting an invalid model would decrease by approximately 38 percent from the original model, or framed another way, going from incorrect acceptance of bad models (by chance) one out of every five times to one out of every two times, assuming the models are bad in the population.

The value of cross-validation depends in part on the falsifiability of the original and cross-validated model. There will be cases where the benefits normally associated with cross-validation do not keep pace with the problems associated with decreasing levels of falsifiability of the tested models. When this is the case, we recommend testing a single model with the collected data and not engaging in any cross-validation efforts with the *original* data (collecting new data to sufficiently cross-validate the model is always welcome however). As is the case with defining "extremely low power", we do not wish to establish rote, mechanistic rules of thumb for when to cross-validate with the original sample. We would simply encourage researchers to conduct a-priori power analysis on what power would be both before and after cross-validation. To the degree power stays *the same* and is *high*; cross-validation should *occur*. To the degree power *drops* and is *low*; cross-validation should be *avoided*. If this situation occurs and new data cannot be obtained to cross-validate the model, the researcher should report the expected cross-validation index (ECVI; Browne & Cudeck, 1989; Browne & Cudeck, 1992), which is computed as an index of how well a solution obtained in one sample is likely to fit independent samples.

Another time when cross-validation should be avoided is when doing so would compromise the stability of parameter estimates. Theoretically, there are cases when splitting a sample size into two or more groups could drive down the sample size enough to cause parameter estimates to go from sufficiently stable to unacceptably unstable. In order to obtain stable parameters estimates, the researcher should aim for a ratio of five units for every free parameter (Bentler & Chou, 1987). Our recommendation is similar to the previous, in that we would encourage researchers to conduct a-priori analysis on the parameter stability of the model parameters before and after cross-validation. To the degree stability stays the *same* and is *high*; cross-validation should *occur* and to the degree stability *drops* and is *low*; cross-validation should be *avoided*.

A final over-arching recommendation is to avoid testing models through "two-step" processes (Anderson & Gerbing, 1988; Anderson & Gerbing, 1992) and having *final* evaluations regarding the utility of models determined through separately analyzing measurement and structural components. The main reason is that in at least some cases, separating one single model evaluation into two smaller models (measurement and structural), has the potential to severely compromise the power to detect model misfit in what would otherwise be a more strongly powered model via reduction in the degrees of freedom of the tested model in much the same way that cross-validation has the potential to compromise power via sample size reduction.

When it comes to ultimate declarations about the utility of a model, we recommend testing the entire model (measurement + structural) in a single step in order to determine overall model fit, as well as individual parameter values. However, one valid criticism of this one-step method proposed by some methodologists (Fornell & Yi, 1992a; Fornell & Yi, 1992b; Hayduk 1996) is that it does not easily lend itself to investigation of model problems and misfit (Bollen, 2000). To that end, we advise researchers to consider the *jigsaw piecewise technique* advocated by Bollen (2000) in combination with the one-step method. Under this technique, researchers fit pieces of the overall model together and then as a whole, ideally evaluating all possible subcomponents of the overall model to assess where and when model misfit experiences radical upward shifts.

In models that combined measurement and structural elements, it is particularly important to test all aspects of the model, as a model's overall fit if judged in a single step, could be disproportionately influenced by the measurement model (Mulaik et al., 1989). In these cases, poor structural fit could be masked by excellent measurement model fit, resulting in an overall model fit statistic that is deemed as acceptable, although the lack of structural fit would still render the model unsound to the researcher. These problems could be addressed from the aforementioned jigsaw piecewise technique approach specified earlier, examination of parameter values, examination of the residual matrix for the overall model, and calculation and examination of fit indices which focus more on path model relationships (such as the RMSEA-P; Williams & O'Boyle, 2011). A more ambitious approach to this issue could involve the *combinatorics* approach taken by Meehl and Waller (2002), whereby within a path analysis framework, path analytic model parameters are estimated using only a subset of the elements of the sample correlation matrix, and the resulting parameter estimates are then tested by determining how well they account for the other unused, elements of the correlation matrix. This procedure is conducted for the original model as well as for a set of similar alternative models, and the original model is then compared with the alternatives with respect to results of the risky tests. Support for *verisimilitude* of the original path analytic model is enhanced to the degree that it outperforms the alternative models. If a researcher finds their original structural model is outperformed by another model, this can provide clues about the validity of the proposed structural model.

Ultimately, it is important to recognize that there is no single procedure, and certainly no single mechanical ritualistic procedure that will address all possible methodological issues within structural equation modeling at once, and that a variety of procedures are needed to test the utility of an SEM in a rigorous manner.

Limitations

The current study has a number of limitations. First, the approach taken by MacCallum et al. (1996) in conceptualizing type I and type II errors can arguably be seen as backwards, given that these conceptualizations run contrary to traditional understandings of type I and type II errors in Psychology for more traditional statistics such as ANOVA. However, we feel that despite the potential confusion over “reversing” the usual terms, there is still a great deal of utility in the overall approach. In discussing these issues with colleagues, we have found it useful to discuss the issues presented in this paper in terms of “liberal” and “conservative” levels of power, rather than in the language of type I and type II errors. Moreover, the general issues that we have discussed and illuminated in this study can also be studied analyzed with type I and type II errors that are aligned in more traditional ways (Hancock 2006; Hancock & Freeman, 2001) if desired.

Another potential limitation of this study is that it necessarily relies on the RMSEA fit index, as well as specific cutoffs for what constitute close fitting models. There are two issues to consider here. The first issue revolves around the use of specific cutoffs for close fitting models under RMSEA. The second issue revolves around the use of approximate model fit indices in SEM in evaluating model fit. We will consider both of these points in turn.

Strict cutoff values for model fit in SEM are oversimplifications of complex statistical situations. We do not see much value in the rules of thumb often used to evaluate model fit, as they tend to be dependent on statistical issues that are often ignored in practice (Marsh, Hau, & Wen, 2004; Nye & Drasgow, 2010; Williams & O’Boyle, 2010). Like Marsh et al. (2004), we feel that despite the cautions offered by Hu and Bentler (1999), problems with the rules of thumb surrounding model fit have been frequently ignored in the practice of SEM. In fact, a current citation count of Hu & Bentler’s study reports approximately 617 overall citations *per yea*r. For the sake of comparison, Marsh et al. (2004), which signifies the very serious problems associated with strict adherence to rules of thumb, averages 59 citations per year, with a large number of these citations occurring in methodological journals (Harzing, 2010).

The reason we chose to use the rules of thumb outlined in the introduction is because we assumed that these would be the values most salient to the greatest number of researchers, a conjecture which has been vetted by the above evidence. Consequently, we thought that our findings would resonate with researchers more when using rules of thumb to which most researchers most likely adhered in order to see the interplay between power and focusing solely on passing the .08 RMSEA “goodness” threshold.

Similar reasoning was used with respect to the second issue: the use of approximate model fit indices to evaluate model fit in the first place. A lively discussion on the internet listserv SEMNET has developed regarding the usefulness of evaluating model fit using goodness of fit (GOF) indices. Some researchers argue that the Chi-Square test is the only acceptable fit index to use in testing SEMs (Barrett, 2007; Hayduk et al., 2007). The argument for the sole use of the Chi-Square in evaluating models is centered on the following points: 1) there are no single thresholds for GOF indices that can be applied to any fit index under all possible measurement and data conditions, 2) GOF indices often allow for researchers to avoid careful model specification and examination, 3) GOF indices can allow rather weak models to make it through the peer-review process, 4) the potential for the degree of casual misspecification of models to be uncorrelated with GOF indices values, and 5) Chi-Square does a better job at detecting model misspecification than does any other fit index.

Readers who are interested in examining the issue of approximate versus exact fit indexes at length can examine the above issues and counterarguments in the special 2007 issue of the *Personality and Individual Differences* journal (42nd volume 5th edition) that summarizes this debate. While we recognize the importance of the fit index debate, the RMSEA fit index is still commonly used and evaluated by reviewers as a basis for the adequacy of model-to-data fit. Indeed, virtually no model in our quantitative review was evaluated strictly on the use of Chi-Square test. It is also worth noting, that even in the event that a researcher only wants to use the Chi-Square to evaluate models, the power analysis of MacCallum et al. (1996) can be extended in such a way, by aligning the null hypothesis value of RMSEA to zero, which represents perfect fit and is analogous to the Chi-Square test (McIntosh, 2007).

Future Research

There are multiple areas for future research as it relates to this study. First, it could be useful for power analysis of SEM to be extended to other areas of psychology. It is possible that the problems presented in this quantitative review are more severe in other areas of psychology that frequently employ SEM. Second, this line of research could potentially extend to comparisons between nested SEMs. Recent work on statistical power as it relates to nested SEM has been proposed by MacCallum and Browne (2006) and Li & Bentler (2011). Third, this line of research could potentially extend to analysis of power about specific parameter estimates. Specifically, using the statistical program Mplus, and work introduced by Muthen & Muthen (2002), one can calculate the power to detect that a specific parameter will be different than zero.

Conclusions

The present study provides strong evidence for the importance of statistical power in SEM in organizational psychology research. Statistical power has been emphasized in experimental and correlational research. We believe the emphasis on statistical power should extend to SEM vis-à-vis overall model fit, with stricter control of model quality in journals via the editors, simple tools that can be used by researchers to actually calculate power coefficients for SEMs, and better quality education with respect to the teaching of power in modeling to peers and students.

References

[1] MacCallum, R.C., & Austin, J.T. "*Applications of Structural Equation Modeling in Psychological Research*", Annual Review of Psychology, 51. pp. 201-226. 2000.

[2] Hermida, R., Conjar, E.A., Najab, J.A., Kaplan, S.A., & Cortina, J.M. *On the Practice of Allowing Correlated Residuals in Structural Equation Models*. Unpublished Manuscript, Department of Psychology, George Mason University, Fairfax, Virginia, United States. 2010.

[3] Landis, R., Edwards, B. D., & Cortina, J. *Correlated residuals among items in the estimation of measurement models*. In C. E. Lance & R. J. Vandenberg (Eds.). Statistical and methodological myths and urban legends: Doctrine, verity, and fable in the organizational and social sciences (pp. 195-214). New York: Routledge. 2009.

[4] Le, H., Schmidt, F., Harter, J.K., & Lauver, K.J. "*The problem of empirical redundancy of constructs in organizational research: An empirical investigation*", Organizational Research Methods, 112. pp. 112-125. 2010.

[5] Pearl, J. "*Causal inference in statistics: An overview*", Statistics Survey, 3. pp. 96-146. 2009.

[6] Pearl, J. "*The causal mediation formula: A guide to the assessment of pathways and mechanisms*", Prevention Science, 13. pp. 426-436. 2012.

[7] Pearl, J. "*Interpretation and Identification of Causal Mediation*", Psychological Methods, 19. pp. 459-481. 2014.

[8] Bagozzi, R.P., & Yi, Y. "*Specification, evaluation, and interpretation of structural equation models*", Journal of the Academy of Marketing Science, 40. pp. 8-34. 2012.

[9] Bentler, P.M. & Bonnett, D.G. "*Significance tests and goodness of fit in the analysis of covariance structures*", Psychological Bulletin, 88. pp. 588-606. 1980.

[10] Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

[11] Cohen, J. "*A Power Primer*", Psychological Bulletin, 112. pp. 155-159. 1992.

[12] MacCallum, R.C., Browne, M.W., & Sugawara, H.M. "*Power analysis and determination of sample size for covariance structure modeling*", Psychological Methods, 1, pp. 130-149. 1996.

[13] Kaplan, D. (1995). *Statistical power in structural equation modeling*. In R.H. Hoyle (ed.), Structural Equation Modeling: Concepts, Issues, and Applications (pp. 100-117). Newbury Park, CA: Sage Publications, Inc.

[14] Specht, D. A. "*On the evaluation of causal models*", Social Science Research, 4. pp. 113-133. 1975.

[15] Mulaik, S. A., James, L. R., Van Alstine, J., Bennett, N., Lind, S., & Stilwell, C. D. "*Evaluation of goodness-of-fit indices for structural equation models*", Psychological Bulletin, 105. pp. 430-445. 1989.

[16] Barrett, P. "*Structural equation modeling: Adjudging model fit*", Personality and Individual Differences, 42. pp. 815-824. 2007.

[17] Hayduk, L., Cummings, G., Boadu, K., Pazderka-Robinson, H., & Boulianne, S. "*Testing! testing! One, two, three-Testing the theory in structural equation models!*", Personality and Individual Differences, 42. pp. 841–850. 2007.

[18] Bagozzi, R.P., & Yi, Y. "*On the evaluation of structural equation models*", Journal of the Academy of Marketing Science, 16. pp. 74-94. 1998.

[19] Bentler, P.M. "*Comparative fit indexes in structural models*", Psychological Methods, 107. pp. 238-246. 1990.

[20] McDonald, R.P., & Ho, M.-H.R. "*Principles and Practice in Reporting Statistical Equation Analysis*", Psychological Methods, 7. pp 64-82. 2002.

[21] Steiger, J.H. "*Structural model evaluation and modification: An interval estimation approach*", Multivariate Behavioral Research, 25. pp. 173-180. 1990.

[22] Steiger J. H., & Lind J. M. (1980). *Statistically based tests for the number of factors*. Paper presented at the annual meeting of the Psychometric Society, Iowa City, IA.

[23] Sugawara, H. M. & MacCallum, R. C. "*Effect of estimation method on incremental fit indexes for covariance structure models*", Applied Psychological Measurement, 17. pp. 365-77. 1993.

[24] Rigdon, E.E. "*Calculating degrees of freedom for a structural equation model*", Structural Equation Modeling, 1. pp. 274-278. 1994.

[25] Patnaik, P.B. "*The noncentral* χ^2 *and F-distributions and their applications*", Biometrika, 36. pp. 202-232. 1949.

[26] Cox, D.R., & Reid, N. "*Approximations to noncentral distributions*", Canadian Journal of Statistics, 15. pp. 105-114. 1987.

[27] Preacher, K. J., & Coffman, D. L. (2006). Computing power and minimum sample size for RMSEA [Computer software]. Available from http://quantpsy.org/.

[28] McQuitty, S. "*Statistical power and structural equation models in business research*", Journal of Business Research, 57. pp. 175-183. 2004.

[29] Chen, F., Curran, P.J., Bollen, K.A., Kirby, J., and Paxton, P. "*An empirical evaluation of the use of fixed cutoff points in RMSEA test statistic in structural equation models*", Sociological Methods and Research, 36. pp. 462-494. 2008.

[30] Kim, K. H. "*The relation among fit indexes, power, and sample size in structural equation modeling*", Structural Equation Modeling, 12. pp. 368-390. 2005.

[31] MacCallum, R.C., Browne, M.W., & Cai, L. "*Testing differences between nested covariance structure models: Power analysis and null hypotheses*", Psychological Methods, 11. pp. 19-35. 2006.

[32] MacCallum, R.C., & Hong, S. "*Power analysis in covariance structure modeling using GFI and AGFI*", Multivariate Behavioral Research, 3. pp. 193-210. 1997.

[33] Williams, L. J. & Holahan, P. J. "*Parsimony-based fit indices for multiple-indicator models: do they work?*", Structural Equation Modeling, 1. pp. 161–189. 1994.

[34] MacCallum, R., Lee, T., & Browne, M.W. "*The issue of isopower in power analysis for tests of structural equation models*", Structural Equation Modeling, 17. pp. 23-41. 2010.

[35] Marsh, H. W., Hau, K. T., & Wen, Z. L. "*In search of golden rules: Comment on hypothesis testing approaches to setting cutoff values for fit indexes and dangers in overgeneralising Hu & Bentler (1999) findings*", Structural Equation Modeling, 11. pp. 320-341. 2004.

[36] Nye, C. D., & Drasgow, F. "*Assessing goodness of fit: Simple rules of thumb simply do not work*", Organizational Research Methods, 14. pp. 548-570. 2011.

[37] Williams, L.J., & O'Boyle Jr., E.H. "*The myth of global fit indices and alternatives for assessing latent relations*", Organizational Research Methods, 14. pp. 350-369. 2011.

[38] Cortina, J.M. & Landis, R.S. "*The earth is NOT round (p = .00)*", Organizational Research Methods, 14. pp. 332-349. 2011.

[39] Browne, M.W., & Cudeck, R. "*Single sample cross-validation indices for covariance structures*", Multivariate Behavioral Research, 24. pp. 445-455. 1989.

[40] Browne, M.W., & Cudeck, R. "*Alternative ways of assessing model fit*", Sociological Methods and Research, 21. pp. 230-258. 1992.

[41] Bentler, P. M., & Chou, C. "*Practical issues in structural modeling*", Sociological Methods & Research, 16. pp. 78-117. 1987.

[42] Anderson, J. C., & Gerbing, D. W. "*Structural equation modeling in practice: A review and recommended two-step approach*", Psychological Bulletin, 103. pp. 411-423. 1988.

[43] Anderson, J. C., & Gerbing, D. W. "*Assumptions and comparative strengths of the two-step approach: Comment on Fornell and Yi*", Sociological Methods and Research, 20. pp. 321-333. 1992.

[44] Fornell, C., & Yi, Y. "*Assumptions of the two-step approach to latent variable modeling*", Sociological Methods and Research, 20. pp. 291–320. 1992.

[45] Fornell, C., & Yi, Y. "*Assumptions of the two-step approach to latent variable modeling: Reply to Anderson and Gerbing*", Sociological Methods and Research, 20. pp. 334-339. 1992.

[46] Hayduk, L. (1996). *LISREL: Issues, debates, and strategies*. Baltimore: Johns Hopkins University Press.

[47] Bollen, K. A. "*Modeling strategies: In search of the Holy Grail*", Structural Equation Modeling, 7. pp. 74-81. 2000.

[48] Meehl, P.E., & Waller, N.G. "*The path analysis controversy: A new statistical approach to Strong Appraisal of Verisimilitude*", Psychological Methods, 7. pp. 283-300. 2002.

[49] Hancock, G. R. (2006). *Power analysis in covariance structure models*. In G. R. Hancock & R. O. Mueller (Eds.), Structural Equation Modeling: A Second Course. Greenwood, CT: Information Age Publishing, Inc.

[50] Hancock, G.R., & Freeman, M.J. "*Power and sample size for the RMSEA test of not close fit in structural equation modeling*", Educational and Psychological Measurement, 61. pp. 741-758. 2001.

[51] Hu, L., & Bentler, P. M. "*Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives*", Structural Equation Modeling, 6. pp. 1-55. 1999.

[52] Harzing, A.W. "Publish or Perish", available at www.harzing.com/pop.htm. 2010.

[53] McIntosh, C. N. "*Rethinking fit assessment in structural equation modelling: A commentary and elaboration on Barrett (2007)*", Personality and Individual Differences, 42. pp. 859–867. 2007.

[54] Li, L., & Bentler, P.M. "*Quantified choice of root-mean-square errors of approximation for evaluation and power analysis of small differences between structural equation models*", Psychological Methods, 16. pp. 116-126. 2011.

[55] Muthen, L.K., & Muthen, B.O. "*How to use a Monte Carlo study to decide on sample size and determine power*", Structural Equation Modeling, 9. pp. 599-620. 2002.

Modelling loans and deposits during electoral years in Romania

Nicolae-Marius JULA*
University Lecturer, PhD - Nicolae Titulescu University, Bucharest, Romania

Abstract

This paper analyzes the effect of electoral years on loans and deposits for population in Romania. Using monthly data regarding the total loans and deposits, we identify the significance of the electoral timing on population´s behavior regarding financial decisions. We estimate that there are small changes in population´s affinity for increase in the indebtedness or for savings.

We use dummy variables for electoral periods, and when these are econometrically significant there is an evidence of the influence of the electoral timings in population´s financial decisions.

Keywords: *electoral years, loans, deposits, econometric model, ARDL model*

JEL Classification: *C32, C52, G10*

Introduction

There are studies that suggest there are different approaches of the banking system across Europe and the rest of the world when electoral years are coming. These studies stress that in well-established democracies such as the in the US or UK, governments have usually only a regulatory role in the banking sector, while in many other countries governments directly control financial resources through ownership of one or more banks in addition to their regulatory functions. In the states where the government highly controls the banking system, there are suspicions of corruption and state-owned banks can be misused by the ruling party, who may direct money to projects which will benefit those who support the government rather than those who serve the greater public interest.

In most democracies, the banking system is independent and usually the government cannot direct money through banks in order to increase the chances of re-election. Nevertheless, we analyze how the banking system and the population react in electoral years.

There are studies analyzing the situation across the world for state owned banks, like Sapienza[1] (2004) – in Italy - provides evidence that state-owned banks charge lower interest rates than do private sector banks, Khwaja and Mian[2] (2005) – Pakistan - provide evidence that low-quality borrowers with political connections can borrow from state-owned banks, Baum et al.[3] (2008) find that politically affiliated banks in Ukraine have significantly lower interest rate margins, Chinese state-owned banks are less profitable, less efficient and have worse asset quality than other types of banks (Lin and Zhang[4] (2009), Berger et al.[5] (2009)).

In Romania, because of the independence of National Bank of Romania, there are not clear evidences of this type of behavior. We focus on pure electorate preferences, meaning the level of confidence in the stability of the economic situation.

* Corresponding author: mariusjula@univnt.ro

[1] Sapienza, P., 2004. The effects of government ownership on bank lending. Journal of Financial, Economics 72 (2), 357–384.

[2] Khwaja, A. I., Mian, A., 2005. Do lenders favor politically connected firms? Rent provision in an emerging financial market. The Quarterly Journal of Economics 120 (4), 1371–1411

[3] Baum, C. F., Caglayan, M., Sch ̈afer, D., Talavera, O., 2008. Political patronage in Ukrainian

[4] Lin, X., Zhang, Y., 2009. Bank ownership reform and bank performance in China. Journal of Banking & Finance 33 (1), 20–29.

[5] Berger, A. N., Hasan, I., Zhou, M., 2009. Bank ownership and efficiency in China: What will happen in the world's largest nation? Journal of Banking & Finance 33 (1), 113–130.

Modelling Loans and Deposits during Electoral Years in Romania

Let variable *dvl_total* be the total deposits in lei and *crl_total* total credits in lei. We assume that in presidential electoral years like 2004, 2009 and 2014 there is a change in these values.

We define 3 dummy variables for these electoral years (dummy04, dummy09 and dummy12), where the values are 1 in electoral years (on monthly level) and 0 otherwise.

As data sources, we use the values from National Bank of Romania, http://www.bnro.ro/Raport-statistic-606.aspx. The modelling was completed using E-Views, version 9.0.

For deposits, the assumption is tested using the following econometric model:

Table 1. Total deposits in lei – author's calculations

Dependent Variable: DVL_TOTAL/IPC
Method: Least Squares
Date: 06/22/15 Time: 19:20
Sample (adjusted): 2002M10 2014M12
Included observations: 147 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	760.4141	312.7101	2.431690	0.0163
DVL_TOTAL(-1)/IPC(-1)	0.984609	0.012514	78.68318	0.0000
RD	-29.58915	15.63153	-1.892915	0.0604
DUMMY09	-658.8270	259.7429	-2.536458	0.0123
DUMMY12	-445.6710	242.9918	-1.834099	0.0687
DUMMY14	-29.49539	254.0484	-0.116101	0.9077

R-squared	0.989246	Mean dependent var	15871.55
Adjusted R-squared	0.988864	S.D. dependent var	7274.270
S.E. of regression	767.6285	Akaike info criterion	16.16445
Sum squared resid	83084754	Schwarz criterion	16.28651
Log likelihood	-1182.087	Hannan-Quinn criter.	16.21404
F-statistic	2593.962	Durbin-Watson stat	1.935352
Prob(F-statistic)	0.000000		

Where IPC represents the Consumer Price Index and RD is the referential interest rate.

The estimators for electoral years 2009 and 2004 are econometrically significant, while the one for 2014 is not.

Moving to credits, the obtained results are:

Table 2. Total credits in lei – author's calculations

Dependent Variable: CRL_TOTAL/IPC
Method: Least Squares
Date: 06/22/15 Time: 19:12
Sample (adjusted): 2002M10 2014M12
Included observations: 147 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	948.1708	264.3648	3.586600	0.0005
CRL_TOTAL(-1)/IPC(-1)	0.987082	0.005216	189.2375	0.0000
RD	-21.27032	12.53150	-1.697349	0.0918
DUMMY09	-483.0364	196.8037	-2.454407	0.0153
DUMMY12	-357.2469	177.5202	-2.012429	0.0461
DUMMY14	-132.4890	185.3929	-0.714639	0.4760

R-squared	0.998464	Mean dependent var	32703.52

Adjusted R-squared	0.998410	S.D. dependent var	14040.17
S.E. of regression	559.9157	Akaike info criterion	15.53341
Sum squared resid	44204281	Schwarz criterion	15.65547
Log likelihood	-1135.706	Hannan-Quinn criter.	15.58300
F-statistic	18332.20	Durbin-Watson stat	0.729741
Prob(F-statistic)	0.000000		

The estimators for electoral years 2009 and 2004 are econometrically significant, while the one for 2014 is not. Also, the model does not pass the autocorrelation test (DW test value is low).

We can conclude that both for deposits and credits, electoral years 2004 and 2009 showed an influence, meaning that there was a decrease in total deposits in these 2 years. For year 2014, the results were not econometrically supported.

We used also another set of dummy variables, where we select only 2 months prior the electoral moment and 2 after.

For deposits, the results are:

Table 3. Total deposits in lei – author's calculations

Dependent Variable: DVL_TOTAL/IPC
Method: Least Squares
Date: 06/22/15 Time: 19:06
Sample (adjusted): 2002M10 2014M12
Included observations: 147 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	872.2507	291.5686	2.991580	0.0033
DVL_TOTAL(-1)/IPC(-1)	0.972677	0.011622	83.69328	0.0000
RD	-32.04046	14.41367	-2.222921	0.0278
DUMMY09M	-273.6475	361.5812	-0.756808	0.4504
DUMMY12M	-277.0588	357.6630	-0.774637	0.4399
DUMMY14M	712.6067	366.0472	1.946762	0.0536

R-squared	0.988951	Mean dependent var	15871.55
Adjusted R-squared	0.988559	S.D. dependent var	7274.270
S.E. of regression	778.0748	Akaike info criterion	16.19148
Sum squared resid	85361445	Schwarz criterion	16.31354
Log likelihood	-1184.074	Hannan-Quinn criter.	16.24108
F-statistic	2524.026	Durbin-Watson stat	1.911667
Prob(F-statistic)	0.000000		

We can observe that only for electoral year 2014 the estimator is econometrically significant.

For total credits, the obtained results are:

Table 4. Total credits in lei – author's calculations

Dependent Variable: CRL_TOTAL/IPC
Method: Least Squares
Date: 06/22/15 Time: 19:22
Sample (adjusted): 2002M10 2014M12
Included observations: 147 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1134.273	242.1076	4.684996	0.0000
CRL_TOTAL(-1)/IPC(-1)	0.981672	0.004757	206.3747	0.0000
RD	-28.56219	11.39351	-2.506882	0.0133

DUMMY09M	-207.5735	269.3308	-0.770701	0.4422
DUMMY12M	-283.7317	264.1095	-1.074296	0.2845
DUMMY14M	-194.9298	269.9848	-0.722003	0.4715

R-squared	0.998387	Mean dependent var	32703.52
Adjusted R-squared	0.998330	S.D. dependent var	14040.17
S.E. of regression	573.7795	Akaike info criterion	15.58233
Sum squared resid	46420431	Schwarz criterion	15.70439
Log likelihood	-1139.301	Hannan-Quinn criter.	15.63192
F-statistic	17455.66	Durbin-Watson stat	0.675040
Prob(F-statistic)	0.000000		

Here none of the estimators for our periods of interest are econometrically significant.

The second analysis, based on dummy variable created to capture the precise months near the electoral event, cannot be econometrically sustained (only for deposits, in electoral year 2014).

As we mention in a previous article[6], analyzing the evolution of credits and deposits suggests a model with autoregressive and lag-distributed factors (an ARDL model). If the series are I(0) – stationary, we can use basic OLS for estimation. If we know the order of integration for the series, and it is the same for all, but they are not cointegrated, we estimate each series independently. If the series are integrated of the same order and are cointegrated, the theory suggest that we estimate, according to Dave Giles[7] "(i) An OLS regression model using the levels of the data. This will provide the long-run equilibrating relationship between the variables. (ii) An error-correction model (ECM), estimated by OLS. This model will represent the short-run dynamics of the relationship between the variables."

For credits, using Eviews software, we obtaine the following results:

Table 5. Total credits in lei – ARDL model - author's calculations

Dependent Variable: CRL_TOTAL/IPC
Method: ARDL
Date: 06/01/15 Time: 23:49
Sample (adjusted): 2003M04 2014M12
Included observations: 141 after adjustments
Maximum dependent lags: 12 (Automatic selection)
Model selection method: Akaike info criterion (AIC)
Dynamic regressors (12 lags, automatic): RD
Fixed regressors: C @TREND
Number of models evalulated: 156
Selected Model: ARDL(2, 7)
Note: final equation sample is larger than selection sample

Variable	Coefficient	Std. Error	t-Statistic	Prob.*
CRL_TOTAL(-1)/IPC(-1)	1.461233	0.076212	19.17316	0.0000
CRL_TOTAL(-2)/IPC(-2)	-0.470285	0.075415	-6.235946	0.0000
RD	-4.739891	40.90288	-0.115882	0.9079
RD(-1)	80.46257	63.79553	1.261257	0.2095
RD(-2)	-194.5636	64.93482	-2.996290	0.0033
RD(-3)	141.0182	65.98484	2.137130	0.0345
RD(-4)	-56.46841	65.48296	-0.862338	0.3901
RD(-5)	-56.88114	65.67987	-0.866036	0.3881
RD(-6)	123.2913	64.77300	1.903436	0.0592
RD(-7)	-101.5646	40.55301	-2.504490	0.0135
C	1766.352	322.8551	5.471035	0.0000
@TREND	-8.803979	2.156076	-4.083335	0.0001

[6] Jula N.M., 2015, Software solutions for ARDL models, CKS 2015, 1001-1006
[7] Giles D., 2013, ARDL Models, http://davegiles.blogspot.com.es/2013/03/ardl-models-part-i.html

R-squared	0.999198	Mean dependent var	33817.94
Adjusted R-squared	0.999130	S.D. dependent var	13225.93
S.E. of regression	390.1402	Akaike info criterion	14.85215
Sum squared resid	19635010	Schwarz criterion	15.10311
Log likelihood	-1035.077	Hannan-Quinn criter.	14.95414
F-statistic	14614.98	Durbin-Watson stat	2.016080
Prob(F-statistic)	0.000000		

*Note: p-values and any subsequent tests do not account for model selection.

This translates in a model like:

$$CRL\_TOTAL/IPC_t = \beta_0 + \beta_1 CRL\_TOTAL(-1)/IPC(-1) + \beta_2 CRL\_TOTAL(-2)/IPC(-2) + \alpha_0 RD_t + \alpha_1 RD_{t-1} + \alpha_2 RD_{t-2} + ... + \alpha_7 RD_{t-7} + \varepsilon_t$$

For deposits, we obtain an ARDL(7,0) model:

Table 6. Total deposits in lei – ARDL model - author's calculations

Dependent Variable: DVL_TOTAL/IPC
Method: ARDL
Date: 06/22/15 Time: 19:50
Sample (adjusted): 2003M04 2014M12
Included observations: 141 after adjustments
Maximum dependent lags: 12 (Automatic selection)
Model selection method: Akaike info criterion (AIC)
Dynamic regressors (12 lags, automatic): RD
Fixed regressors: C @TREND
Number of models evalulated: 156
Selected Model: ARDL(7, 0)
Note: final equation sample is larger than selection sample

Variable	Coefficient	Std. Error	t-Statistic	Prob.*
DVL_TOTAL(-1)/IPC(-1)	1.008354	0.085511	11.79208	0.0000
DVL_TOTAL(-2)/IPC(-2)	-0.041918	0.120099	-0.349027	0.7276
DVL_TOTAL(-3)/IPC(-3)	0.104406	0.119382	0.874557	0.3834
DVL_TOTAL(-4)/IPC(-4)	-0.096606	0.118184	-0.817418	0.4152
DVL_TOTAL(-5)/IPC(-5)	0.005022	0.117421	0.042766	0.9660
DVL_TOTAL(-6)/IPC(-6)	0.411911	0.117306	3.511413	0.0006
DVL_TOTAL(-7)/IPC(-7)	-0.416436	0.082093	-5.072705	0.0000
RD	-50.09462	21.26648	-2.355566	0.0200
C	1120.154	410.7344	2.727199	0.0073
@TREND	-2.396577	3.497914	-0.685145	0.4945

R-squared	0.989945	Mean dependent var	16397.15
Adjusted R-squared	0.989254	S.D. dependent var	6954.140
S.E. of regression	720.8813	Akaike info criterion	16.06711
Sum squared resid	68076759	Schwarz criterion	16.27624
Log likelihood	-1122.731	Hannan-Quinn criter.	16.15209
F-statistic	1433.032	Durbin-Watson stat	1.842568
Prob(F-statistic)	0.000000		

*Note: p-values and any subsequent tests do not account for model selection.

A problem when using complex models to overcome the econometric tests is that it is difficult to keep a reasonable economic interpretation for the created variables. As Mayumi and Gianpietro stated in *Dimensions and logarithmic function in economics: A short critical analysis* (2010), the analysts should "know the importance of "dimensional homogeneity" in daily life which is an arithmetic principle: 4 m^2 plus 4m^3does not make any sense; one dollar plus one dollar makes perfect sense, but one dollar times one dollar does not make any sense at all. So, economists concerned with the biophysical and monetary aspects of ecological and economic interactions must understand the importance of "dimensional homogeneity".

Conclusions

There are articles that suggest in some countries where the central bank is not independent or the state has enough banks under its control, in electoral years can be observed an increase in credits, especially for political supports. In Romania, the results indicate that in 2004 and 2009 the credits and the loans decreased. Less credits may be interpreted as an insecure period for both the creditors and debtors. Less number of deposits suggests that the population may have opted for other saving options.

For further development of this analysis, we should compare also the results for credits and loans in foreign currency. Even if the National Central Bank encouraged population to avoid credits in foreign currency, there is a lot of demand for these types of credit lines.

Also, for deposits and loans we suggest using ARDL models, as these time series variables are most suitable for these models and using OLS can raise a succession of hypothesis testing issues.

"This work was financially supported through the project "Routes of academic excellence in doctoral and post-doctoral research - READ" co-financed through the European Social Fund, by Sectoral Operational Programme Human Resources Development 2007-2013, contract no POSDRU/159/1.5/S/137926."

References

[1] Baum, C. F., Caglayan, M., Sch ̈afer, D., Talavera, O., 2008. Political patronage in Ukrainian.
[2] Berger, A. N., Hasan, I., Zhou, M., 2009. Bank ownership and efficiency in China: What will happen in the world's largest nation? Journal of Banking & Finance 33 (1), 113–130.
[3] banking. The Economics of Transition 16 (3), 537–557.
[4] Fidrmuc J. (2000), Political support for reforms: Economics of voting in transition countries, European Economic Review 44, 1491-1513.
[5] Frances, P. H. & R. van Oest, 2004. On the econometrics of the Koyck model. Report 2004-07, Econometric Institute, Erasmus University, Rotterdam.
[6] Frey B. (1996). Political Business Cycles. Aldershot: Edgar Elgar.
[7] Gaertner M. (1994). Democracy, Elections and Macroeconomic Policy; Two Decades of Progress, European Journal of Political Economy, 10: 85-110.
[8] Giles D., 2013, ARDL Models, http://davegiles.blogspot.com.es/2013/03/ardl-models-part-i.html.
[9] Hallenberg M., & Souza L.-V. (2000). The Political Business Cycles in EU Accession Countries, Tinbergen Institute Discussion Paper, TI-2000-085/2.
[10] Jula N.M., 2015, Software solutions for ARDL models, CKS 2015, 1001-1006.
[11] Jula N., Jula D., 2015, The Impact Of Compensation Payments On Employment In Regional Structures, in CKS2015, 675-681.
[12] Khwaja, A. I., Mian, A., 2005. Do lenders favor politically connected firms? Rent provision in an emerging financial market. The Quarterly Journal of Economics 120 (4), 1371–1411.
[13] Koyck, L. M., 1954. Distributed Lags and Investment Analysis. North-Holland, Amsterdam.
[14] Lemann L., Bochsler D. (2014), A systematic approach to study electoral fraud, Electoral Studies, 35, 33-47.
[15] Lin, X., Zhang, Y., 2009. Bank ownership reform and bank performance in China. Journal of Banking & Finance 33 (1), 20–29.
[16] Mayumi K., Gianpietro M., (2010), Dimensions and logarithmic function in economics: A short critical analysis, in Ecological Economics, 69, 1604–1609.
[17] Nannestad P., & Paldam M. (1999). The Cost of Ruling. A Foundation Stone for Two Theories, Working Paper 19/5, Department of Economics, Aarhus University, Denmark.
[18] Pacek A.C., (1994), Macroeconomic Conditions and Electoral Politics in East Central Europe, in American Journal of Political Science 38, 723-44.
[19] Paldam M. (1997), Political Business Cycles, cap. 16 in Mueller (ed.), Perspectives in Public Choice. A Handbook, Cambridge UP: Cambridge, UK & NY.
[20] Persson T., & Tabellini G., 1990, Macroeconomic Policy, Credibility, and Politics, Chur, Switzerland: Harwood Academic Publishers.
[21] Sapienza, P., 2004. The effects of government ownership on bank lending. Journal of Financial, Economics 72 (2), 357–384.

Is Africa's current growth reducing inequality? Evidence from some selected african countries

Alege P.O.
George E.O.
Ojeaga P.I.*
Oluwatimiro Q.
Covenant University Nigeria P.M.B 1023 Ota
Department of Economics and Development Studies

Abstract

Is Africa's current growth reducing inequality? What are the implications of growth on output performances in Africa? Does the effect of Africa's growth on sectorial output have any implication for inequality in Africa? The study investigates the effect of shocks on a set of macroeconomic variables on inequality (measured by life expectancy) and the implication of this on sectors that are perceived to provide economic empowerment in form of employment for people living in the African countries in our sample. Studies already find that growth in many African countries has not been accompanied with significant improvement in employment. Therefore inequality is subject to a counter cyclical trend in production levels when export destination countries experience a recession. The study also provides insight on the effect of growth on sectorial output for three major sectors in the African economy with the intent of analyzing the impact of growth on sectorial development. The method used in this study is Panel Vector Autoregressive (PVAR) estimation and the obvious advantage of this method lies in the fact that it allows us to capture both static and dynamic interdependencies and to treat the links across units in an unrestricted fashion. Data is obtained from World Bank (WDI) Statistics for the period 1985 to 2012 (28 years) for 10 African Countries. Our main findings confirm strong negative relationship between GDP growth and life expectancy and also for GDP and the services and manufacturing sector considering the full sample.

Keywords: *Growth, Sectorial Performances, Inequality, Panel VAR and Africa*

JEL Classification: *C33, E30, F62*

1. Introduction

In this section a brief introduction into growth and inequality is presented. According to the World Bank press release October 2013, Africa's economic growth outlook continues to remain strong with an estimated forecast of 4.9% growth rate for 2013, it is expected that the African economy will grow by 6% in 2014, depicting that Africa will continue to experience strong economic growth in the years to come. The African region is also expected to remain a strong magnet for tourism and investment due to the attractiveness of the African business environment despite problems of high political instability, business environment risks and poor economic policies. Strong government investments, higher production in the mineral, agricultural and services sectors are also boosting growth in many African countries. Private investment and regional remittances are also on the increase, with remittances alone now worth over 33 billion dollars supporting household income. It is clear that almost a third of the countries in the region are now experiencing growth rates of over 6% making African countries to be among the fastest growing economies in the world. This increasing growth trends however, have also been found not to translate to poverty reduction in many African countries. Inequality and poverty has remained quite high despite strong growth and the rate of poverty reduction has remained quite sluggish, with Africa still accounting for the highest proportion of un-enrolled school children in the World. Africa's Pulse (2014) a World Bank yearly Journal, also states that despite the global economic improvement in Africa, poverty will continue to remain a strong concern on the continent. Forecasting that between 16 to 33% of the entire World's poor, will reside in Africa by 2030 presenting once again a future demographic challenge that can be an impediment to future development of the continent. The vulnerability of economic growth in Africa to capital flow and commodity price reduction also makes it imperative for many African countries

* Corresponding author drpaulojeaga@gmail.com

to invest in times of growth in other non-performing sectors with prospects for cushioning their economies from global shocks (i.e. shocks associated with a sudden reduction in commodity prices and capital flows to the continent).

This paper investigates the effects of growth on inequality in Africa by studying the implication of growth for sectors in the African economy that are labor intensive particularly the agricultural and services sector with meaningful use for economic empowerment and inequality reduction. It also investigates the effect of growth on the manufacturing sector that is less labor intensive with the intent understanding the impact of growth on the manufacturing sector. Incite is gained on the implication of growth on inequality reduction in general using panel vector auto regression (PVAR) which allows us to study dynamic interdependencies between growth and inequality reduction with the intent of establishing a link between growth and inequality the impact of growth on sectorial output particularly for sectors with capability for employment are also considered. Data for some ten selected African countries (they include Algeria, Egypt, Ghana, Nigeria, Kenya, Uganda, Cameroon and Congo) two from each major economic regions (i.e. North, West, East, Central, and Southern Africa) in Africa is utilized in the study for the period of 1985 to 2012 a period of 28 years. The rest of the paper is divided into the scope and objective, review of literature, empirical analysis and results and the concluding sections.

2. Scope and Objectives of the Study

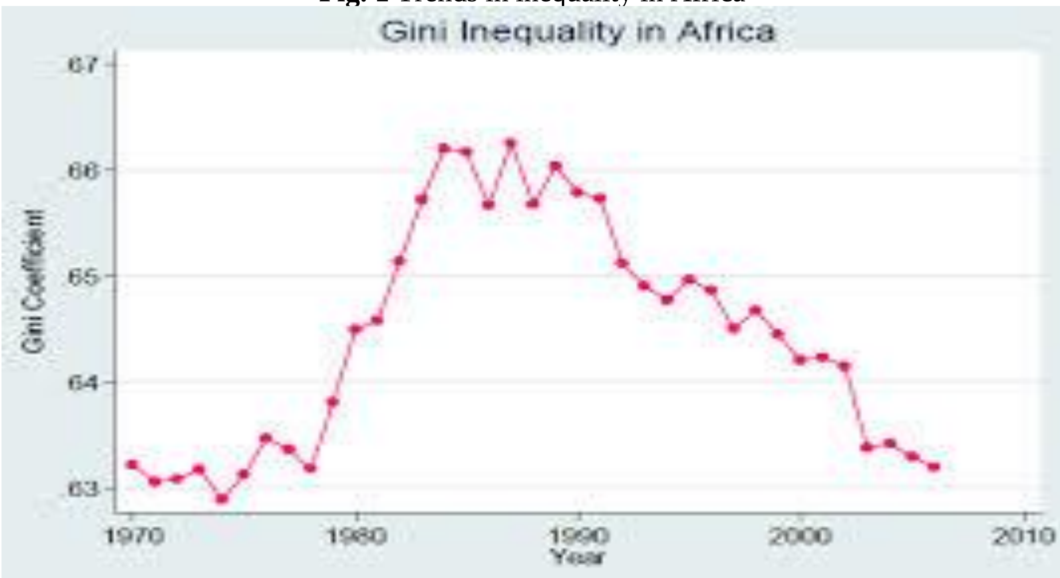
In this section the scope and objectives of the study are presented. The study investigates the implications of the current growth trend in Africa on inequality reduction, by studying the implications of growth on life expectancy (as the measure for inequality) and output production in three sectors namely agricultural, services and the manufacturing sectors, the first two being labour intensive and the last a technological driven sector termed high-tech under the assumption increased productivity in sectors will mean higher levels of economic empowerment through employment. The objectives of the study are:

a. To determine the effect of Africa's current growth on inequality reduction
b. To evaluate the effect of Africa's growth on sectorial performances in three major sectors in the African economy i.e. agricultural, services and manufacturing sectors.
c. And to investigate if the impact of Africa's current growth has an effect on labor intensive sectors with the capability to reduce inequality.

3. Stylized Facts on Growth and Inequality

In this section stylized facts on growth and inequality is presented. Poor macroeconomic management, political instability, corruption and a host of other factors are responsible for poor results in infrastructural development in many developing countries. Poverty also contributes to poor production output in many African countries since poor incomes means less access to quality educational and health facilities for a sizeable percentage of their population resulting in poor skill development.

Fig. 1 Trends in inequality in Africa

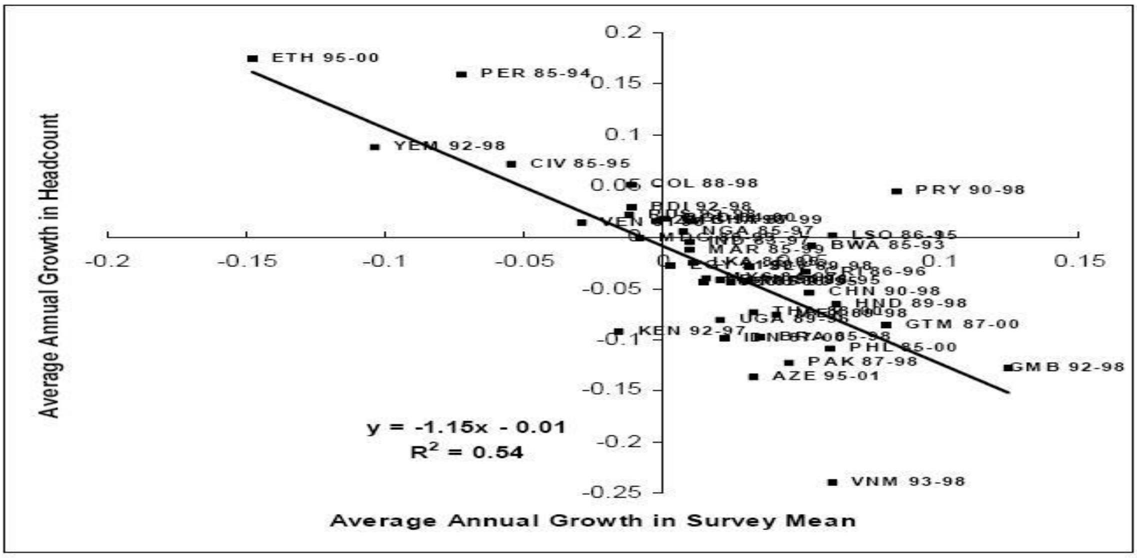


Note: The graph above depicts that inequality is on the decrease this is particularly noticeable from the early 1990s. However Africa still has the World's highest percentage of people living below the poverty line.

Source: World Bank Gini-coefficient on inequality in Africa

Trends already show that there is still a wide gap between the rich and the poor in Africa; although inequality is reducing (See Trends in inequality in Fig. 1), Africa still has the highest per cent of the world's people living below the poverty line. This depicts the sluggishness of government policies in yielding results that can have meaningful effect for job creation and skill improvement. The paper by Art Kraay (2004) after studying the implication of growth for household income in some selected developing countries also state that at best, there is a negative relationship between annual average growth and annual growth in household in many developing countries depicting that national growth does not often translate to household growth in developing countries thereby suggesting a counter-cyclical relationship between growth and annual household growth. This shows that growth in many developing countries are often not inclusive, and are associated with joblessness therefore such economic expansions are not characterized

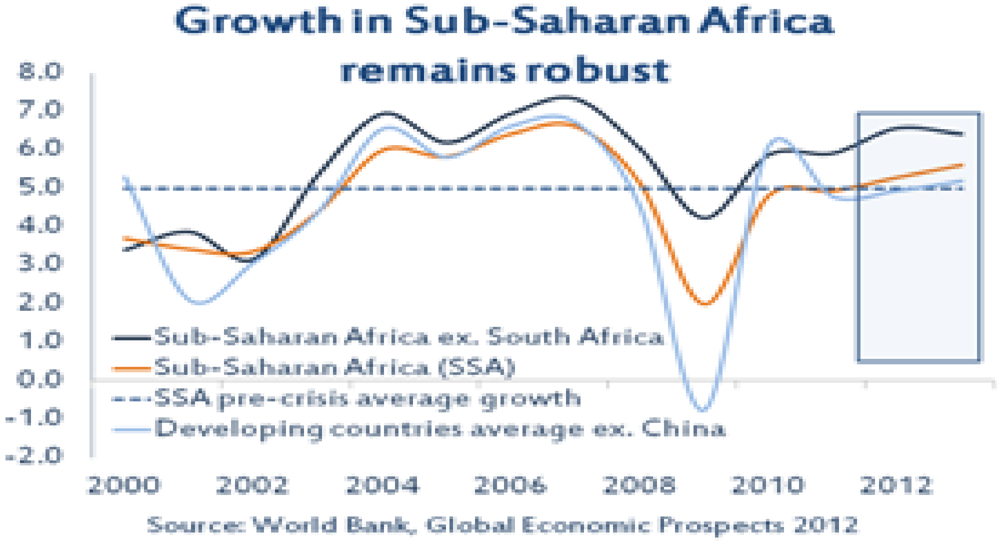
Fig.2 Relationship between growth rate and poverty reduction



Note: The fig above depicts that national growth does not often translate to household growth in many African countries suggesting a counter-cyclical relationship between growth and annual household growth.

Source: The paper by Art Kraay (2004) "When Growth is Pro-Poor" with improved skill development and training of indigenous manpower in many developing countries. Trends also show that growth rate in sub Saharan Africa remains modestly high particularly in the last decade according to the World Bank statistics 2013 (see fig. 3), Africa's growth rate has been on the average at approximately 6% annually. Pundits also state that though the growth trend has lasted for close to a decade many African countries have failed to take advantage of the current trend to diversify their economics from simple raw material exporting economies to middle level manufacturing economies. Despite the less developed nature of the African banking system compared to those of the developed North, the 2008 financial meltdown had strong effects on the economies of many African countries. Depicting an interdependent relationship, between Africa and the rest of the World.

Fig. 3 Growth in Sub- Saharan Africa in the last decade

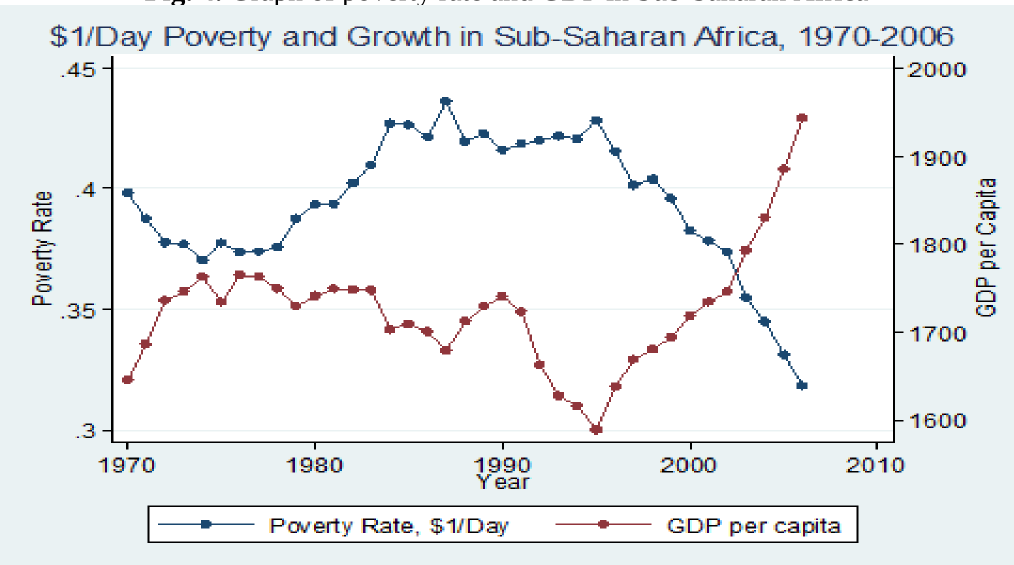


Note: The figure above shows recent trends in growth for the last decade in Africa. Africa's growth rate has been on the average at approximately 6% annually. This growth is often associated with sustained commodity price boom. It also shows the interdependency between the developed North and developing South particularly the effect of the sub-prime mortgage crisis of 2008 on growth in Africa.

Growth in the last decade has also managed to surpass that of the 1980s (see Fig. 4), this increasing trend in growth has not translated to improved earnings for the mass of the population in many African countries. The long run implication of such growth for Africa is that productive capabilities are going to be limited in the future as natural resources dwindle. It also means that in the short run many African countries are going to continue to trade in primary goods (raw material exports), obviously missing out from gains often associated with product differentiation that skill and development of a robust domestic industrial base can provide.

The quality of manpower is also a source of concern since many African economies are plagued with poor incomes and poor educational and other socio infrastructural amenities such as power, roads and housing. The overall implication for poor income increases on the continent is that it will have strong effects on domestic consumption and access by the poor to social amenities that could otherwise have long run implications on the economy. For instance strong domestic consumption could mean an insulation from global financial shocks as in the case of China and access to social amenities could mean people could live longer and transfer savings to their offspring.

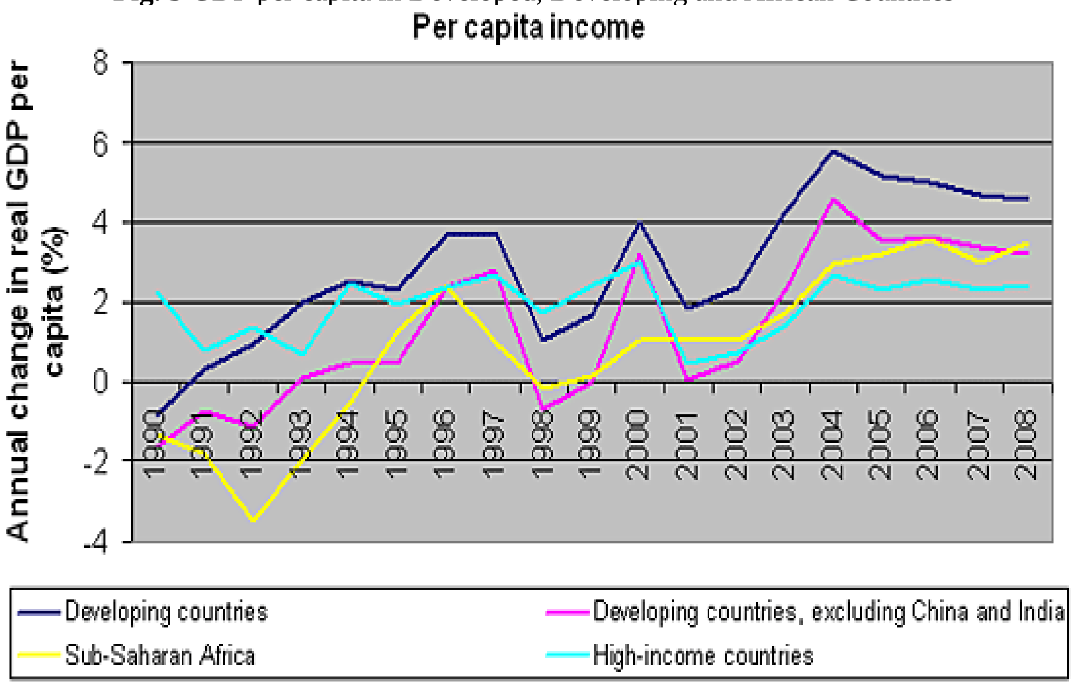
Fig. 4. Graph of poverty rate and GDP in Sub-Saharan Africa



Note: The above graph depicts the counter-cyclical nature of Poverty on GDP in Africa. This depicts that improved growth showing in the last decade i.e. year 2000 onwards has not translated to strong reduction in the income gap between the rich and poor in many African countries.
Source: World Bank data

Trends also suggest that there is a relationship in movements in GDP across the different regions of the world see Fig. 5. Financial shocks are also seen to be transmitted from high income countries to other regions in subsequent periods. This once again depicts the interdependent nature of the global economy. Capital flows in times of economic shock can have strong effects for many developing countries receiving foreign aid, so also can it affect foreign direct investment to the private sector of the economies of many developing countries since investors are often known to repatriate funds back to their domestic economies in times of crisis. The implication of such negative capital flows is that a reduction will in turn affect the volume of their expenditure spending. This reduces their capability of providing social infrastructure and other social amenities which have the tendency to reduce the gap between the rich and poor.

Fig. 5 GDP per capita in Developed, Developing and African Countries



Note: The figure above compares GDP in high income countries to those of other developing countries and Africa. Financial shocks are seen to be transmitted from high income countries to other regions in subsequent periods. This once again depicts the interdependent nature of the global economy.
Source: World Bank Statistics 2010

4. Review of Literature

In this section we review some past literature on the subject under study. Past studies have utilized panel regressions in studying, the effect of capital flows particularly in times of banking crisis on aid supply and find that aid supply decreases after the first two periods in times of crisis Laeven and Valencia (2010). Historical evidences of the effect of capital flows have found that recessions are likely to have deep and prolonged effects on growth and fiscal balance and cause significant disturbances to government revenue and expenditure Reinhart and Roghoff (2008).

The variables used in the study are based on instrumental variables for poverty (life expectancy, Misery index etc) and aid Hansen and Tarp (2001), Rajan Subramanian (2008) and Ojeaga (2012). VAR models are also employed, the models employed in the paper by Frot (2009) is extended for the purpose of the study. The variables employed in the analysis in the study have been found to have significant relevance for life expectancy in one or more past literature. They include GDP per capita, fiscal variables such as government expenditure spending and sectorial output from manufacturing, services and agriculture see Chong and Granstein (2008), Faini (2006), Boschini and Olofsgard (2007) Dang et al (2009) and Frot (2009).

The aim of the study is to investigate the extent to which capital flows affect inequality in Africa using data from some selected countries and compare the response of life expectancy and output productivity particularly in three sectors that can influence economic empowerment which include the agricultural, manufacturing and services sectors to unexpected shocks. It is worthy of note to state also that other studies have presented counter argument rejecting the possible relationship between decreases in capital flows and economic recessions stating that capital flows does not depend solely on economic factors, arguing that political factors and strategic decisions about where to invest were more relevant see Paxton and Knack (2008) for such a critical position. However, in this study, it was found that economic factors influence capital flows and significantly affects fiscal spending which in turn have grave consequences for inequality.

5. Empirical Analysis and Results

5.1. Theoretical Framework

In this subsection the theoretical framework of the study is presented. We assume that there are i=1,…..t sectors in the economy of countries, contributing to their aggregate output production of which export is a useful fraction and that exports from countries will flow to different export destination countries j= 1,……j. Private sector production is also not purely to promote welfare and production of satisfactory public goods in countries but mainly for the private interest (profits for the private entrepreneur) which is the returns on invested revenue i.e. firms profit maximization ends. Therefore firms in countries are indirect consumers of production. The framework portrays aggregate production therefore as a private rather than a public good. Large firms will therefore produce fewer goods per total revenue than small firms, since large numbers of shareholders will mean more profits shared, rather reinvested in the production process. The aim of this paper is to investigate, the effect of the choice of producers to consume indirect production or maximize welfare. Therefore we can let the producers in sectors have the utility function expressed in equation 1 as

$$u_{i,t} = f\,(P_{ij,t}C_{i,t}) \tag{1}$$

Where $P_{ij,t}$ is aggregate production in a country across sectors j in firm i at time t, and $C_{i,j}$ is the total consumption in firm i at time t. Individual preferences for firm goods can be written as expressed by Chong and Gradstein (2008) as

$$U=U_A\ (P_{ij,t}) + U_c\ C_{i,t} = \frac{1}{1-\sigma}\,\alpha\, P_{ij,t}{}^{1-\sigma} + \frac{1}{1-\sigma}\, C_{i,t}{}^{1-\sigma}\ ,\ \alpha > 0 \tag{2}$$

The parameter α is the preferences for goods produced and σ is the elasticity of substitution between two goods. Income is also allocated between consumption, government expenditure and firms, therefore if price is numeraire, then firms budget constraint will be

$$Y_{i,t} = C_{i,t}\ P_{ij,t}{}^{\alpha}\ D_{i,t}{}^{\beta} \quad where \quad \beta > 0 \tag{3}$$

Where β represents preferences for internal expenditures and $D_{i,t}$ represents cost of transaction between firms and markets (external expenditure). Revenues $R_{i.t}$ come from sales $S_{i.t}$ and from bank credit $B_{i.t}$ that are used to finance production and other firm internal costs, expressed as

(4) $$R_{i.t} = S_{i.t} + B_{i.t}$$

(5) $$R_{i.t} = \alpha P_{ij,t} + \beta D_{i,t} = T_{i,t}$$

Therefore firm output will follow production targets across sectors representing aggregate output in countries which will be subject to external shocks and deviation, where the adjustment to such shocks will take longer than one period. The production of goods by firms i in countries will be subject to available resources for production and other internal cost incurred by firms in their day to day production. Such cost related to unstable economic conditions will affect production levels $P_{ij,t-S}$ and could also be associated with other long run impacts expressed as the lagged variables of the internal cost of firms $D_{i,t-s}$. Allowing us to state production below as

(6) $$P_{iJ,t} = P_{iJ,t-S} \sum_{\substack{i=1 \\ T=1}}^{I,T} D_{i,t} \sum_{\substack{i=1 \\ T=1}}^{I,T} D_{i,t-S} \ (Y_{i,t})(\varepsilon_{i,t})$$

With $\varepsilon_{i,t}$ representing, other country or time specific shocks, and s indicating the number of lagged periods. The impact of crisis shocks will be function firms internal needs, financial condition of consuming countries and exporting countries, social conditions in producing countries and other political preferences.

(7) $$D_{i,t} = f(d_{i,t}\ ;\ f_{i,t};\ p_{i,t})$$

Production will be an increasing function of good financial conditions, political concerns, social conditions and available resources and decreasing function of social needs since firms having their own profit maximizing interests expressed below as

(8) $$\frac{\partial D}{\partial d} < 0\ , \qquad \frac{\partial D}{\partial p} > 0\ , \qquad \frac{\partial D}{\partial f} > 0, \qquad \frac{\partial D}{\partial Y} > 0$$

5.2. Empirical Analysis

In this subsection the intuition for the study is presented. The analysis is based on VAR, it adequately stems from the fact that it studies interdependencies among variables without worrying about the direction of causality. It is flexible and the method treats all variables in the system as endogenous and independent, each variable is explained by its own lagged values and those of the other variables.

It is also a system of equations and not a one equation model. Panel VAR also allows for the investigation of unobservable individual heterogeneity and improve asymptotic results. The results provide useful insights which go beyond coefficients to reveal the adjustment and resilience of unexpected production shocks as well as the importance of other different shocks. Canova and Ciccarelli (2004), give a brief description of the PVAR analysis expressing the general form as

(9) $$y_{i,t} = P_0 p_{i,t} + L_1 y_{i,t-1} + \ldots\ldots + L_p y_{i,t-p} + u_t$$

where $y_{i,t}$ is a kx1 vector of k panel data variables, and i= 1,……, I, $p_{i,t}$ is a vector of deterministic terms such as the linear trend, dummy variables or a constant, P_0 is the associated parameter matrix and the L's are k x k parameter matrices attached to the lagged variables $y_{i,t}$. The lag order is represented by p, the error process is represented by three components, $\mu_{i,t}$ the country specific effect, γ_t the yearly effect and $\varepsilon_{i,t}$ the disturbance term. Two restrictions are imposed by the specification: a.) It assumes common slope coefficients, and it does not allow for interdependencies across units. Therefore the estimates L are interpreted as average dynamics in the response to shocks. All variables depend on the past of all variables in the system as with the basic VAR model with the individual country specific terms been the difference.

This study tries to establish that movements in growth have an intrinsic effect on inequality (life expectancy to be specific) and production across sectors particularly those that have the capability of employing a sizeable amount of the population in sub-Saharan Africa. The study applies panel data analysis to past production volumes. There are other studies that have studied the effect of aid in times of crisis e.g. Gravier-Rymaszewska (2012), Hansen and Heady (2010) also study the effect of aid on net imports and spending using PVAR.

The study uses PVAR approach to estimate the effect on inequality and sectorial production output of unexpected shocks to a set of variables that are responsive to economic upturns. The method is suitable since the VAR method does not require the imposition of strong structural relationship and another merit is that only a minimal set of assumptions are needed to interpret the impact of shocks on each variable. The reduced form equation allows for the implementation of dynamic simulations once the unknown parameters are estimated. However the method only allows for the analysis of short run adjustments effects and not the long run structural effects.

The results come in form of the impulse response functions (IRFs) and their coefficients analysis as well as their forecast error variance decompositions (FEVDs) which allows for the examination of technological innovations or shocks to any variable in question to other variables. Orthogonalizing the response allows us to identify the effect of one shock at a time while others are held constant. The Choleski decomposition method of variance covariance matrix of residuals is adopted, the identification is based on the premise that variables which appear earlier in the system are more exogenous than those which appear later and are assumed endogenous. Implying that the variables that follow are affected by the earlier variables contemporaneously with lags and the later variables affect previous variables only with lags.

The simple VAR model is presented below with three variables: GDP per capita, government spending (govspend) and sectorial output (Sec.Out/GDP) as a percentage of GDP interchangeably with inequality although we emphasize on sectorial output for brevity in explanations, in the above order required for the VAR system estimation. Therefore GDP per capita is the most exogenous variable and production output from sectors as a percentage of GDP and inequality as the case maybe are the most endogenous variables. Output from sectors is endogenously affected by GDP and government spending (particularly on infrastructural development which has the capability of attracting FDI through the provision of enabling environment); higher GDP will mean probably higher output from sectors ordinarily.

A sector is not likely to affect GDP adversely particularly in economies with multiple sectors however diminished social infrastructural provision due to diminished government spending on social infrastructure will mean poor FDI inflow is likely to affect GDP making capital inflow into the economy a buffer for effects of shocks from sectorial output to aggregate GDP. The model interpretation requires a delay in the direct observation of sectorial output and profits attributable to firms given the business environment, therefore GDP will only respond to sectorial performances with lag. The three variable model is a simple model that contains GDP per capita, government spending and sectorial outputs Sec. out/GDP expressed in this particular order for the identification of the VAR system.

$$\text{GDP per capita}_{\,i,t} \rightarrow \text{govspend}_{i,t} \rightarrow (\frac{Sec.Out}{GDP})_{i,t}.$$

This allows us to state that a set of endogenous equations influence each other therefore sectorial output is contemporaneously affected by GDP and government Spending. Lower GDP will therefore result in lower output in firms and lower FDI inflows due to poor social infrastructural provision will affect firm capacity to produce also. Theoretically therefore GDP will respond only to sectorial outputs from past periods. The three variable PVAR model is presented below as

$$\begin{bmatrix} 1\ \alpha_{12} & \alpha_{13} \\ \alpha_{21}\ \ 1\ \alpha_{23} \\ \alpha_{31}\ \ \alpha_{32}\ \ 1 \end{bmatrix} \begin{bmatrix} \left(\Delta\frac{GDP}{POP}\right)_{i,t} \\ (\Delta gov.spend)_{i,t} \\ \left(\Delta\frac{\text{SEC}\,OUT}{GDP}\right)_{i,t} \end{bmatrix} = \begin{bmatrix} \alpha_{10} \\ \alpha_{10} \\ \alpha_{10} \end{bmatrix} + \begin{bmatrix} L_{11}\ L_{12}\ L_{13} \\ L_{21}\ \ L_{22}\ \ L_{23} \\ L_{31}\ \ L_{32}\ \ L_{33} \end{bmatrix} \begin{bmatrix} \left(\Delta\frac{GDP}{POP}\right)_{i,t-p} \\ (\Delta gov.spend)_{i,t-p} \\ \left(\Delta\frac{\text{SEC}\,OUT}{GDP}\right)_{i,t-p} \end{bmatrix} + \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix}$$

Where $y_{i,t}$ is a 3 variable vector including 3 endogenous variables: GDP per capita $\Delta\frac{GDP}{POP}$, government spending $\Delta gov.spend$ and sectorial outputs $\Delta\frac{\text{SEC}\,OUT}{GDP}$. The coefficients of the contemporaneous relationship are given by L, which is a 3x3 matrix that depicts the relationship between the 3 variables. The impulse response of sectorial outputs to shocks in GDP and government spending are subjects of strong interests in the study see Gravier-Rymaszewska (2012) for further discussion.

5.3. Data Presentation

In this subsection the data for the study is presented. The VAR estimation technique requires that the data is transformed to remove the trend and only keep data with variations. Employing the use of panel data ensures that the underlying structure is the same for each cross sectional unit .i.e. the matrices L coefficients are the same for all the countries in the sample. Fixed effects (μ_i) are introduced to overcome the restriction of the above constraint and allow for country heterogeneity. The limitations that the fixed effects are correlated with the regressors due to the use of lags of the dependent variables (Arellano and Bond 1991; Blundell and Bond 1998), makes us adopt a procedure called the Helmert transformation, a forward mean-differencing to eliminate the fixed effects (Arellano and Bover 1995) to keep the orthogonality between variables and their lags so that lags can be employed as instruments.

The issue of cross-sectional autocorrelation is dealt with by subtracting from each series at any time the average of the group see Levin and Lin (2002), for cross-sectional auto-correlation related to the common factors. The model is run in first difference to emphasize on the dynamics of sectorial output (and life expectancy as the case maybe) adjustments to and short run effects of shocks. The data is tested for stationarity in order to proceed with panel VAR. The data is in fact stationary as they are in first differences although the test is carried out for scrutiny. The main variables of interest GDP per cap, government spending, and sectoral output from sectors. Data for some ten selected African countries (they include Algeria, Egypt, Ghana, Nigeria, Kenya, Uganda, Cameroon and Congo) two from each major economic regions (i.e. North, West, East, Central, and Southern Africa) in Africa is utilized in the study for the period of 1985 to 2012 a period of 28 years all obtained from World Bank Data, are found to be stationary after conducting the Levin and Lin (2002), the Breitung (2001) and the Im, peaseran snd Shin (2003) unit root test . These are reported in the table below. It is therefore appropriate based on these test to proceed by estimating the model with panel VAR models.

Table 1. Panel Unit Root Tests

Variables→ Test ↓	Life exp.	Man.Sec. Out	Agr.Sec.Out.	Ser.Sec.Out.	Gov.spending	GDP per capita
Levin-Lin- Chu Adjusted t* p-value	-17.6118 0.0001	-18.4072 0.0000	-13.0876 0.0000	15.0543 0.0000	-12.0567 0.0000	-0.6510 0.0000
Breitung unit-root test Lambda p-value	-3.8734 0.0000	-2.5412 0.0002	-8.8761 0.0000	-2.1265 0.0000	-3.7645 0.0016	-2.8712 0.0000
Im-Peasaran-Shin z-tilde-bar p-value	-4.2176 0.0000	-5.3127 0.0000	-3.7167 0.0018	-4.7123 0.0000	-2.6519 0.0000	-6.8712 0.0000

Note: H_0: Panels contain unit roots H_α : Panels are stationary Common AR parameter

Number of panels = 10

Number of periods = GDP per capita (27) le (27) Gov. spending (25) Agr. Sec. out. (25) Man. sec. out.(24) Ser. Sec. out (24)

Source: Authors Compilations

5.4. Discussion of Results

In this subsection a discussion of the results is undertaken. The study investigates the effect of shocks on a set of macroeconomic variables on inequality (measured by life expectancy) and the implication of these on three sectors (i.e. the agricultural, services and manufacturing sectors) that are perceived to provide economic empowerment in form of employment for people living in the African countries in our sample. The reason for this is to examine the impact of shocks on sectorial output since increased output might mean improved economic empowerment. Our main findings confirm strong negative relationship between GDP growth and life expectancy considering the whole sample. The response of life expectancy to GDP shocks is stronger and significant in the second lag of GDP. This suggests that improvement in GDP growth does not cause any reasonable improvement in inequality reduction since government spending were not sufficiently reducing mortality rates in countries.

While GDP explains more of the government expenditure spending pattern in countries, negative GDP shocks are likely to account for up to 15% of government spending reduction in countries. The impulse response function gives us information on the short run dynamics of shocks impact. Most shocks start to have noticeable influences on the economy after the third lag and are likely to be absorbed probably 4 to 5 periods later. Our analysis of results suggests that shocks trigger structural changes, while government spending is negatively affected by GDP shocks, spending are likely to become more resilient after adjustments to shocks, therefore in times of growth expenditure spending are also likely to increase. The transmission of GDP shocks to inequality therefore is likely to be through expenditure spending on social welfare and infrastructural provision which despite increased growth in recent time has not sufficiently improved living conditions in many African countries

Finally, on extending the model to three sectors (the agricultural sector, services sector and manufacturing sectors) that have the capability to provide economic empowerment, we find that economic fluctuations decreases government spending and introduces a level of uncertainty to output production in sectors, government fiscal and monetary policies were found to have strong consequences on inequality and expenditure spending decisions, therefore these economic variables and government decisions were largely shaping inequality in countries.

System GMM Main Results for the Three Variable PVAR Model

Table 2. Full Sample Regression for Life Expectancy

SHOCKS Response of	Response to GDP_{t-1}	Response to $Gov.spend_{t-1}$	Response to GDP_{t-2}	Response to $Gov.spend_{t-2}$
Life.exp.	-.00002 (.00002) t=-.9023	-.0013 (.0008)* t=-1.7051	-.00010 (.00002)*** t=-3.8957	.0006 (.0009) t=.6934

Notes: ***indicates 1 percent significance level t-test$>$ 2.35: ** 5 percent significance level t-test $>$ 1.96 , * 10 percent significance level t-test $>$ 1.65 respectively. All standard errors are in parenthesis. The model is estimated by system GMM, while the country fixed effects and common factors are remover before estimation.

Table 3. Full Sample Regressions for Manufacturing Sector Output

SHOCKS Response of	Response to GDP_{t-1}	Response to $Gov.spend_{t-1}$	Response to GDP_{t-2}	Response to $Gov.spend_{t-2}$
Manufacturing Output	.0001 (.0100) t=.01	-.0065 (.5030) t=.0130	-.0010 (.0010) t=.01	-.0267 (.0854) t=.313

Notes: ***indicates 1 percent significance level t-test$>$ 2.35: ** 5 percent significance level t-test $>$ 1.96 , * 10 percent significance level t-test $>$ 1.65 respectively. All standard errors are in parenthesis. The model is estimated by system GMM, while the country fixed effects and common factors are remover before estimation.

Table 4 Full Sample Regression for Agricultural Sector Output

SHOCKS Response of	Response to GDP_{t-1}	Response to $Gov.spend_{t-1}$	Response to GDP_{t-2}	Response to $Gov.spend_{t-2}$
Agricultural Output	-.0017 (.0014) t=-1.21	.0210 (.1789) t=.12	.0013 (.0015) t=.86	.1944 (.3348) t=.581

Notes: ***indicates 1 percent significance level t-test$>$ 2.35: ** 5 percent significance level t-test $>$ 1.96 , * 10 percent significance level t-test $>$ 1.65 respectively. All standard errors are in parenthesis. The model is estimated by system GMM, while the country fixed effects and common factors are remover before estimation.

Table 5 Full Sample regression for Services Sector Output

SHOCKS Response of	Response to GDP_{t-1}	Response to $Gov.spend_{t-1}$	Response to GDP_{t-2}	Response to $Gov.spend_{t-2}$
Services Output	.0012 (.0038) t=.32	.1883 (.1780) t=1.05	-.0007 (.0045) t=-.1568	-.3092 (.1794)* t=1.72

Notes: ***indicates 1 percent significance level t-test> 2.35: ** 5 percent significance level t-test > 1.96 , * 10 percent significance level t-test > 1.65 respectively. All standard errors are in parenthesis. The model is estimated by system GMM, while the country fixed effects and common factors are remover before estimation.

Source: Authors Compilation

The analysis of above results is as follows, the effect of shocks in the first period does not significantly affect life expectancy. This is however significant in the second period -.0001 see table 2. Government fiscal spending had a decreasing effect on life expectancy -.0013 (t=-1.7051) see table 2, in the first period but dies away in subsequent periods, this depicts that government often adjust budget deficit and seek alternative ways to fund socio infrastructure. See also tables 6 to 9 to see the effect of shocks in subsequent periods.

For sectors GDP and fiscal shocks are not noticeable in the first periods for manufacturing and services see table 3 and 4 .0001 and .0012 respectively, however these have negative effects on the sectors in the second period. The agricultural sectors in many African economies is characterized by large informal subsistence cultivation GDP shocks are not noticeable in the second period.

Persistence of shocks was found to have strong negative effects on life expectancy (depicting increases in inequality). Shock persistence was found to also have negative implications for the services and manufacturing sectors leading to contraction in output productivity from these sectors. Decreases in sectorial output will mean less capacity for sectors to create meaningful employment even though this may not suggest a high level of staff disengagement.

Table: 6 Forecast Error Variance Decomposition for the full Sample with Life expectancy Variance of life expectancy as explained by shocks in each variable

Full Sample Variables	t=2	t=3	t=4	t=6	t=8
GDP per capita	0.0003	0.0013	0.0023	0.0023	0.0023
Gov.spend	0.0674	0.0946	0.1218	0.1218	0.1218
Life. Exp.	0.9322	0.9042	0.8762	0.8762	0.8762

Source: Authors Compilations

Table: 7 Forecast Error Variance Decomposition for the full Sample with Agricultural sector output Variance of agricultural output as explained by shocks in each variable

Full Sample Variables	t=2	t=3	t=4	t=6	t=8
GDP per capita	.0112	.0101	.0090	.0079	.0079
Gov. spend	.0135	.0114	.0093	.0072	.0072
Agr. out	.9746	.9784	.9862	.9940	.9940

Source: Authors Compilations

Table: 8 Forecast Error Variance Decomposition for the full Sample with Service Sector Output Variance of services output as explained by shocks in each variable

Full Sample Variables	t=2	t=3	t=4	t=6	t=8
GDP per capita	.0548	.9999	1.4500	1.4500	1.4500
Gov. spend	.1273	.1517	.1517	.1517	.1517
Ser. Out.	.8178	.7484	.6792	.6792	.6792

Source: Authors Compilations

Table: 9 Forecast Error Variance Decomposition for the full Sample with Manufacturing Sector Output Variance of life expectancy as explained by shocks in each variable

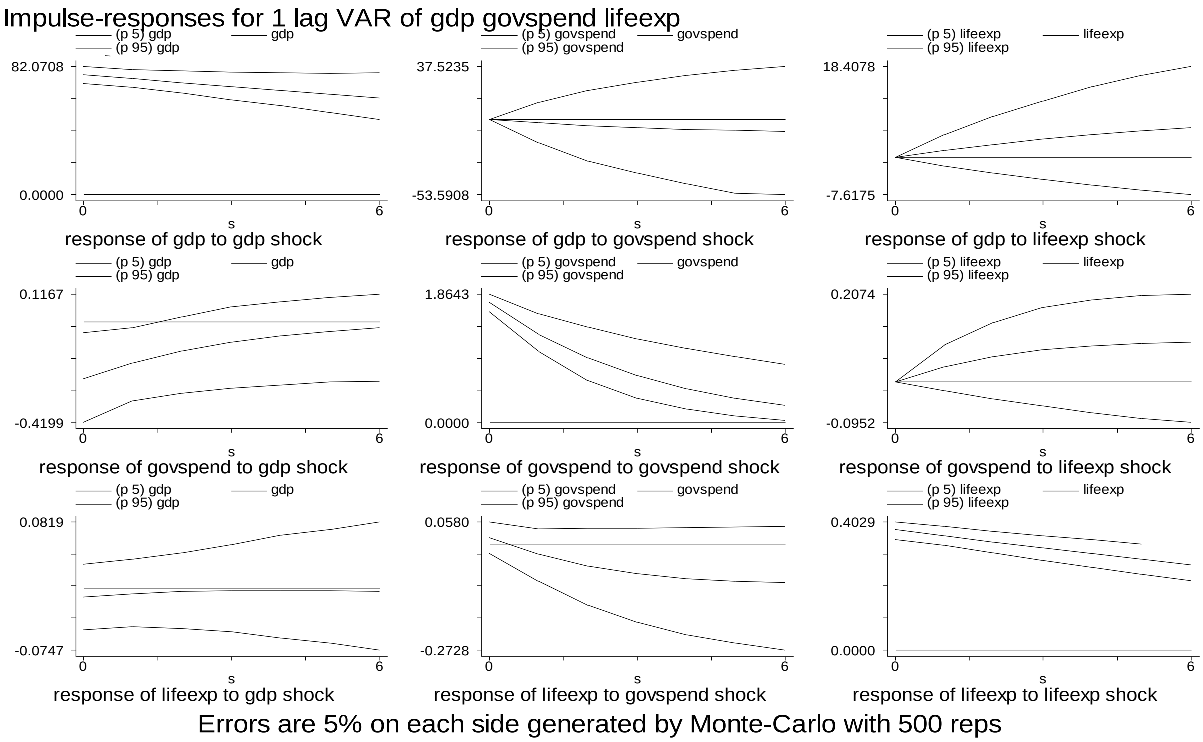
Full Sample Variables	t=2	t=3	t=4	t=6	t=8
GDP per capita	.8673	.8867	.8860	.8860	.8860
Gov. spend	.1123	.1124	.1125	.1125	.1125
Man. out	.0203	.0209	.0215	.0215	.0215

Source: Authors Compilations

The response of government to shocks to GDP and fiscal spending in their decision to improve welfare were found to be interesting. It was observed that while decreases in welfare were noticeable and affected inequality in subsequent periods. Fiscal spending decreases were only noticeable initially. In subsequent periods governments probably adjusted budgets and sourced for alternative funds to finance infrastructural provision.

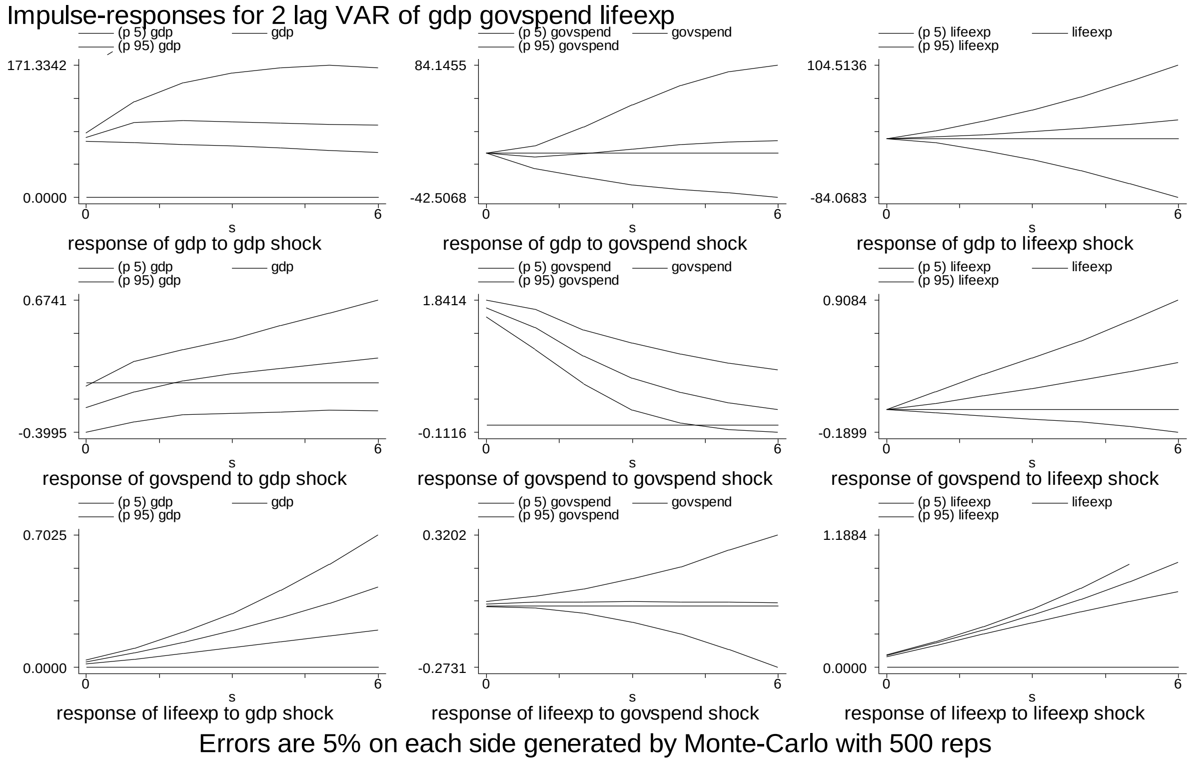
The variance decomposition for sectors yield that shocks to the manufacturing and services sectors affect output production for sectors. The negative effects of shocks to the agricultural sector are less; this is due to the informal nature of the sector.

Fig. 6 One lag impulse response function of life expectancy to shocks in GDP and government Spending



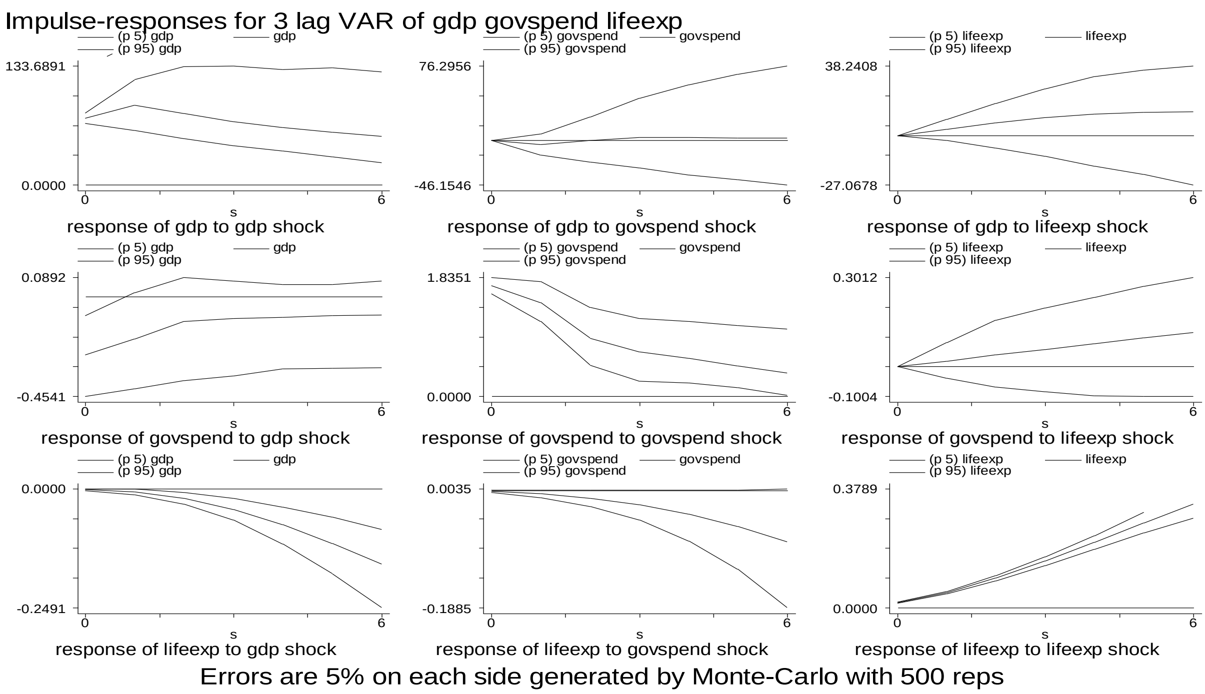
Source: Author's computations

Fig. 7 Two lag impulse response function of life expectancy to shocks in GDP and government Spending



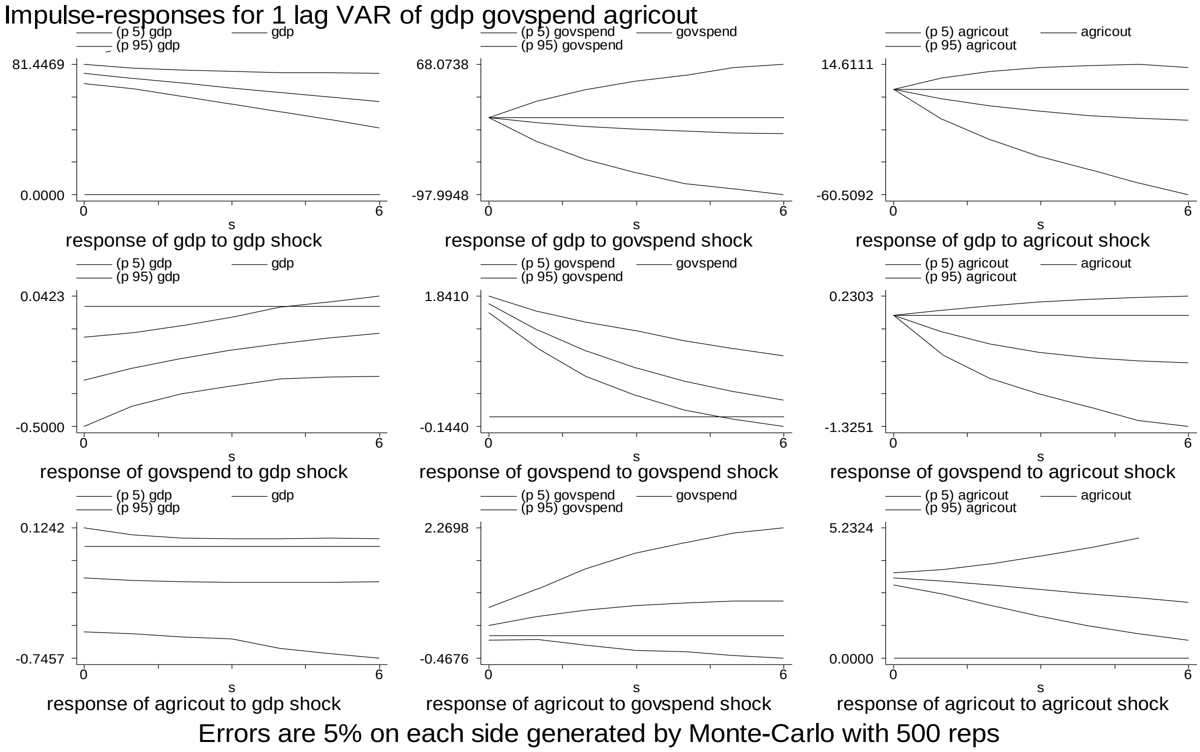
Source: Author's computations

Fig. 8 Three lag impulse response function of life expectancy to shocks in GDP and government Spending



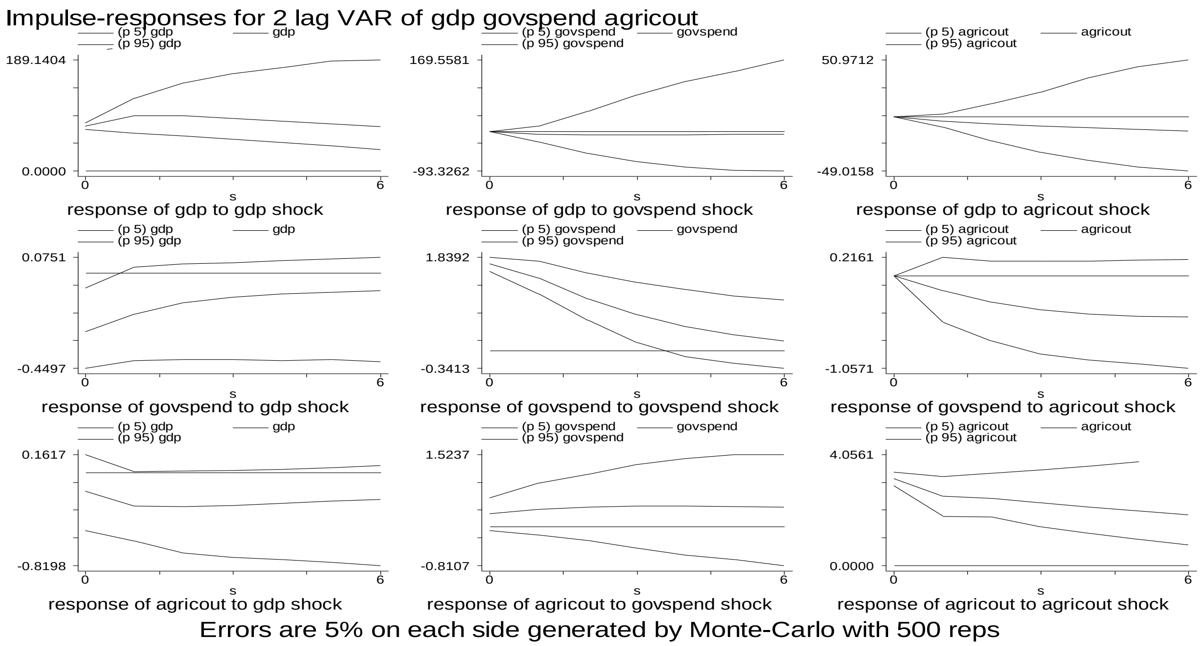
Source: Author's computations

Fig. 9 One lag impulse response function of agricultural output to shocks in GDP and government Spending



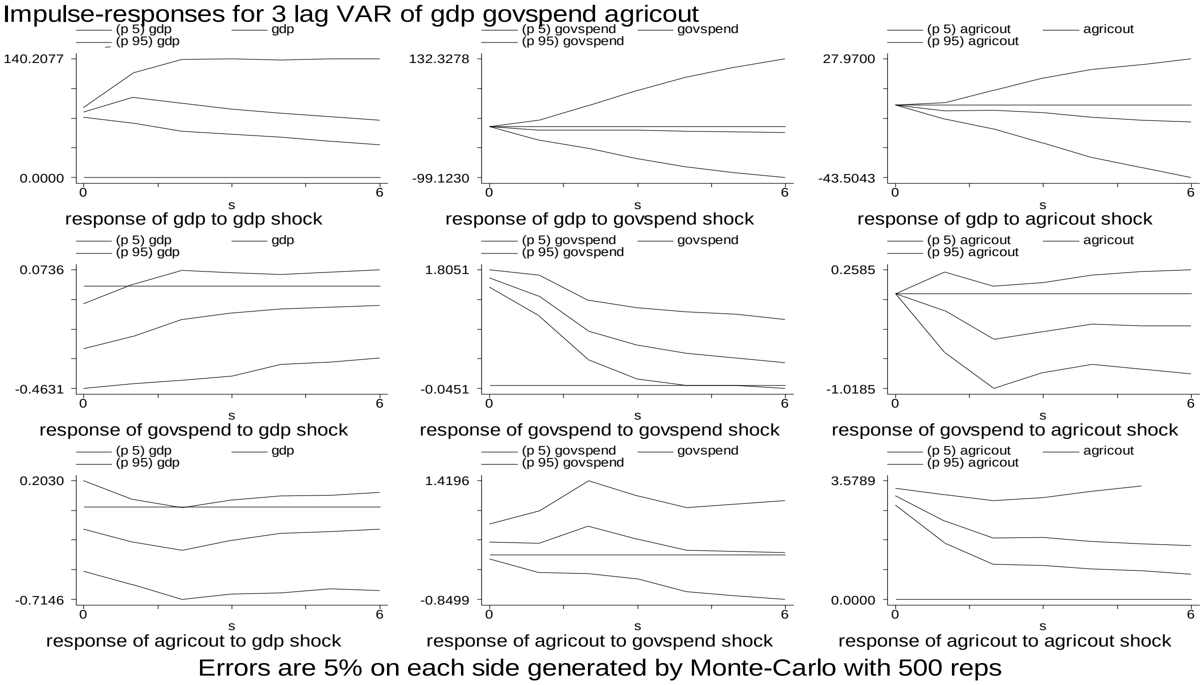
Source: Author's computations

Fig. 10 Two lag impulse response function of agricultural output to shocks in GDP and government Spending



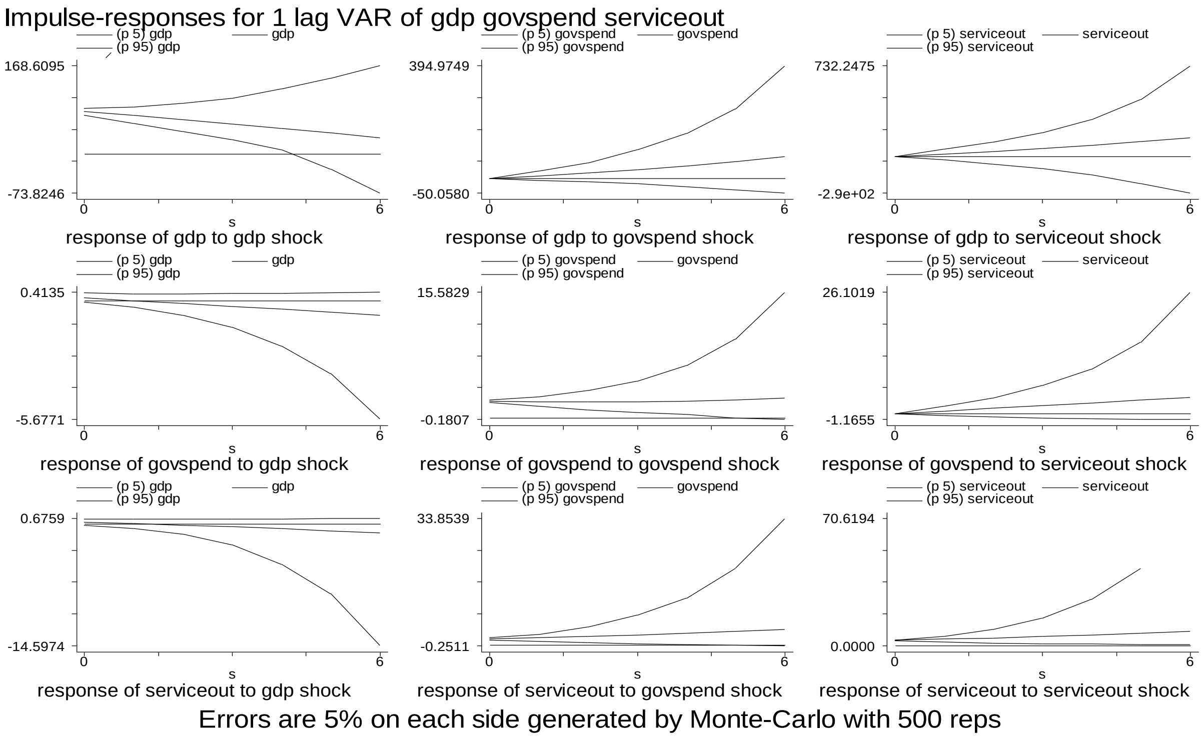
Source: Author's Computations

Fig. 11 Three lag impulse response function of agricultural output to shocks in GDP and government Spending



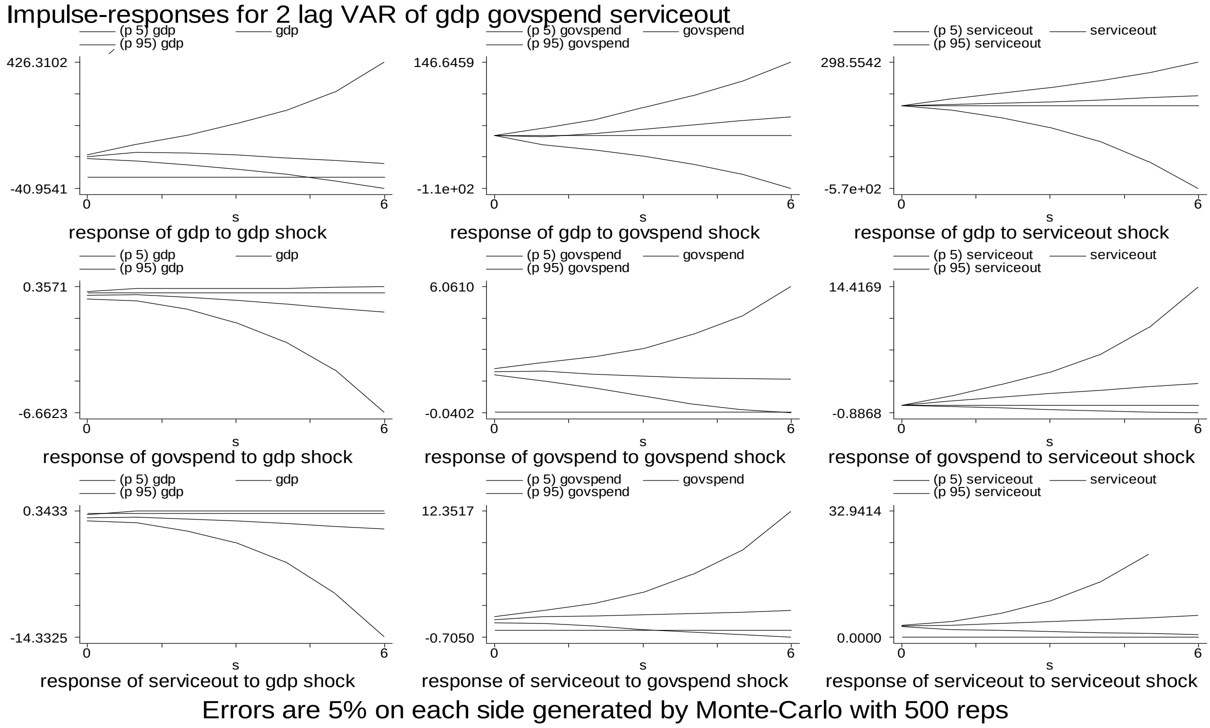
Source: Author's Computations

Fig. 12 One lag impulse response function of services output to shocks in GDP and government Spending



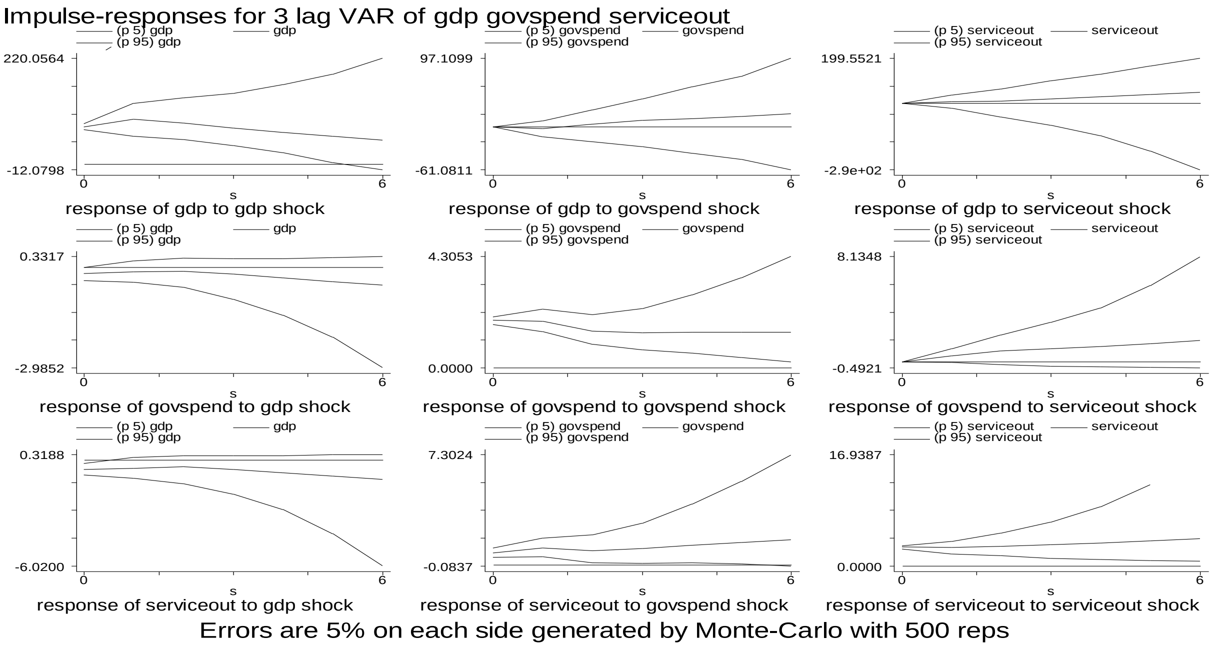
Source: Author's Computations

Fig. 13 Two lag impulse response function of services output to shocks in GDP and government Spending



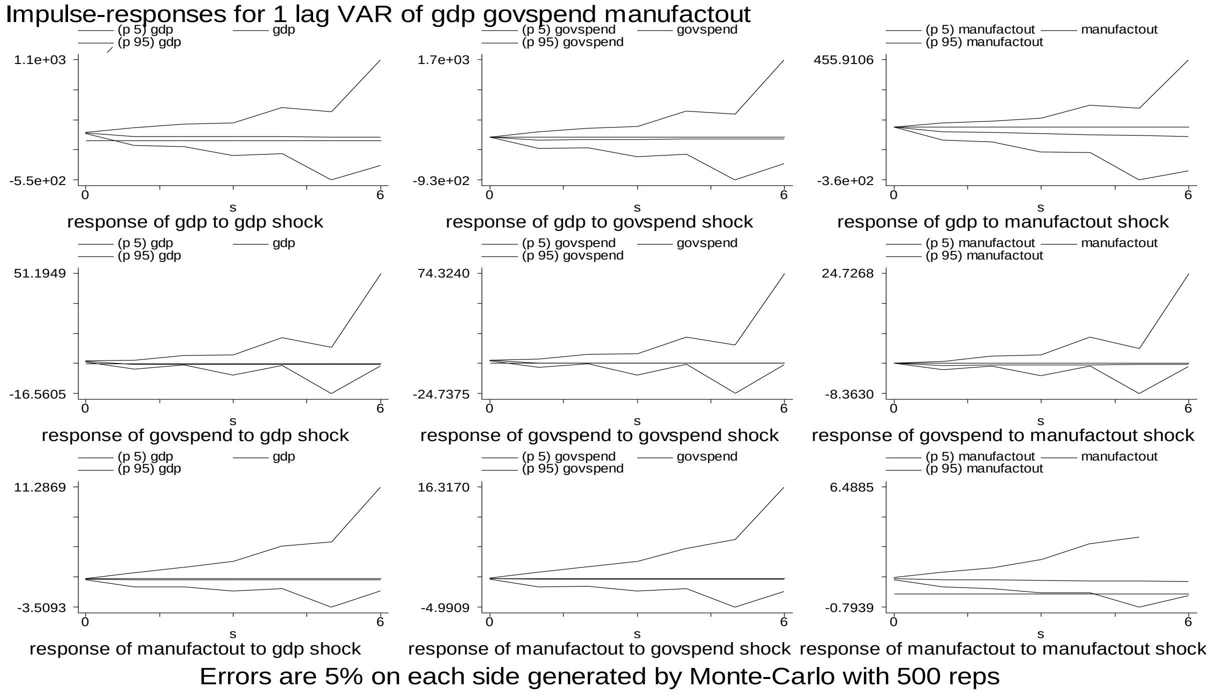
Source: Author's Computations

Fig. 14 Three lag impulse response function of services output to shocks in GDP and government Spending



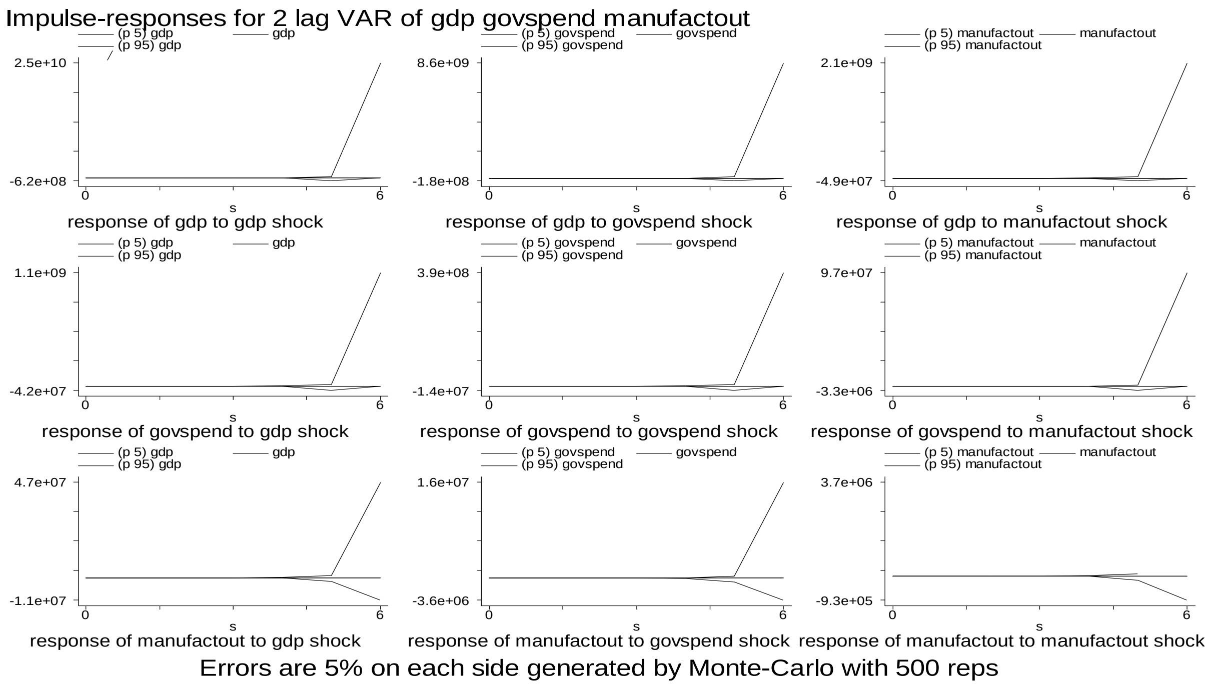
Source: Authors Computations

Fig. 15 One lag impulse response function of manufacturing output to shocks in GDP and government Spending



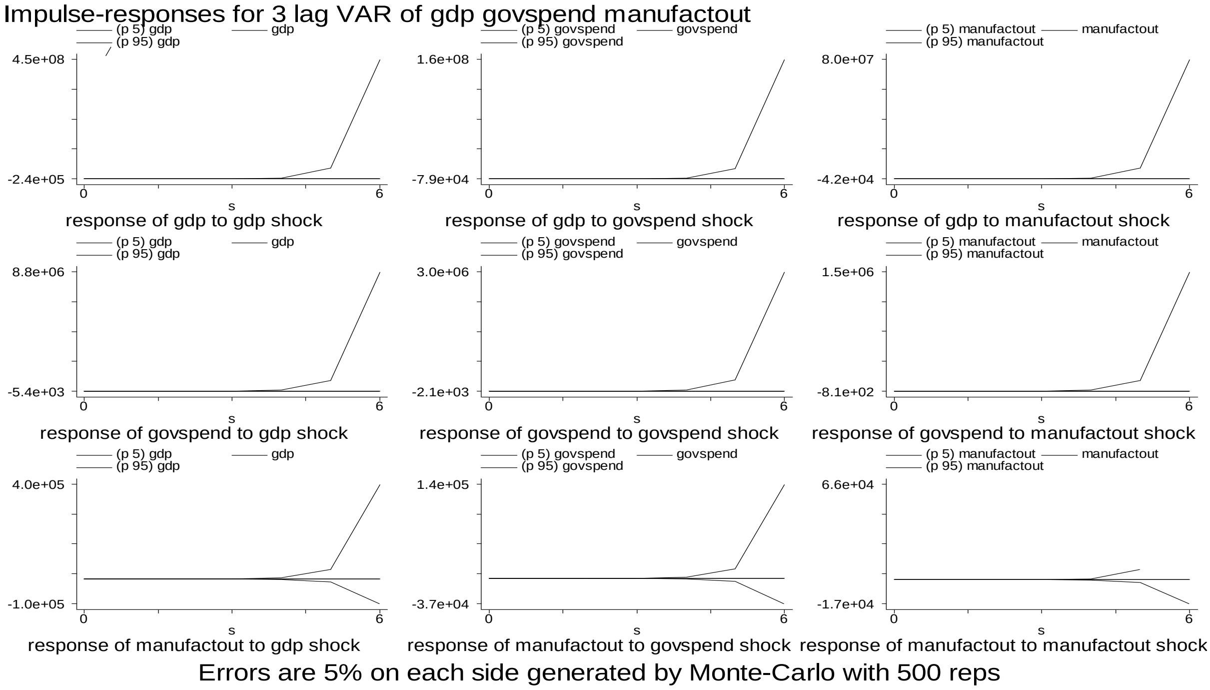
Source: Authors Computations

Fig. 16 Two lag impulse response function of manufacturing output to shocks in GDP and government Spending



Source: Authors Computations

Fig. 17 Three lag impulse response function of manufacturing output to shocks in GDP and government Spending



Source: Authors Computations

In providing answers to the study objectives we rely on the results outcomes:

- Although the effect of GDP shocks to life expectancy were not immediate, it was found to be transmitting reducing effects to life expectancy and increasing inequality in countries.
- Negative shocks were also exerted on sectorial output production for the services and manufacturing sectors although these did not have significant implications for manufacturing.
- Negative shocks were found to have weak implications for the services sector, this were not noticeable for the agricultural sector, these are two labour intensive sectors with significant implicative effects for employment.

Conclusion

In this section we conclude. Negative shocks to GDP were found to significantly increase inequality in countries particularly after the first periods. The same were observed for sectors although the results were only weakly significant for the services sector. The impulse response function of life expectancy to GDP and fiscal spending had strong negative implications for life expectancy. For sectors these were not immediately noticeable for the manufacturing and services sector outputs and no decreases were observed for the agricultural sector.

The shock to fiscal spending on life expectancy increased inequality after the first period and were found to frizzle out in the subsequent periods as government adapted to shocks, through budget adjustments.

GDP and fiscal shocks due to financial volatility were found to have negative impacts on life expectancy and also across sectors indicating that a high level of uncertainty due to financial friction can have strong consequences for inequality in Africa, although this effect was not significant for the manufacturing and agricultural sector outputs. In concluding, the assertion earlier made that a sector is not likely to affect GDP adversely particularly in economies with multiple sectors but that diminished social infrastructural provision due to reduction in government spending on social infrastructure, will mean poor FDI inflow which is likely to affect GDP appeared to be quite plausible, therefore improved capital inflow into the economy could act as a buffer for the effects of shocks from sectors to aggregate GDP.

Recommendation

In this subsection we make useful recommendation for policy purposes. It is necessary for government to provide basic social security blankets for people living below the poverty line in many African countries by making basic medical facilities more accessible and easily affordable particularly in rural communities to help reduce poverty and mortality rates in general, since it was discovered that GDP shocks to life expectancy was transmitting reducing effects to life expectancy and increasing inequality in countries.

Sectoral performances also show poor ability of labour intensive sector to withstand the negative shocks in GDP. Adequate attention should be paid to socio infrastructural challenges as this could reduce the transaction cost of private firm activity. Since government also seem to be the highest employer of labour in the services sector e.g. schools, hospitals, airports and other social services, encouragement of other sectors such as manufacturing and agriculture where such negative implicative job reducing effects are likely to occur should be boosted.

Finally effective use of government funds as an intervention mechanism particularly in short term sectoral improvements such as business information provision, reduction in business permits processing time and avoidance of multiple taxation could encourage private investment in the manufacturing and agricultural sectors

The Authors Express Thanks to Inesa Love for PVAR Codes, and express heartfelt gratitude to Gravier-Rymaszewska for making his paper available.

References

[1] Africa's Pulse 2014 "World Bank Journal on Economic Growth" Vol. 9 April 2014 Available on http://www.worldbank.org/content/dam/Worldbank/document/Africa/Report/Africas-Pulse-brochure_Vol9.pdf
[2] Arellano, M., and Bond, S. (1991). 'Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations'. The Review of Economic Studies, 58(2): 277.
[3] Arellano, M., and Bover, O. (1995). 'Another look at the instrumental variable estimation of error-components models'. Journal of Econometrics, 68(1): 29-51.
[4] Blundell, R., and Bond, S. (1998). 'Initial conditions and moment restrictions in dynamic panel data models'. Journal of Econometrics, 87(1): 115-43.
[5] Boschini, A., and Olofsgård, A. (2007). 'Foreign aid: An instrument for fighting communism?'. Journal of Development Studies, 43(4): 622-48.
[6] Breitung, J. (2001). 'The local power of some unit root tests for panel data'. Bingley: Emerald Group Publishing Ltd.
[7] Canova, F. and M. Ciccarelli, (2004), Forecasting and turning point predictions in a Bayesian panel VAR model, Journal of Econometrics 120, 327-59
[8] Chong, A., and Gradstein, M. (2008). 'What determines foreign aid? The donors' perspective'. Journal of Development Economics, 87(1): 1-13.
[9] Dang, H., Knack, S., and Rogers, F. (2009). 'International aid and financial crises in donor countries'. Policy Research Working Paper 5162. Washington, DC: World Bank.
[10] Faini, R. (2006). Foreign aid and fiscal policy. London: Centre for Economic Policy Research.
[11] Frot, E. (2009). 'Aid and financial crisis: shall we expect development aid to fall?'. Working Paper. Stockholm: Stockholm School of Economics.
[12] Hansen, H., and Headey, D.D. (2010). 'The short-run macroeconomic impact of foreign aid to small states: An agnostic time series analysis'. The Journal of Development Studies, 46(5): 877-96.
[13] Hansen, H., and Tarp, F. (2001). 'Aid and growth regressions'. Journal of Development Economics, 64(2): 547-70.
[14] Gravier-Rymaszewska(2012) "How Aid Supply Responds to Economic Crises A Panel VAR Approach" UNUWIDER Working Paper Available on http://www.wider.unu.edu/publications/working-papers/2012/en_GB/wp2012-025/
[15] Im, K., Pesaran, M., and Shin, Y. (2003). 'Testing for unit roots in heterogeneous panels'. Journal of Econometrics, 115(1): 53-74.
[16] Knack, S. (2001). Aid dependence and the quality of governance: Cross-country empirical tests. Southern Economic Journal, 68(2), 310–29.
[17] Kraay, Aart (2004). "When Is Growth Pro-Poor? Cross-Country Evidence". IMF Working Paper Available on https://www.imf.org/external/pubs/ft/wp/2004/wp0447.pdf
[18] Laeven, L., and Valencia, F. (2010). 'Resolution of banking crises: The good, the bad, and the ugly'. Working Paper 10: 146. Washington, DC: IMF.
[19] Levin, A., Lin, C., and James Chu, C. (2002). 'Unit root tests in panel data: asymptotic and finite-sample properties'. Journal of Econometrics, 108(1): 1-24.
[20] Ojeaga (2012) "Economic policy, does It help life expectancy? An African Evidence of the Role of Eonomic Policy on Longevity." Available on https://mpra.ub.uni-muenchen.de/secure/40199/1/MPRA_paper_40199.pdf
[21] Rajan, R., and Subramanian, A. (2008). 'Aid and growth: What does the cross-country evidence really show?'. The Review of Economics and Statistics, 90(4): 643-65.
[22] Reinhart, C., and Rogoff, K. (2008), 'Banking crises: an equal opportunity menace'. Technical report, Cambridge, Mass: National Bureau of Economic Research.
[23] World Bank Press Release 2012.
[24] World Bank Statistics 2013.

Bayesian modelling of real GDP rate in Romania

Mihaela SIMIONESCU[1*]
[1] Romanian Academy, Institute for Economic Forecasting, Bucharest, Calea 13 Septembrie, no. 13, District 5, Bucharest[1]

Abstract

The main objective of this study is to model and predict the real GDP rate using Bayesian approach. A Bayesian VAR (BVAR), a Bayesian linear model and switching regime Bayesian models were employed for the real GDP rate, inflation rate and interest rate. From the set of variables that were connected to real GDP, for identifying the most relevant ones using the data for Romanian economy, we applied the selection algorithm based on stochastic search. Weight of revenues in GDP, weight of budgetary deficit in GDP, investment rate and inflation rate are the most correlated variables with the real GDP rate. The averages of posterior coefficients of models were used to make forecasts. For Romania on the horizon 2011-2014, the unrestricted switching regime models generated the most accurate forecasts.

Keywords: *Bayesian model, forecasts, GDP rate, switching regime*

1. Introduction

The main aim of this study is to propose various types of Bayesian models to predict the real GDP in Romania. Moreover, a Bayesian algorithm was applied to select the variables that explain better the evolution of real GDP rate in Romania. The advantages of Bayesian approach are essential for the case of Romanian economy when the lack of long time series is a serious problem. Bayesian methods have a general character that does not require special regularity conditions, concepts like confidence interval and assumptions' testing. The Bayesian approach conducts us to the evaluation of the VAR and linear regression models by using Bayesian principles. These types of models were employed in this study for annual data. Moreover, the proposed econometric models were used to construct some predictions during 2011-2014.

After this Introduction, a literature review is made, being followed by the estimation of Bayesian models that are used in making forecasts. The results indicated a good accuracy of real GDP forecasts during 2011-2014.

2. Literature review

In Romania a BVAR model for quarterly GDP was built by a researcher who made the comparison with unrestricted VAR model, a OLS regression and random walk in terms of forecasts accuracy. The results put into evidence a slow recovery of the Romanian economy, for the next quarters (2009Q4-2010Q4) the Bayesian model predicting a negative gap [1].

The real GDP growth rate depends on investment rate, following the relationship [2]:

$$\mathrm{rGDP_t} = \mathrm{a} + \mathrm{b} \cdot \mathrm{rinv_t} + \mathrm{eps_t} \tag{1}$$

$rGDP_t$- real GDP growth rate at time t

$rinv_t$- investment rate at time t

eps_t- error term

a,b- parameters to be estimated

[*] Corresponding author: mihaela_mb1@yahoo.com.

This paper has been financially supported within the project entitled "*Routes of academic excellence in doctoral and post-doctoral research,* contract number *POSDRU/159/1.5/S/137926,* beneficiary: Romanian Academy, the project being co-financed by European Social Fund through Sectoral Operational Programme for Human Resources Development 2007-2013.

Some authors more multiple regression models for explaining the GDP growth rate [3]:

$$\mathrm{rGDP_t} = \alpha_0 + \alpha_1 \cdot \mathrm{income\_tax_t} + \alpha_2 \cdot \mathrm{expenses\_weight_t} + \alpha_3 \cdot \mathrm{deficit_t} + \varepsilon_{1t} \quad (2)$$

$$\mathrm{rGDP_t} = \alpha_0 + \alpha_1 \cdot \mathrm{income\_tax_t} + \alpha_2 \cdot \mathrm{expenses\_weight_t} + \alpha_3 \cdot \mathrm{deficit_t} + \alpha_4 \cdot \mathrm{BCF_t} + \alpha_5 \cdot \mathrm{rpop_t} + \varepsilon_{2t} \quad (3)$$

$$\mathrm{rGDP_t} = \alpha_0 + \alpha_1 \cdot \mathrm{income\_tax_t} + \alpha_2 \cdot \mathrm{expenses\_weight_t} + \alpha_3 \cdot \mathrm{other\_revenues_t} + \alpha_4 \cdot \mathrm{GCF_t} + \alpha_5 \cdot \mathrm{rpop_t} + \varepsilon_{3t} \quad (4)$$

$rGDP_t$- real GDP growth rate

$income\_tax_t$- weight of tax in GDP

$expenses\_weight_t$- weight of total expenses in GDP

$deficit_t$- weight of deficit or surplus in GDP

$other\_revenues_t$- weight in GDP of net revenues from other sources except for dues

GFC_t- weight in GDP of gross fixed capital formation

$rpop_t -$ population growth rate

$\varepsilon_{1t}, \varepsilon_{2t}, \varepsilon_{3t}$- errors terms

Marcellino, Porqueddu and Venditti predicted the GDP in euro zone using the Bayesian indicator of growth. EURO-BIG is actually a mixed frequency data small scale dynamic factor model with stochastic volatility [4].

Recently, a Bayesian variant of global vector autoregressive model (B-GVAR) was proposed. The predictive performance of B-GVAR models was compared for variables like inflation rate, real GDP, real exchange rate and interest rates. By considering the international linkages the inflation, real GDP and the real exchange rate predictions are improved [5].

In order to predict the GDP growth in euro zone, some authors employed bridge models which were applied for aggregate GDP. The national bridge model predicted better the GDP rate than univariate and multivariate models that were used as benchmark. The aggregation of national predictions generated an improvement in accuracy [6]. It was proved that an improvement in GDP forecasts accuracy was brought by the use of monthly data regarding the current activity more than the use of financial variables from surveys. A large number of time series with monthly and quarterly frequency in real-time was used to predict the output growth. The performance of the factor model forecasts was assessed by comparison with the GDP of Germany. The importance of predictions revisions was analyzed in detail [8].

The real GDP rate was nonlinearly correlated with oil prices, this relationship being used to predict the real GDP growth in USA. The symmetric nonlinear processes seemed to provide more accurate forecasts than the asymmetric models [9]. The logistic-growth equations were used to predict the real GDP per capita and long-run inflation [10].

3. Modelling real GDP rate in Romania

For constructing the econometric models that explains the evolution of GDP growth rate, the data series for the following variables have been used: GDP growth rate, the weight in GDP of revenue tax, budget expenses, budget revenues, budget deficit, gross fixed capital formation, population growth rate, inflation rate, unemployment rate, interest rate, investment rate. The sources of data are Eurostat and National Institute of Statistics and the period is 1991-2013.

A Bayesian VAR (BVAR) of order 2 was employed for the real GDP rate, inflation rate and interest rate. In Appendix 1 the posterior coefficients, the posterior covariance matrix and constant terms were presented. For making predictions of real GDP growth, we used the average of posterior coefficients. 10 000 replications were saved for this application, the total number of iterations being 50 000. The form of BVAR model is the following one:

$$Y(t) = c + Phi(1)*Y(t-1) + ... + Phi(p)*Y(t-p) + ut \quad (5)$$
ut ~ MVN(0,Sigma), Y(t) inclused d variables

The Gibbs sampler algorithm with prior values as in Lindley and Smith (1972) was applied:

Phi ~ N(mu,V), Sigma ~ W(Omega,df)

The results consist in:

- Phi_draws = posterior values of phi(1);
- Sigma_draws = posterior values for covariance matrix;
- constant_draws = posterir values for constant.

Posterior conditional distribution Sigma follows an inverse Wishart distribution.

With Lindley and Smith (1972) proper priors, draws from posterior conditionals of Beta and Sigma2 are obtained successively. When the priors are flat, the posterior means of Beta and Sigma2 are similar to the OLS estimator.10 000 replications were saved from this application, the total number of iterations being 50 000.

The GDP growth rate is explained using the weight of deficit in GDP, the weight of tax in GDP and the weight of gross capital formation in GDP using linear Bayesian models. In the Appendix 2 posterior mean and posterior standard deviations were displayed for coefficients and for the errors variance.

The switch of regime breaks the regression into two regression models with different slope and disturbances variance. The regime switch time is unknown and has a uniform prior. Then the posterior is proportional to the likelihood in which switch occurs at a given time. Priors and posteriors of other parameters follow the standard linear regression model.

For the real GDP growth a Bayesian switching regime model was proposed. In Appendix 3 posterior mean and posterior standard deviations were displayed for coefficients and for the errors variance before and after the regime change. Two variants of the model were proposed: unrestricted regime and connected regime.

From the set of variables that were connected to real GDP, for identifying the most relevant ones using the data for Romanian economy, we applied the selection algorithm based on stochastic search, being proposed by George and McCulloch (1997).

The form of the model is:

$$Yi = Xi * Betai + ui \text{ , where } ui \sim N(0,s^2) \quad (6)$$

where Betai | taui ~taui * N(0,V1) + (1-taui) * N(0, V2) , V1 > V2

taui = 1 suggests a variable is chosen, while taui = 0 implies Betai is close to zero and can be excluded

Gibbs sampler with hierarchical proper priors was used.

At level one: s^2 ~ IG(a,b), Betai | taui ~ taui * N(0,V1) + (1-taui) * N(0, V2)

Level two supposes: taui | pi ~ Bernoulli (pi)

Level three imples: pi ~ Beta (a', b')

Hyperparameters are specified below. Conditional posteriors of Betai, s^2, pi have conjugate forms, while conditional posterior of taui is updated by the Bayes formula.

Y = dependent variable (n * 1 vector)

X = regressors (n * k matrix)

ndraws = number of draws in MCMC

burn_in = number of burn-in draws in MCMC

tol = the critical probability to accept of variable add_constant = whether to add a constant to X (default = 0)

The results are:

Beta= posterior draws of coefficients corresponding to the k regressors

Sigma2 = posterior draws of variance of disturbances

Tau = posterior draws of the variable inclusion indicators

Beta_refine = posterior draws of coefficients of the refined regression model

Sigma2_refine = posterior draws of disturbances variance of the refined regression model

The prior of regression coefficients have a Gaussian mixture, one with large variance and one with small variance. If a coefficient resides on the latter, it is an indication of exclusion since the coefficient is close to zero. The acceptance critical probability is 0.3.

In the following table the final results of the algorithm application were presented, the details being in Appendix 4.

Table 1. The results of the application of Bayesian algorithm for selecting the determinant factors for real GDP rate

Current no.	Variable	Decision(excluded/included)
1	Rate of population growth	Excluded
2	Weight of revenues in GDP	**Included**
3	Weight of expenses in GDP	Excluded
4	Weight of budgetary deficit in GDP	**Included**
5	Weight of gross capital formation in GDP	Excluded
6	Weight of tax on income in GDP	Excluded
7	Investment rate	**Included**
8	Interest rate	Excluded
9	Inflation rate	**Included**
10	Unemployment rate	Excluded

Source: own computations

From the table, we can see that only four variables were considered in the final model: weight of revenues in GDP, weight of budgetary deficit in GDP, investment rate and inflation rate. The predictions based on the proposed Bayesian models were displayed in the following table.

Table 2. Forecasts of real GDP growth (%) using Bayesian models (horizon: 2011-2014)

Year	Bayesian linear regression model	BVAR(2) model	Bayesian model with unrestricted switching regime	Bayesian model with connected switching regime	Registered values
2011	2.87	2.83	2.75	3.4	2.3
22012	4.04	1.07	2.7	4.8	0.6
22013	4.79	1.46	2.8	5.8	3.5
22014		0.48			

Source: own computations

In the category of Bayesian model, we observed that for 2011 and 2013, the Bayesian model with unrestricted switching regime generated the closest value of the prediction for these years. For 2012, the BVAR model determined the most accurate prediction, while for 2014 the unrestricted

Conclusion

The Bayesian approach is useful for modeling and predicting macroeconomic variables. In the context of economic crisis, many researchers raised questions regarding the ability of the economists to better predict the macroeconomic indicators. The failure of usual econometric model that did not anticipated the economic crisis made the economists to consider better solutions like Bayesian models.

For Romania, during 2011-2014, the unrestricted switching regime models generated the most accurate forecasts.

Acknowledgement

This paper has been financially supported within the project entitled "*Routes of academic excellence in doctoral and post-doctoral research,* contract number *POSDRU/159/1.5/S/137926,* beneficiary: Romanian Academy, the project being co-financed by European Social Fund through Sectoral Operational Programme for Human Resources Development 2007-2013.

APPENDIX 1

BVAR model for real GDP growth, inflation rate and interest rate

Constant

1.8778	5.6555	-2.5507

Posterior Phi1 coefficients

0.21	0.02	-0.21
-0.35	0.62	0.27
-0.10	-0.01	0.99

Posterior Phi2 coefficients

0.09	0.01	0.17
-1.75	-0.11	0.07
0.50	0.08	-0.12

Posterior covariance matrix of the VAR system

9.93	-38.69	-14.28
-38.69	1572.85	269.61
-14.28	269.61	113.74

APPENDIX 2

Bayesian linear regression model

The dependent variable : real GDP rate

The regressors: budgetary deficit, tax, gross capital formation.

A constant is added to regressors.

'Coeff.'	'Post. mean'	'Post. std'
'C(0)'	[-11.1979]	[5.2553]
'C(1)'	[-0.9120]	[0.4991]
'C(2)'	[0.0018]	[0.0374]
'C(3)'	[0.7078]	[0.2366]
's^2'	[22.7571]	[7.0078]

APPENDIX 3

Unrestricted regime change

------- Before Regime Change ----------

'Coeff.'	'Post. mean'	'Post. std'
'C(0)'	[-23.5789]	[1.1645]
'C(1)'	[7.4562]	[0.7588]

's^2'	[0.2302]	[0.2762]

------- After Regime Change ----------

'Coeff.'	'Post. mean'	'Post. std'
'C(0)'	[2.7298]	[1.9614]
'C(1)'	[0.0074]	[0.1374]
's^2'	[14.7063]	[4.5556]
[]	[]	[]
'switch'	[2.0967]	[0.3012]

Connetcted regime change

'Coeff.'	'Post. mean'	'Post. std'
'Const'	[-16.0164]	[5.6841]
'C(1) Before'	[4.9024]	[2.3379]
'C(1) After'	[-0.0883]	[0.3328]
's^2'	[16.4120]	[5.5463]
'Switch'	[5.0249]	[3.1956]

APPENDIX 4

Bayesian algorithm for selecting the real GDP rate determinants

Coef.	Post.mean	Post.std
C(0)	-1.657	3.087
C(1)	-0.078	0.646
C(2)	0.241	0.652
C(3)	-0.094	0.588
C(4)	-0.444	0.757
C(5)	0.002	0.082
C(6)	-0.000	0.006
C(7)	1.659	2.538
C(8)	-0.004	0.020
C(9)	-0.038	0.040
C(10)	0.053	0.281
s^2	19.037	7.555

Variable Inclusion Probabilities

Coef.	Post.mean	Post.std
Tau(0)	0.521	0.500
Tau(1)	0.271	0.445
Tau(2)	0.302	0.459
Tau(3)	0.298	0.457
Tau(4)	0.426	0.494
Tau(5)	0.072	0.259
Tau(6)	0.015	0.120
Tau(7)	0.719	0.450
Tau(8)	0.052	0.223
Tau(9)	0.500	0.500
Tau(10)	0.213	0.410

------ Refined Regression Model ------

Coef.	Post.mean	Post.std
C(0)	-3.381	3.020
C(1)	0.235	0.112

C(2)	-0.984	0.402
C(3)	2.987	1.791
C(4)	-0.054	0.022
s^2	14.019	4.472

References

[1] Caraiani, P., *"Forecasting Romanian GDP Using a BVAR Model",* Romanian Journal of Economic Forecasting 13 (4). pp. 76-87. 2010.

[2] Albu, L.-L., Scenarios for post-crisis period based on a set of presumed changes in the interest rate – investment – GDP growth relationship. *MPRA Paper*, 32753. pp.1-25. 2010.

[3] Martin, R., Fardmanesch, M., "*Fiscal variables and growth: A cross-sectional analysis*". Public Choice, 64. pp. 239-251.1990.

[4] Marcellino, M., Porqueddu, M., Venditti, F., "*Forecasting euro area GDP with EURO-BIG (Bayesian Indicator of Growth)"*. 2011.

[5] Feldkircher, M., Huber, F., Cuaresma, J. C., "*Forecasting with bayesian global vector autoregressive models: A comparison of priors"*, 189. 2014.

[6] Baffigi, A., Golinelli, R., "*Bridge models to forecast the euro area GDP*", International Journal of Forecasting, 20(3), pp. 447-460.2004.

[7] Diron, M., "*Short-term forecasts of euro area real GDP growth: an assessment of real-time performance based on vintage data*", Journal of Forecasting, 27(5). pp. 371-390.2008.

[8] Schumacher, C., Breitung, J., *"Real-time forecasting of German GDP based on a large factor model with monthly and quarterly data",* International Journal of Forecasting, *24*(3). pp. 386-398.2008.

[9] Kilian, L., Vigfusson, R. J., *"Do oil prices help forecast us real gdp? The role of nonlinearities and asymmetries*", Journal of Business & Economic Statistics, *31*(1), pp. 78-93.2013.

[10] Modis, T., "*Long-term GDP forecasts and the prospects for growth*", Technological Forecasting and Social Change, 80(8). pp. 1557-1562.2013.

Individual contributions to portfolio risk: risk decomposition for the BET-FI index

Marius ACATRINEI[*]
Institute of Economic Forecasting

Abstract

The paper applies Euler formula for decomposing the standard deviation and the Expected Shortfall for the BET-FI equity index.

Risk attribution allows the decomposition of the total risk of the portfolio in individual risk units. In this way we can compute the contribution of each company to the overall standard deviation/Expected Shortfall of the portfolio.

Keywords: *risk attribution, marginal contributions, Expected Shortfall*

JEL: *C1, G11*

1. Introduction

A portfolio contains a large number of positions on different financial instruments ranging from stock, bonds to derivatives instruments. In order to model the risk and return of the portfolio it is necessary to map the portfolio to its risk factors. The risk factors for a linear portfolio may include the prices of general market indices, foreign exchange rates or zero coupon market interest rates of different maturities to which the portfolio is exposed.

The risk factor sensitivities of an asset or portfolio measure the change in price when a factor risk changes while holding constant the other factors. In a stock portfolio the risk factor sensitivities are called betas (factor betas). In a linear portfolio, such as a stock portfolio, the mapping of the risk factors is carried out with factor models.

Risk is expressed as a sum of the contribution from each factor contributing to the overall risk evaluation. If a portfolio of securities may be mapped to a set of factors then the factors should explain most of the variation in the portfolio. APT factor models explain the mapping by segregating between systematic and idiosyncratic components. By using Euler formula it is possible to decompose the contribution of each factor and to assess the specific contribution of each factor.

Yamai (2002) and Hallerbach (2003) showed that Value at Risk can be decomposed in several components: marginal VaR, component VaR and incremental VaR assuming Gaussian distribution. The marginal VaR is the marginal contribution of the individual portfolio component to the diversified portfolio VaR, component VaR is the proportion of the diversified portfolio VaR that is attributed to the individual components and incremental VaR is the effect on the VaR of the portfolio by adding a new financial instrument.

Zhang and Rachev (2004) criticized the beta coefficient from the Capital Asset Pricing Model (CAPM) since it is neither translation-invariant nor monotonic, properties that any coherent risk measure should display. The authors define “risk attribution’ as “a process of attributing the return of a portfolio to different factors according to active investment decisions”. They show that by using Euler’s formula it is possible to identify the main sources of risk in a portfolio.

Scherer (2005) implemented risk budgeting with multiple benchmarks and rival risk regimes for accommodating the different objectives demanded by investors.

Darolles and Gouriéroux (2012) applied the risk contribution restrictions on a portfolio of futures on commodities and compared the performance of the associated portfolios in terms of risk contributions, performance and budget allocations.

[*] Corresponding author: marius.acatrinei@gmail.com

2. Factor Models

A factor model allows the analysis of the returns of a portfolio and computation of the portfolio risk. Factor models are based on univariate or multivariate linear regression. Capital Asset Pricing Model (CAPM) is a Single Factor Model (Single Index Model) which assumes a linear dependence between the expected excess returns of a single asset and the expected excess return on the market portfolio and allows to investigate the risk and return characteristic of an asset relative to the market index.

The Single Index Factor Model uses a broad market index (F_M) as a proxy for the market portfolio which is unknown. Roll (1977) showed that the market portfolio is not observable. Fama and French (1992) showed that beta and long-run average return are not correlated.

$$E(R_i) = \alpha + \beta_i E(F_M) \tag{1}$$

$$\beta_i = \frac{Cov(R_i, F_M)}{\sigma_F^2} \tag{2}$$

where F_M represents the ordinary return (not the excess returns) on the market portfolio.

$$R_i = \alpha + \beta_i F_{M,t} + \varepsilon_{it}, \varepsilon_{it} \sim i.i.d(0, \sigma_i^2) \tag{3}$$

where β_i is the risk factor sensitivity of the asset *i*, $\beta_i \sigma_F$ is the systematic volatility of the asset *i*, σ_i is the specific volatility of the asset *i* and σ_F is the volatility of the equity index.

In a Single Factor Model the total risk is decomposed in systematic risk and specific risk. The volatility of the portfolio return can be decomposed in three risk sources: 1) sensitivity to the market factor beta 2) volatility of the market factor and 3) specific risk.

Beta is a measure of risk in a portfolio since the weighted sum of individual betas equals the portfolio beta.

$$\beta_P = \sum_{i=1}^{N} w_i \beta_i \tag{4}$$

When considering a portfolio with *k* factors, the linear Multifactor Model may be written as

$$R_t = \alpha + \beta_{i1} * F_{1t} + \cdots \beta_{ik} * F_{kt} + \varepsilon_{it} \tag{5}$$

$$\begin{aligned} R_t &= \alpha + \beta_i * F_t + \varepsilon_{it}, F_t \sim (\mu_F, \Sigma_F \\ \varepsilon_{it} &= (0, \sigma_{\varepsilon,i}^2) \end{aligned} \tag{6}$$

Since the asset returns are typically non-normal, not i.i.d, there are several distributions that may fit the data better than the Gaussian distribution. Among the distributions usually used in practice for fitting financial returns are Student's-*t*, skewed Student-*t*, GED, generalized Pareto, etc. Since in practice the portfolios may include large number of assets sometimes in small samples and with missing data, the multivariate modelling involved by the portfolio analyses are difficult and may require a non-parametric fitting of the multivariate distribution.

We may use statistical factor model for quantifying the portfolio exposures to risk factors. Our drive is to map the risk factors in equity portfolio and to decompose their contribution to the portfolio risk.

In order to quantify the portfolio risk, we use the following risk measure: standard deviation (SD) and Expected Shortfall also known as conditional Value at Risk (cVaR) or Expected Tail Loss (ETL).

The risk measures (RM) are defined as:

Active risk (SD) is derived from the variance of the portfolio

$$SD = \sqrt{\beta_i' \Sigma_F \beta_j + \sigma_\varepsilon^2} \tag{7}$$

Value at Risk (VaR)

$$VaR_\alpha = F^{-1}(\alpha), where\ F\ is\ the\ c.df.of\ R_t \tag{8}$$

Expected Shortfall (ES)

$$ES_\alpha = E[R_t | R_t \leq VaR_\alpha] \quad (9)$$

The risk decomposition is performed by highlighting the contribution of each risk factor and the contribution of constituent assets to portfolio risk. Euler formula shows that if R(w) is the scalar risk measure associated with allocation and if the risk measure is homogeneous of degree 1 that is $RM(\lambda w) = \lambda RM(w)$ then by differentiating the homogeneity condition with respect to λ we get:

$$\sum_{i=1}^{N} w_i \frac{\partial RM(\lambda w)}{\partial w_i} = RM(w) \quad (10)$$

and deduce the Euler formula by setting $\lambda = 1$

Reverting to the factor model, we get additive decomposition by using Euler theorem

$$RM_p(w) = \sum_{j=1}^{k+1} w_i \frac{\partial RM(w)}{\partial w_i} \; where\ RM\ is\ SD, VaR_\alpha, ES_\alpha \quad (11)$$

where *w* are the portfolio weights and RM denotes a portfolio risk measure that is a homogenous function of degree one in the portfolio weight vector. RM may be standard deviation, Value at Risk or Expected Shortfall.

Since portfolio volatility is a linear homogenous function of portfolio weights, the portfolio volatility may be re-written as the weighted sum of marginal risk contribution.

By taking into account the weights of the asset in the composition of the portfolio, the contributions of the individual asset to the portfolio risk are quantified as:

MCRs: Asset *i* marginal contribution to portfolio risk:

$$\frac{\partial RM(w)}{\partial w_i} \quad (12)$$

CRs: Asset *i* contribution to portfolio risk:

$$w_i \frac{\partial RM(w)}{\partial w_i} \quad (13)$$

The asset *i* contribution to risk are weighted marginal contributions.

PCRs: Asset *i* percent contribution to portfolio risk:

$$w_i \frac{\partial RM(w)}{\partial w_i} \Big/ RM(w) \quad (14)$$

The asset percent contributions (PCR) to portfolio risk measure are the contributions to risk divided by the risk measure (RM).

Each component contribution to risk of asset *i* represents the amount of risk contributed to the total risk by investing a certain weight (w_i) in asset *i*. The sum of all contributions equals the total risk. By rescaling the contribution to risk of asset I, we get the percentage of the total risk which is contributed by asset *i*. The sum of asset *i* percent contribution to portfolio risk sum up to 1 (100%).

The marginal contribution to risk of asset *i* represents the marginal impact in the total portfolio risk which comes from a small change in the weight attributed to asset *i*. If the sign of the marginal risk is negative, then by increasing the position size of the asset we increase the total portfolio risk and vice versa.

An incremental change in the allocation of asset *i* is offset by a corresponding change in the allocation of asset *j* such as $\Delta w_i = -\Delta w_j$.

Therefore the change in portfolio volatility is about:

$$\Delta\sigma_{porfolio} = (MCR_i^{\sigma} - MCR_j^{\sigma})\Delta w_i \quad (15)$$

Scaillet (2002) and Meucci (2007) showed that we if assume a multivariate Gaussian distribution for the financial returns, then the partial derivative that give the assets' contribution to risk can be computed analytically while in non-normal markets their joint distribution can be represented through Monte Carlo simulations.

3. Data

Our dataset includes the component companies of the BET-FI index: Fondul Proprietatea (symbol:FP), SIF1, SIF2, SIF3, SIF4, SIF5. The closing prices for BET-FI index and Fondul Proprietatea (symbol:FP), SIF1, SIF2, SIF3, SIF4, SIF5 were extracted from Bucharest Stock Exchange website from January 2014 to April 2015. Daily returns were calculated from the closing prices according to the formula $R_t = \ln(P_t/P_{t-1})$ where P_t is the daily closing price of the index.

Since the weights of the component companies were often changed during the sample time period although with small differences, we have constructed a proxy portfolio for the BET-FI index using the average weights displayed in Table 1. All the calculations were carried out on the proxy portfolio.

Table 1. BET-FI component' weights

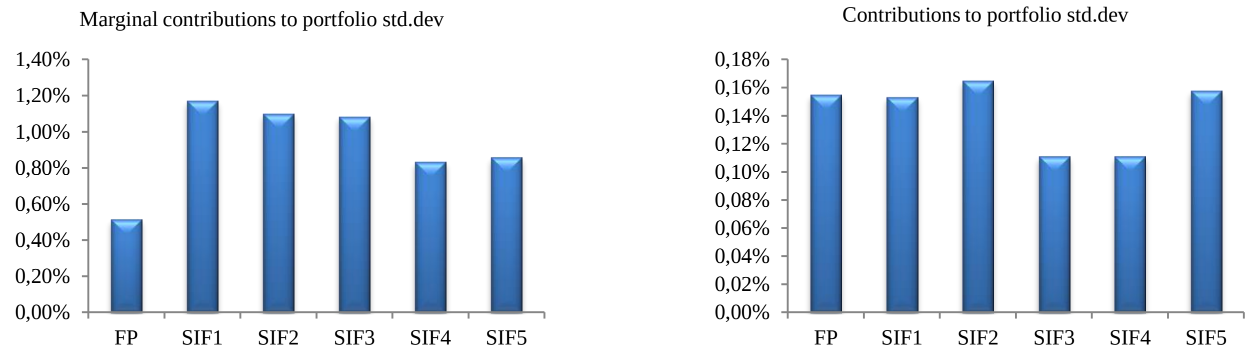
	FP	SIF5	SIF4	SIF2	SIF1	SIF3
14.03.2014	0.2989	0.2145	0.1361	0.1241	0.1182	0.1082
11.06.2014	0.3127	0.1995	0.1257	0.1252	0.119	0.117
13.06.2014	0.2996	0.2016	0.1297	0.129	0.1212	0.1189
01.08.2014	0.2998	0.1995	0.1299	0.1257	0.123	0.1221
12.09.2014	0.2996	0.1947	0.1418	0.1246	0.1236	0.1157
12.12.2014	0.2994	0.1842	0.1501	0.133	0.1309	0.1024
13.03.2015	0.2989	0.1789	0.1492	0.1388	0.1276	0.1065
Average	0.30127	0.19612	0.1375	0.128629	0.12335	0.11297

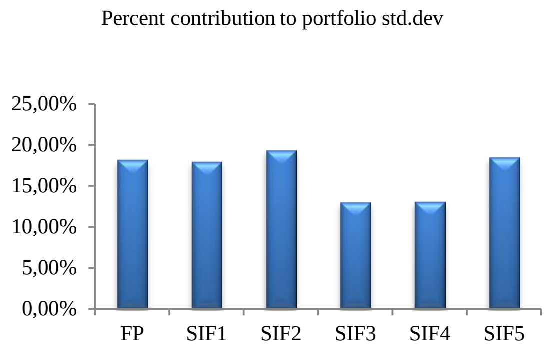
4. Results

Since the marginal contribution to risk of FP is lowest among all the other companies, any incremental change in the weight allocated to FP will decrease the overall portfolio volatility. On the other hand, since the marginal contributions of the other five companies are almost the same, any change in them will increase the volatility in a smaller degree.

Given the high weight of the FP the percent contribution of all six companies to the portfolio volatility is close, ranging from 13% to 18%. SIF1 has the highest marginal contribution although it has a small weight.

Figure 1. Standard deviation decomposition

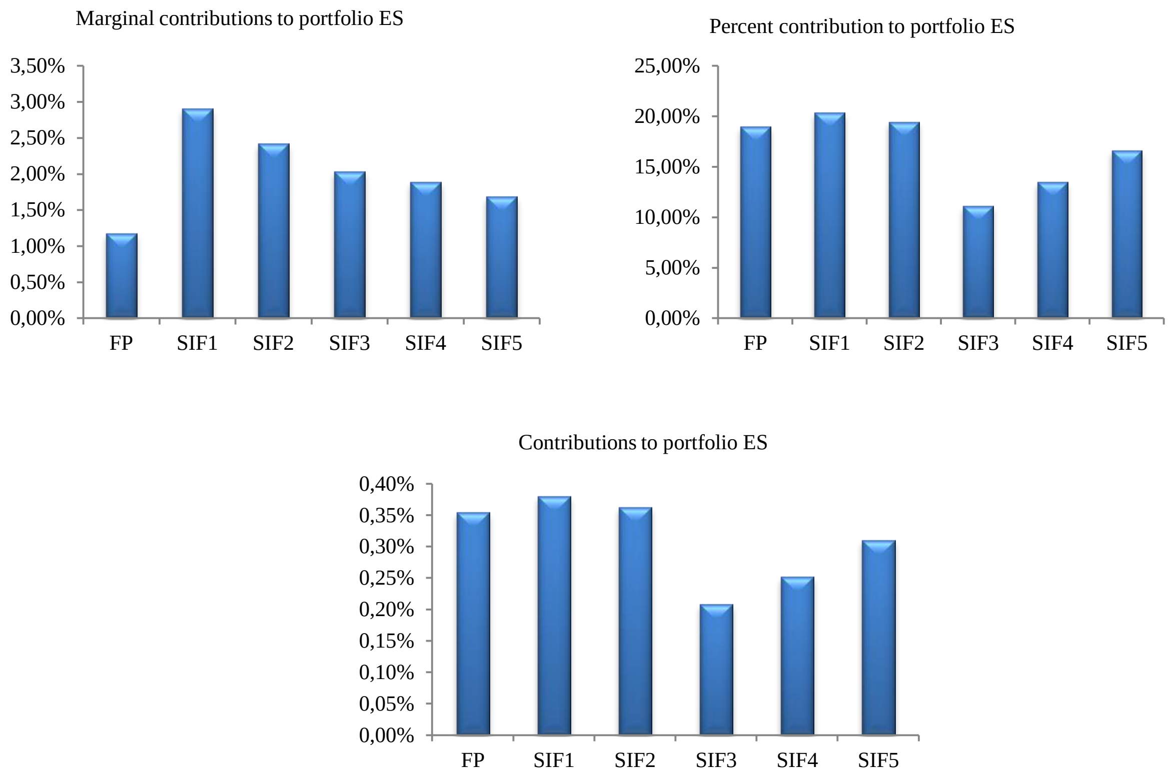




Source: Author's calculations

The VaR of BET-Fi portfolio at 5% is 1.15% and the Expected Shortfall at 5% is 1.86%. Using the Euler formula in the same way as for the portfolio standard deviation, we can compute the MCR, CR and PCR of the contribution of the constituent companies of the BET-FI index to the Expected Shortfall. The results show that in the case of PCR, FP, SIF1 and SIF2 contribute with about the same percent, around 20%. SIF3 and SIF4 have the lowest contribution with 11.15% and 13.5% and SIF4 contributes with 16.6%. The marginal contributions to portfolio ES is similar to the marginal contributions to standard deviation.

Figure 2. Expected Shortfall decomposition



Source: Author's calculations

5. Conclusion

Risk attribution is a method of decomposing the portfolio risk and attributing the return of a portfolio to its risk factors. Thus it is possible to calculate the assets' return contribution to the portfolio standard deviation/Expected Shortfall.

If the chosen risk measure is a homogenous function of degree one in the portfolio weight vector, then we can apply the Euler formula to decompose it into individual contributions.

We have applied Euler formula in order to decompose the standard deviation and the Expected Shortfall of the BET-FI equity index into individual risk contribution.

Our dataset included the component companies of the BET-FI index: Fondul Proprietatea (symbol:FP), SIF1, SIF2, SIF3, SIF4, SIF5. We have constructed a proxy portfolio for the BET-FI index for taking by taking into account the average weight of the companies included in the index portfolio.

The contributions of the individual asset to the portfolio risk are quantified as asset *i* marginal contribution to portfolio risk (MCR), asset *i* contribution to portfolio risk (CR), asset *i* percent contribution to portfolio risk (PCR).

The results showed that in average the marginal contribution to risk of FP was the lowest among all the other companies, meaning that an incremental change in the weight allocated to FP will decrease the volatility of the BET-FI index. The marginal contributions of the other five companies (SIFs) were similar implying that any change in their allocation will increase the portfolio volatility in a smaller degree and smaller degree.

The percent contribution of all six companies to the portfolio volatility is close, ranging from 13% to 18%. SIF1 has the highest marginal contribution although it has a small weight.

We computed the MCR, CR and PCR of the contribution of the constituent companies of the BET-FI index to the Expected Shortfall. FP, SIF1 and SIF2 contributed with about the same percent (20%), SIF3 and SIF4 had the lowest contribution (11.15% and 13.5%) and SIF4 contributed with 16.6%.

By using the risk budgeting framework it is possible to decompose the risk measure(s) calculated on a portfolio and compare the performance of any portfolio in terms of risk contributions, performance and budget allocations.

Acknowledgement

This work was financially supported through the project "Routes of academic excellence in doctoral and post-doctoral research - READ" co-financed through the European Social Fund, by Sectoral Operational Programme Human Resources Development 2007-2013, contract no POSDRU/159/1.5/S/137926.

References

[1] Darolles, S., Gouriéroux, C., & Jay, E, “Portfolio allocation with budget and risk contributions restrictions” CREST DP, 2012.

[2] Hallerback, W.G.,*”Decomposing Portfolio Value-at-Risk: A General Analysis”,* The Journal of Risk, 5(2), 1-18, 2003

[3] Meucci, A., “*Risk contributions from generic user-defined factors”,* RISK-LONDON-RISK MAGAZINE LIMITED-, 20(6), 84, 2007.

[4] Roll, R., “*A Critique of the Asset Pricing Theory.s Tests: Part I”,* Journal of Financial Economics, 4, 129-176,1977.

[5] Scaillet, O., ”Nonparametric estimation and sensitivity analysis of expected shortfall”,. Mathematical Finance, vol. 14, 74-86, 2002.

[6] Scherer, B., Martin, R. D., “*Modern Portfolio Optimization with NuOPT™, S-PLUS®, and S+ Bayes™*”, Springer Science & Business Media, 2005.

[7] Yamai, Y., & Yoshiba, T., “Comparative analyses of expected shortfall and value-at-risk: their estimation error, decomposition, and optimization”, Monetary and economic studies, 20(1), 87-121, 2002.

[8] Zhang, Y., Rachev, S., *“Risk Attribution and Portfolio Performance Measurement-An Overview*”, University of Karlsruhe, D-76128 Karlsruhe, Germany and Department of Statistics and Applied Probability University of California, Santa Barbara, CA93106, USA, 2004.

Human resources in the economic crisis

Carmen RADU[1], Liviu RADU[2]
[1] University Lecturer, PhD - Nicolae Titulescu University, Bucharest, Romania
[2] University Lecturer, PhD - Nicolae Titulescu University, Bucharest, Romania

Abstract

This paper is meant to be an extension of our studies published over recent years, which were meant to seek some answers regarding the existing cause-effect relationship between the economic, financial, demograph and food crisis. In the said article we place the human resource, in its sense of labor force and demograph potential as well, in the middle of the economic, financial, demography and even the food crisis. Provided that in the previous case we demonstrated the hypothesis according to which a food crisis can be caused as well as lead to the migration of the active population to other countries (especially from the rural area) and the agglomeration of underprivileged population in certain geographic areas, we are currently resuming to the mutations recorded by the human resource, as part of the active population, under the aspect of social and economic disequilibrium.

Keywords: *human resources, total population, active population, usually active population, unemployed persons, public policies, public expenses.*

1. Introduction

In 19th century England, the renowned mathematician and demography specialist, Thomas Robert Malthus, made some predictions regarding the intensification of the economic crisis caused by overpopulation. In his paper, "An Essay on the Principle of Population" from 1798, Malthus foresees mankind's collapse through the fact that "the population's force surpasses so much the earth's capacity to insure the resources required for man's sustenance, that, (…) our vices are active and able agents of depopulation[1]. If the vices are not sufficient, than diseases, wars or extreme meteorological conditions arise and ultimately lead to hunger, which will put the equality sign between the ever-evolving population and the diminishing resources. Whilst for that matter, we can also find in the previous article entitled "Food Crisis Overlapping the Economic Crisis" that we are presently facing o lack of food supplies due to the excessive consumption imposed by the multitude of economic agents.

The population's law specifies the hypothesis according to which as long as an income improvement exists, as an effect of the economic development, the population tends to increase in a geometrical progression[2]. Simultaneously, the food supplies offer is increasing more slowly, in an arithmetic progression. The consequence is inevitably the incomes' decrease and the domination of generalized poverty. Unfortunately, Malthus' theory has been confirmed by some developing countries which in the post-war have assisted to causing the "demograph explosion" simultaneously with the decrease of incomes per inhabitant. The fact that an increasingly acute lack of specialized labor force emerges in the given circumstances of an increasing population worldwide, in situations of economic recession, is paradoxical.

From the beginning of history humanity required 2 thousand years to reach the number of one billion people. Afterwards, only one hundred years were needed to double the number, thus arriving at two billion people in the year 1920. Following this interval only 50 years were required to double it again and reach four billion, and according to statistics we will be nine billion people in the year 2050.

Table 1. Demograph growth and growth rate recorded in the 1950-2011 interval and the tendency for 2050

Year	Population (mil. inhab.)	Annual medium rhythm of growth (%)
1750	600-900	0,4
1820	1000	0,5
1927	2000	0,65

[1] R. Malthus (translation by Victor Vasiloiu and Elena Angelescu), *Essay on the principle of population,* Bucharest, Scientific Publishing House, 1992.
[2] Economy dictionary, 2nd edition, Economic Publishing House, Bucharest, 2001.

1960	3000	1,4
1974	4000	1,9
1987	5000	1,7
1999	6000	1,3
2011	7000	1,2
Estimation: 2050	9500	0,51

Source: *The World Factbook,* July, 2011, Central Intelligence Agency,

*** *World Demograph Estimated and Projections,* United Nations, New York, 2011

The population's growth in a geometrical progression between the years 1900 and 2000 has also drawn the same rhythm for the increase of some indicators such as: carbon dioxide emissions, forests' destruction rate, extinction of some species of plants and animals, consumption of water and paper, fishing and ozone layer's destruction. The indicators are, on one hand, in correlation with the population's growth but with the development of technical progress as well and, on the other hand, with the growing number of motorized vehicles, enhancement of foreign investments or increases on a macroeconomic level.

This period of economic growth was anticipated by researchers to last until 2005, hypothesis confirmed as a matter of fact when it reached a peak in 2006-2007 and that followed in 2008. In the economic growth phase effects are being produced at the level of combining production factors, production structure but also consumption, changes which are concretized in an economic leap. Through the profit expectations generated by the economic growth, investments are increasing, fabrication structures along with professional qualification and management are being modified, simultaneously with the development of the old ones. All these transformations are produced until a saturation point that marks a structural crisis.

2. Theoretical background

Humanity's evolution has a certain rhythm and constancy: people generate consumption, consumption indicates production, production attracts financial funds and these funds are used for consumption, investments and speculations. Ultimately the speculations inevitably lead to unjustified earnings or massive losses. Translated on a global level, the equation determines that certain countries will win for a short term while others will lose. That is to say the rich will be richer and the poor will have something more than the liberty to die of hunger, as economists Paul and Ronald Wonnacott[3] figuratively estimated. In this context, the emergence of an economic crisis becomes almost predictable, as the Russian economist Nikolai Kondratieff appreciated, but are the people ready to face them? Here is a question for which the answer can be found in our opinion by studying the human resource. We are presently in an era full of knowledge, in which technology, innovations and the complex structure of production factors impose the rethinking of the human resource's place and role, of its behavior and actions.

In time the human resource is physically and especially morally eroded. The moral usage appears due to the faster development of technique and technology in rapport with the evolution of labor resources. If man requires 15, 20 years of continuous study in order to specialize, in the moment when he is ready to work in the field for which he has prepared he will notice he has insufficient knowledge for the job's maximum exploitation. Perhaps this explains the fact that the European Union is currently facing an acute lack of specialists in the information technology. Although it is one of the most required specializations among young people, the domain imposes a highly advanced level of knowledge and abilities, which makes it more useful for those who wish to work from home. Unfortunately, the specialists in the information technology are not the only specialists who are not willing to respect the strict laws of a working place.

This is why the analyses presented in the rapport entitled "An agenda for a Growing Europe" and presented in 2003 by a group of specialists from the European Commission, which was coordinated by Professor André Sapir (European Center Ad.Research in Economics and Statistics) are undoubtedly real. According to the theories presented in 2003 „a system built around the assimilation of existing technologies, the mass production that generates large economies and an industrial structure dominated by large and very large firms with stable markets and long term employment of personnel schemes, is no longer efficient in today's world characterized

[3] Paul Wonnacott, Ronald Wonnacott, Economics, 3rd edition, McGraw – Hill Book Co., 1986, pages 67-68

through economic globalization and powerful external competition. Nowadays we require more opportunities for the market noobs ☺), more mobility inside the firm and between firms for employees, more requalification, a stronger dependency on financial markets and more investments in research development as well as in education. All these ask for an urgent and massive change in Europe's economic politics."

The Sapir Rapport was published in 2003 and the authors came back in 2005 with "Globalization and the reform of European Social Models (ECOFIN Informal Meeting in Manchester)[4]. In 2005 the unemployment rate at the European Union's level recorded a decreased level and maintained itself in a decreasing trend, but with a great distribution difference between the four country categories (Anglo-Saxon, continental, Nordic and Mediterranean).

A series of predictions were made regarding the future of the European Union confronting the wave of adhering, trying to somehow delimitate the Union's old members from the new labor force that arrived with another mentality, training, level of culture and civilization.

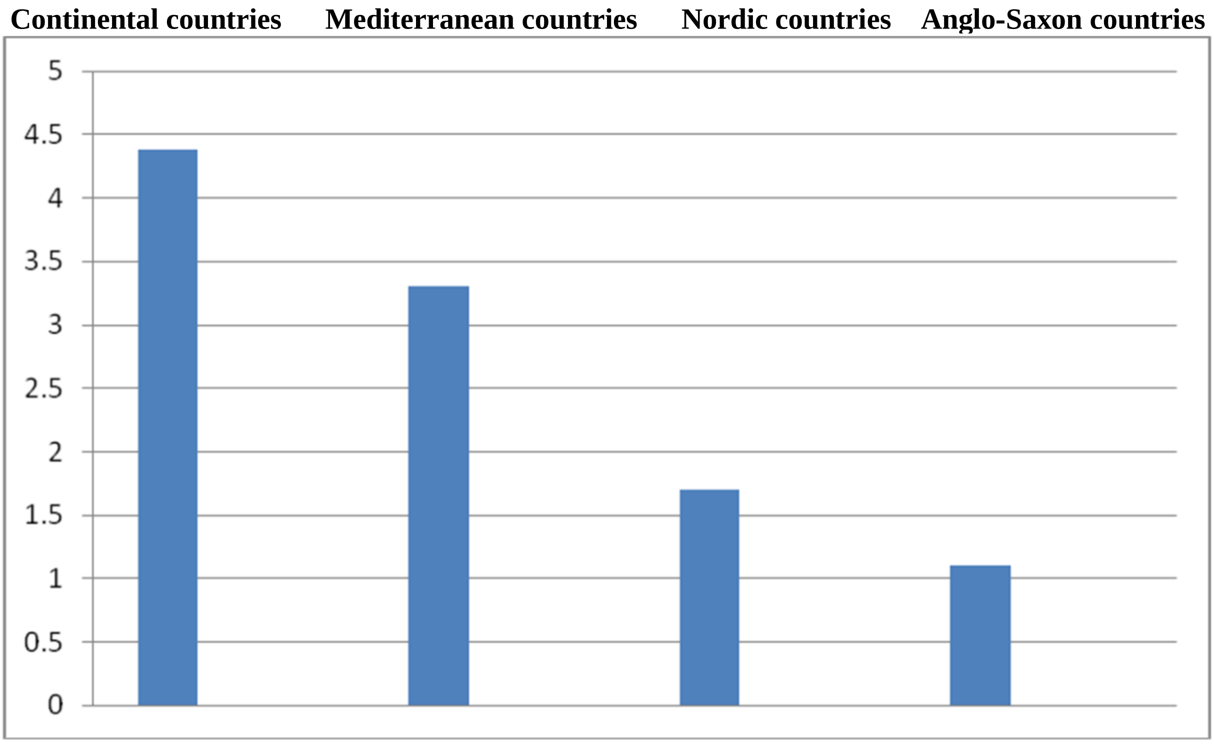


Fig. 1. Long-term unemployment level in different categories of European countries
Source: Eurostat 2005

In the graph published by the Eurostat in 2005 we can notice that the countries with an Anglo-Saxon origin have a long-term level of unemployment (over 12 months) that is little over 1%, while the European Continental countries have a rate of almost 4.5%. The same source also publishes how the poverty reduction level is distributed (measured as a mass of the people with an available income under 60% of the national average income) in the above mentioned countries.

[4] André Sapir, Globalisation and the reform of European Social Models, JCMS 2006 Volume 44. Number 2. pp. 369–90

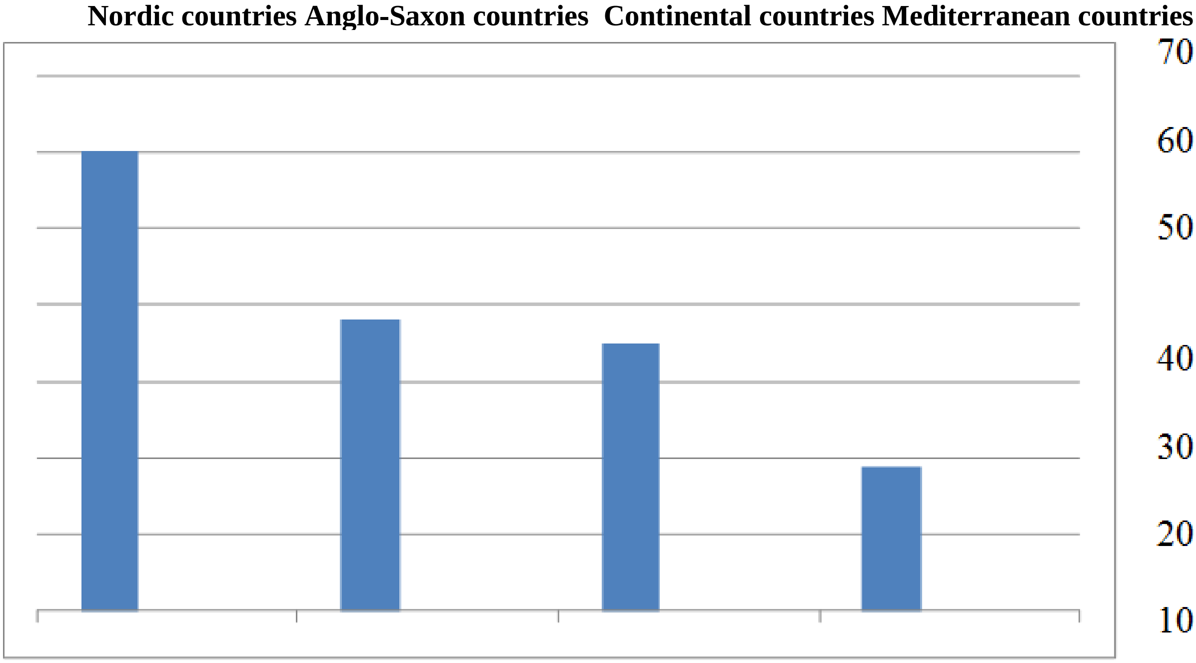


Fig. 2. Poverty reduction level in different categories of European countries (EU-15)
Source: Eurostat 2005

The graph reveals the fact that the richest countries before the beginning of the economic crisis were the Nordic countries, while the poorest were the Mediterranean ones. Otherwise said, between countries such as Sweden or Denmark and Italy or especially Greece, a discrepancy of 50% exists between the available incomes. In addition, studies demonstrate that an indissoluble connection exists between the level of poverty, level of education and mass of expenditures with social protection.

In 2007 the unemployment rate recorded the lowest level, furthermore being a year of economic boom as we have shown. As it can be observed in figure 4, in 2008 the crisis and unemployment suddenly grow.

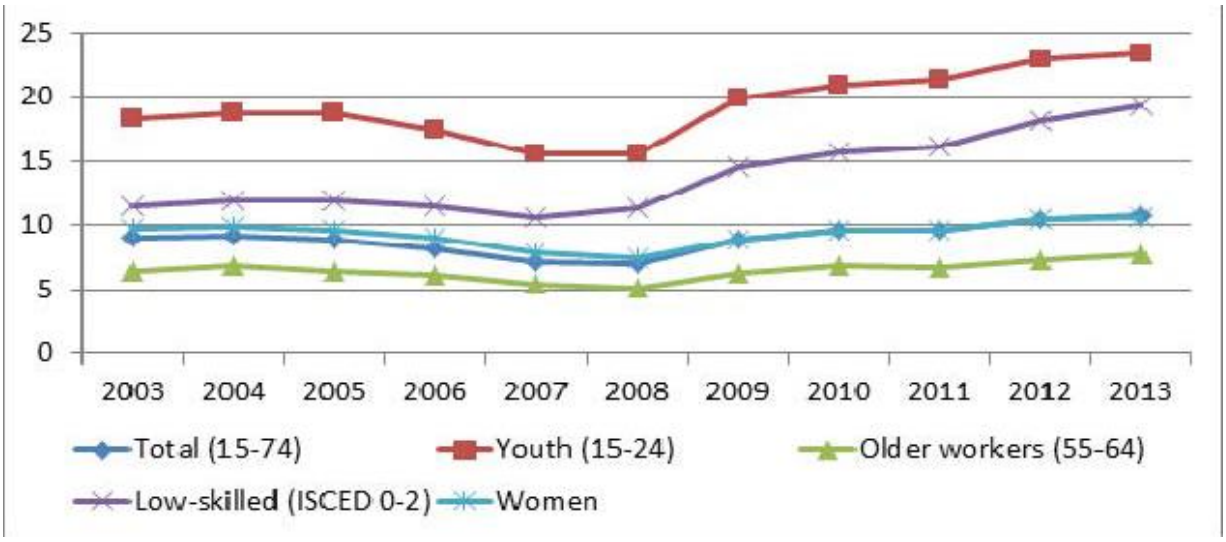


Fig. 3. Unemployment level in the EU by categories of people
Source: "EU Employment and Social Situation, Quarterly Review", March 2013

The graph shows a distribution of the unemployment on the following categories: working population (between 15 and 74 years old), young people (between 15 and 24), elderly (between 55 and 64), population with a low level of qualification and women. A reduced rate of unemployment can be observed among the elderly and women, rate that follows the working total population’s trend. On the other hand the rate of unemployment is emphasized among young people and the population with a low level of qualification.

If the unemployment rate has dropped between 2003 and 2008 with more than two percentage points, we can observe that after the outbreak of the economic crisis the unemployment begins to rapidly grow. Therefore, in five years, from the middle of the year 2008 up to the second trimester of the year 2013, the unemployment rate is growing from 7,1% to 10,9%. The unemployment’s further evolutions along the crisis period have been more or less similar for different categories of people in the labor force market, with a couple of exceptions. First of all, unemployment among young people is much more receptive to the business cycle in general. Furthermore, unemployment among men is higher in the sectors of activities which are dominated by them, as a matter of fact between 2008 and 2009 unemployment rises almost insignificantly among women, remaining under the general level.

As far as structural differences are concerned, young people and workers with a medium level of training have the highest risen rate due to the similarities regarding the specialization degree. At 24 years old a young person finishes the complete studies (bachelor’s and master’s degree) which only imply a theoretical qualification and very few practical skills.

Unemployment stops from rising at the middle of the year 2013 and somewhat remains constant at the beginning of 2014, reaching a number of nearly 27 million people in the European Union. Interesting in this sense is also the distribution of unemployment on different countries members of the union. In comparison to 2012 the unemployment rate has very much increased in Greece, Cyprus, Italy and Holland and has recorded a regression in Ireland and Hungary. These variations are also caused by the GDP’s recorded level in that period. Per ensemble, long term unemployment continues to grow on account of the long duration of economic crisis. At the end of the previous year long term unemployment reached the record of 12.5 million which represents 5% of the active population in the European Union.

If we relate to the year 2008 long term unemployment almost doubled, with the exception of Germany where unemployment rate drops from 4% to 2.5% between 2008 and 2012 and Luxembourg where it maintains relatively constant around 1.5%.

Unemployment presents great divergences especially between countries members of the euro zone[5] as well. From the economic crisis’ debut we can notice a massive rise of unemployment in the countries situated in the south and periphery of the euro zone as opposed to the rest of the member countries. Thus, the variations are between approximately 5% in Austria, Germany and Luxembourg and over 25% in Greece and Spain. Also, the unemployment with rates of over 16% recorded in Portugal, Cyprus and newly joined Croatia, is considered to be above average. As a matter of fact, Cyprus has the largest accession recorded from one year to another (September 2012, September 2013) of over 4.4 percentage points.

Unemployment among young people remains very high, of 23.5% at the EU level but with surprising variations between states, from 7 or 8% in Germany and Austria to 56 or 57% in Spain and Greece.

In spite of the economic crisis we can detect an improvement of the activity rates in several member states. These improvements were possible on account of the increased activity rates among the population with ages between 55 and 64 years old but also of women, which led to a general growth of the activity rates on a European level from 70.7% to 71.9%. There have obviously existed here as well high variations between member states. The highest growth rate recorded the Czech Republic, Malta and Latvia, and at the opposite pole the lowest rate was recorded in Denmark (but from a very high level), Ireland and Croatia. Although women’s rate of employment has continuously improved, a large gap still exists in opposition to the masculine population. The difference is the working schedule, women preferring a part-time schedule (Holland 77,3% or Germany and Austria with over 45%) as opposed to men who normally wish for a full norm. As far as financial wins or loses are concerned, no great differences have existed between the genders. Big differences appear in the labor force’s distribution on branches and sectors of activity. The highest losses are in the constructions domain (-4,5%), agriculture (-1,5%) and processing industry (-1.2%). The highest increases of the occupied population are in the information’s technology domain (+2.5%) but here also without a permanent contract. In this domain the temporary or non-renewable jobs had a priority. As we showed, although attractive from a financial point of

[5] *Quarterly Report on the Euro Area,* European Commission, Volume 12, No. 3, 2013

view, working places in the IT domain are appreciated by young people and women that have other activities in parallel.

Still on the European Union's level, per ensemble, the vacant working places' rate has not recorded great modifications although, as we showed, unemployment has increased pretty much, in different regions and states. This fact can also be explained by the lack of interest that either the vacant jobs present, or their remuneration, either the employing firm's perspectives. Here is why a highlighted increase of qualification and professional training can also mean a lack of attractiveness of the opportunities in the labor force market, and this fact associated with the manifestations prolonged by the economic crisis can induce a state of depression among the available active population. This fact can be countered through serious investments in the human capital.

The economic crisis has produced significant mutations on the population's migration fluxes as well. Thus three levels are differentiated:

1. migration from the third countries to the EU is reduced with 3.7% between 2010 2011;
2. migration from the EU to the third countries is increased by 14% in the same interval. Here we can notice that the majority of those who wish to leave from inside the union come from Spain, England, France, Ireland, Portugal and the Czech Republic;
3. migration of those returning home.

We are thus assisting, in consequence of the economic crisis' effects, to the change of models regarding migration in the European Union. The economic crisis triggers a lack of trust in the opportunities that the EU has left to offer. Numerous individuals categorized as currently active population either choose to return to their country of origin or leave from inside the Union to other countries. Those who wish to leave also originate from countries such as England and France, where the perspectives of overcoming the crisis' effects seem to be farfetched. Unfortunately, recent[6] studies show that in the European Union there is a lack of competitiveness regarding the currently active population's abilities. Truthfully the human resource is increasingly more literate, possesses a large amount of knowledge, but with a deficit of practical experience. Therefore approximately 20 percentages of the population that is able to work do not possess the abilities required to occupy a working place. Unfortunately this percentage is higher in Italy and Spain, as opposed to countries that insist upon learning practical skills such as the Nordic countries. Nonetheless, not even countries such as Finland, Holland or Norway can rise to the level of countries from outside of Europe, such as Japan and Australia. Unfortunately, the dates confirm the fact that Europe was unable of efficiently investing in education and professional competences. It is dramatic that this state of fact continues to worsen through the fact that no less than 10 European states have reduced the expenditures for education in absolute terms (Denmark, Ireland, Greece, Spain, Italy, Cyprus, Hungary, Portugal, Czechoslovakia and the UK) and 20 other members are reducing the mass of education expenditures from the gross domestic product.

[6] DRAFT JOINT EMPLOYMENT REPORT accompanying the Communication from the Commission on Annual Growth Survey 2014

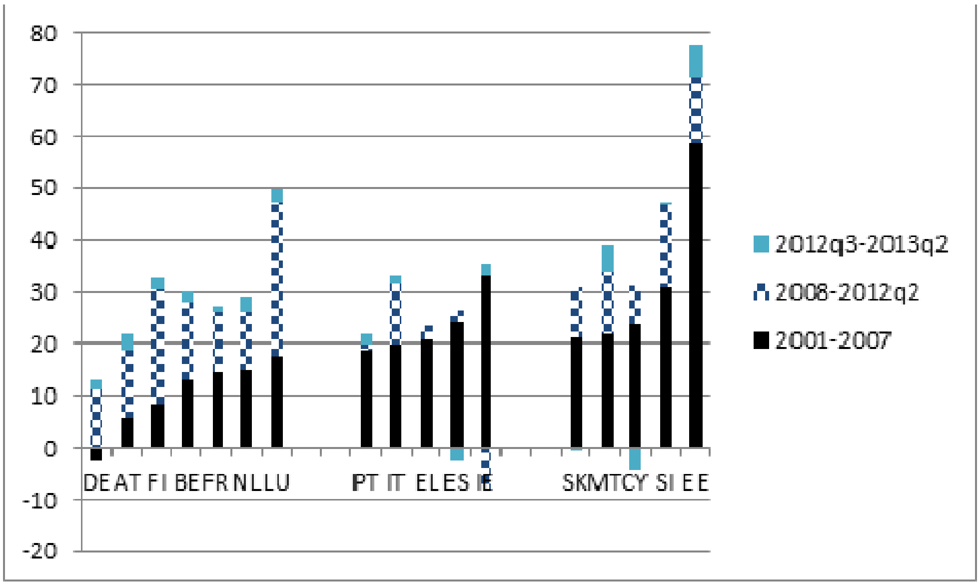


Fig. 4. Countries in the euro zone and education investments
Source: EUROSTAT

A great difference can be observed between Germany, which has reduced expenditures for education between 2001 and 2007 and raised them between 2008 and 2012, on one hand and countries such as Luxembourg, Slovenia and Estonia, whose expenditures have and continue to increase.

Reducing expenditures for education is overlapping in the European Union with the early school leaving phenomenon (ESL), which was of 12.7% in 2012. In this context it is not surprising that average costs and earnings are reduced but the prices for finished products are rising due to the increased indirect taxes and administrative prices. Reducing labor's unit costs as opposed to the sustained growth of prices leads in time to increased profit margins which unfortunately are not accompanied by a rise of investments. At the same time the fiscal burden remains elevated in many member states and implicitly causes the population with low incomes to reach the sustenance limit, in the context in which income taxes have been reduced in the two years following the crisis outbreak.

The economic crisis has substantially modified the inequality between the states members of the European Union. These inequalities are reflected in the slow development of the union's southern countries – Spain, Greece, Italy and Cyprus but also in the discrepancy of distributing the human resources' incomes in countries such as Bulgaria, Latvia and Romania. Furthermore, the access to medical assistance is more difficult to gain in countries with reduced incomes, due to the decreased expenditures for health.

3. Conclusions

As we have already estimated, a system built around the assimilation of existing technologies is no longer efficient in a world dominated by a powerful external competition. Presently, the globalization phenomenon leads to the economy being managed by large and very large firms, with traditional and stable market outlets and with human resources that are both prepared and loyal to the company. Nowadays we require more opportunities for the investors and initiative entrepreneurs, a higher mobility inside the firm and between firms for employees (based on the model of switching the favored personnel especially by French companies). But on top of everything we require more requalification through branch operational programs and more investments in research development and education as well. All these require an urgent and massive change in Europe's economic politics and especially in Romania.

The low income countries inside the European Union will feel the economic crisis' effects more strongly. The population's growth in these countries will determine a powerful unemployment increase on a European level.

Very large companies such as Microsoft, Webhelp, Evoline, Google, TeamNet etc. create thousands of working places in countries marked by an almost record level of unemployment, but then again the number of qualified people for those certain jobs is insufficient. The disparity between the candidates' preparation and the type of qualification required by the employer is increasingly worse on the economic crisis background and is changing into one of Europe's most important problems. One of the solutions found by companies is to employ qualified personnel from outside the said country.

In this context, people who have lost their jobs as well as young people in search for a starting point in their career are discovering they do not possess the necessary training for the existing working places. Glenda Quintini, economist from the OECD, affirms that "we are assisting to an alarming lack of adequate aptitudes, which means that a significant number of unemployed people are not ready for the new working places." In consequence, the majority of young people need to try, in this case, to reorient from a professional point of view towards other domains.

Paradoxical is the fact that more than 2 million jobs remain unoccupied in the conditions in which Europe is facing unemployment rates that are maintained at over 12%. A study[7] published in November 2013 by the Eurofound, the EU's research division, indicates that in spite of the recession, approximately 40% of the companies are facing difficulties in finding employees with suitable aptitudes, in comparison to 37% in 2008 and 35% in 2005. The European Commission has recently warned that approximately 900.000 working places in the IT and communications technology domains may remain unoccupied until 2015 in the EU[8].

The unemployment rate in Romania at the end of last year (2013) was of 5,60%, namely 507607 people, with 0.20 more percentage points in rapport to the same period of the year 2012. As the genre repartition is concerned, the male unemployment rate is of 6.01% and the female unemployment rate has risen from 4.98% to 5.14%. The number of unemployed women at the end of last year was of 219288 people. Represented by age groups, 93639 unemployed people were under 25 years old, 37504 people had the ages between 25 and 30, 109.124 were of ages between 30 and 40 and 133.392 were included in the 40-50 years old category. At the same time, 60.489 unemployed people were of ages between 50 and 55 and 73.459 surpassed 55 years old. From the total of unemployed people recorded in the ANOFM (National Agency of Occupying the Labor Force), those with no studies or only with a primary level of training, grade school and technical school, represented 69.02%, those with a high-school and post high-school education level – 23.75%, and the ones with university studies – 7.23%.

The European Commission has already allocated Romania funds of approximately 100 million Euros through the Youth Employment Initiative program with the purpose of combating unemployment especially among young people. At the same time, our country needs to allocate the European Social Fund at least 30.8%, which implies 4.77 billion Euros from the cohesion funds, in the 2014-2020 budgetary practice. The money from the European Social Fund is allocated for programs dedicated to increasing the occupation degree in the member states.

4. References

[1] R. Malthus (1992), translation by Victor Vasiloiu and Elena Angelescu). *Essay on the principle of population*, Bucharest, Scientific Publishing House.

[2] André Sapir, *Globalisation and the reform of European Social Models*, JCMS 2006 Volume 44. Number 2. pp. 369–90.

[3] Paul Wonnacott, Ronald Wonnacott (1986), *Economics*, 3rd edition, McGraw – Hill Book Co.

[4] Summit of the Millenium, "*The role of the United Nations Organization (UNO) in the XXIst century*", The Millennium declaration adopted by the United Nation's General Assembly. World Population Ageing 1950-2050"- United Nations –Economic and Social Affairs, New York 2002.

[5] United Nations Development Programme (September 2005). *Human Development Report 2005: International Cooperation at a Crossroads-Aid, Trade and Security in an Unequal World*. United Nations Development Programme. ISBN 978-0-19-530511-1.

[6] *Population Division of the Department of Economic and Social Affairs of the United Nations Secretariat (2005).* World Population Prospects: The 2004 Revision, Highlights. New York, United Nationation, www.un.org.

[7] The World Factbook, July, 2011, Central Inteligence Agency, *World Demograph Estimated and Projections*, United Nations, New York, 2011.

[8] ***"EU Employment and Social Situation, Quarterly Review", March 2013.

[9] DRAFT JOINT EMPLOYMENT REPORT accompanying the Communication from the Commission on Annual Growth Survey 2014Quarterly Report on the Euro Area, European Commission, Volume 12, No. 3, 2013.

[10] *** "World Population Prospects : The 2000 Revision ", United Nations, New York, 2002.

[11] *** *Economy dictionary*, 2nd edition, Economic Publishing House, Bucharest, 2001.

[12] *** Eurostat.

[7] Quality of employment conditions and employment relations in Europe.

[8] http://www.eurofound.europa.eu/pubdocs/2013/67/en/1/EF1367EN.pdf

The relationship between tourism and economic growth in greece economy: a time series analysis

Turgut Bayramoğlu[1*]
Yılmaz Onur Arı[2**]
[1]Assistant Professor Doctor Bayburt University, Economics Department
[2]Research Assistant Bayburt University, Economics Department

Summary

In this study it's analyzed how and in what way the expenditures of foreign visitors who came to Greece between 1980 and 2013 affected economic growth for Greece. For this purpose Granger Causality Test was used, the results of unit root tests such as Augmented Dickey Fuller (ADF) and the Philips Perron (PP) were tested and because it's a time series analysis, unit root and co-integraion tests were applied. At this point lag coefficient was obtained using by Akaike Information Criteria (AIC). Using five different criteria, it was confirmed that the best suited lag period is 2. GDP and tourism data were obtained from the World Bank Statistical Data. The result showed that there was a strong unidirectional causality from the expenditures of foreign tourists who visited Greece to the growth of Greece at 1 % level of significance.

1. Introduction

As accepted by World Travel and Tourism Council, tourism has been the most rapidly growing sector. Tourism shows that it is one of the most important sectors by providing employment to 255 million people and supporting 6 billion dollars, which is 9 % of the total revenue of the world (Chou, 2013: 226). Greece, a Mediterreanean country, reached 13 billion dollar revenue with 17 million foreign visitors annually (UNWTO, 2014: 8). Because tourism makes a great contribution to economy, it is considered as an important tool for both growth and development. On the other hand, tourism has crucial functions such as eliminating external deficit, obtaining finance, creating new employment opportunities and reducing unemployment rate (Yavuz, 2006: 162; Çoban ve Özcan. 2013:244).

Although tourism has some differences from country to country, socio- economic contribution for national economies can be classified as follows (Pao, 2004: 81; Akan ve Işık, 2009: 198);

- Balance of payments: can be seen as a main source of foreign capital inflows.
- Regional development: provides for spreading economic activities into the country.
- Variety of economics: makes a contribution to economy by affecting different areas.
- Income level: provides many people income opportunities.
- Job opportunities: provides employment especially for areas where unskilled labor force is available.
- Government revenues: assures funds for certain expenditures.

Tourism is an important sector for Greece, which has been struggling with economic crisis. According to Greek Tourism Business Association's data (SETE), it's expected that tourism revenues of Greece will reach 13 billion euros in 2015. Greece ranks at 17 in the world in tourism revenues gained. According to data of 2012, tourism's contribution to employment was 18.3 % of the total employment in Greece, and it's contribution to GDP is approximately 16.4 %. This data shows that tourism is one of the competitive sectors in global extent for Greek economy.

2. Literature study

In this part of the study previous studies about how the expenditure of foreign visitors affected economic growth were presented with a table. In literature the effect of tourism on economic growth was analyzed by using methods like Granger Causality Test, Toda-Yamamato, Error-Correction Model and Panel Data Analysis, and so

[*] Corresponding author: tbayramoglu@bayburt.edu.tr
[**] Corresponding author: yoari@bayburt.edu.tr

the direction of this relationship between tourism and economic growth was determined. It was also shown that the direction between variables were unidirectional or bidirectional.

Despite tourism has a residual importance, the development of tourism sector related to economic growth hasn't been searched yet in economic literature. In the light of previous studies; if there is a relationship between tourism and economic growth, it affects both the economy and tourism positively.

Table 1. Literature Summary

Sample	*Authors*	*Method*	*Period*	*Countries*	*Causality Relationship*
One Country	Dritsakis (2004)	Error Correction Model	1960-2000	Greece	Tourism ⇔ Growth
	Oh (2005)	Granger Causality Test	1975-2001	Korea	Growth ⇒ Tourism
	Özdemir and Öksüzler (2006)	Granger Causality Test	1963-2003	Turkey	Tourism ⇒ Growth
	Yavuz (2006)	Granger Causality Test	1992-2004	Turkey	None
	Vanegas et al. (2007)	Granger Causality Test	1980-2005	Nikaragua	Tourism ⇒ Growth
	Kızılgöl and Erbaykal (2008)	Toda- Yamamato Causality Test	1992-2006	Turkey	Growth ⇒ Tourism
	Akan and Işık (2009)	Granger Causality Test Johansen Cointegration Test	1970-2007	Turkey	Tourism ⇒ Growth
	Brida et al. (2010)	Granger Causality Test	1980-2006	İtaly	Tourism ⇒ Growth
	Kapiki (2011)	Field Research	2010	Greece	-
	Polat and Günay (2012)	Error Correction Model	1969-2009	Turkey	Tourism ⇒ Growth
	Çoban and Özcan (2013)	Johansen Cointegration Method	1963-2010	Turkey	
More Than One Country	Gökovalı and Bahar (2006)	Panel Data Analysis	1987-2002	Mediterreanean Countries	Tourism ⇒ Growth
	Holzner (2011)	Panel Data Analysis	1970-2007	134 countries	Tourism ⇒ Growth
	Chou (2013)	Panel Data Analysis	1988-2011	10 Transition Countries	Causality in 7 countries

Some of the previous studies show that tourism affects economic growth unidirectionally. Özdemir and Öksüzler (2006), explained that tourism affected economic growth in Turkey between 1963 and 2003 by using Granger Causality Test. Similiarly Brida et al. (2010), in their study in İtaly's Trentino Alto-Adige Region proved that tourism affected economic growth with Granger Causality Test. In a study which error-correction model was applied, Polat and Günay (2012), concluded that tourism affected the economic growth in Turkey between 1969-2009 unidirectionally.

There are also some views which rejects that tourism causes economic growth in the literature. Oh (2005), analyzed the relationship between tourism and economic growth in Korea by using Engle and Granger and bivariete VAR Approach. As a result of the study it's revealed that there wasn't a causality between tourism revenues and economic growth. Contrary to other studies, it's claimed that tourism was a reason of economic growth in this study. It's clearly understood that tourism had no impact on Korean economy. Similiarly, Kızılgöl and Erbaykal (2008), argued that economic growth had an impact on tourism in Turkey between 1992 and 2006, by using Toda Yamamato Causality Test, therefore they claimed that tourism was the result of economic growth.

Dritsakis (2004), analyzed the impact of tourism on Greek economic growth in the long run using Johansen Cointegration and ECM tests with 1960-2000 data. As a result of the study it's claimed that there was a bidirectional relationship between international tourism and economic growth in the long run. For another example about Greek tourism, Kapiki (2011) studied the impact of economic crisis on tourism and hospitality industry. As a result of his survey study, due to the financial crisis in Greece, profits of hotels in Greece has lowered and therefore operating costs such as electricity and wages has increased. Despite the crisis, number of accomodation in 5-star hotels has decreased only about 1.7 % in 2009-2010 when compared with the before-crisis period.

When it comes to multiple country applications, Chou (2013), used panal data analysis on 10 transition countries, and claimed that tourism didn't affect economic growth only in 3 transition countries (Bulgaria, Romania and Slovenia). Tourism was one of the main reasons of economic growth in other 7 countries, according to the results of the study. Gökovalı and Bahar (2006), concluded that tourism was a triggering factor of economic growth for Mediterranean countries between 1987 and 2002, as a result of panel data analysis. Holzner (2011), who analyzed the Dutch Disease Effect[17] on countries which were dependent upon tourism, concluded that tourism had positive impact on countries' output levels and Dutch Disease couldn't jeopardize in countries which were dependent upon tourism in the long run, according to his panel data analysis.

3. Econometric method and empirical results

In this study, we investigate the interaction between variables, which are Economic Growth of Greece and a variable of foreign visitor's expenditure by employing Granger Causality Test. The study is based upon time series data between the years 1980 to 2013. On the grounds that the time series data has been used in this study, stationary and cointegration tests were implemented. At this point, the lag criterion was obtained by using 'Akaike Information Criteria' (AIC).

In this paper, variables of Economic Growth and Expenditure of Foreign Visitors were inclusive of the periods of 1980 to 2013. We procured the data of Economic Growth of Greece and Tourism from The World Bank. To analyze the study, we made use of E –Views 8 econometric software.

Table 2. Results of ADF and PP Unit Root Test

	ADF UNIT ROOT TEST			PHİLİP PERRON		
Variables	Constant adf t statistics	Constant and trend adf t statistics	Without constant and trend adf t statistics	Constant pp t statistics	Constant and trend pp t statistics	Without constant and trend pp t statistics
lngdp	-1.411638	-2.367683	1.283650	-0.897125	-1.903325	1.698096
lntourism	-0.329243	-2.429080	1.852193	-0.224590	-2.571546	1.883251
dlngdp	-3.452813**	-3.499312*	-3.048711***	-3.406145**	-3.271036*	-3.016161***
dlntourism	-4.884152***	-4.806298***	-4.313786***	-4.816250***	-4.719211***	-4.324275***

* statistically significant at 0.10 significance level
** statistically significant at 0.05 significance level
*** statistically significant at 0.01 significance level

In this analysis, economic growth variable and expenditures of foreign visitors were indicated "GDP", and "TOURISM", respectively. ADF and PP Unit Root Test Results was shown as constant, constant and trend, and without constant and trend at Table-2. For both two variables, unit root test results in their levels showed that variables were not stationary, in other words these series contained unit root.The non-stationary series were tested again by taking their first difference in order to make them stationary. In this respect, both of the series were again subjected to the stationary test by taking the difference. TOURISM as series which were taken difference was stationary at 1% significance level after ADF and PP statistics tests which are constant, constant and trend, and non-constant and non-trend were employed. In addition, GDP was statistically significant at 5 % level in constant, at 10 % level in constant and trend, and at 1 % level in non-constant and non-trend ADF and PP tests.

[17] Dutch Disease; can explain as a decrease in production as a result of a country change their production factors to a new resource, which provides a sudden and high wealth level. For the first time in history it occured in Holland in 1960s, due to rich natural gas resources. That's why it's called '*'Dutch Disease''*.

Table 3. Determining Lag Length Upon VAR Model

Lag	LogL	LR	FPE	AIC	SC	HQ
0	-26.60317	NA	0.026447	2.043083	2.138241	2.072174
1	42.59717	123.5720	0.000251	-2.614083	-2.328611	-2.526812
2	**56.60962**	**23.02046***	**0.000124***	**-3.329259***	**-2.853471***	**-3.183806***
3	57.34024	1.095929	0.000158	-3.095731	-2.429629	-2.892097
4	57.89870	0.757904	0.000208	-2.849907	-1.993490	-2.588092
5	65.25967	8.938320	0.000171	-3.089976	-2.043244	-2.769980

In the first difference of the series which are at stationary state must have the proper lag length for future analysis. In Table-3, LR (Likelihood), FPE (Final Prediction Error), AIC (Akaike Information Criterion), SC (Schwarz Information Criterion), HQ (Hannan- Quinn Information Criterion) were investigate to find the most proper lag length. According to this, we estimated 2 (as a value) which is the most appropriate lag length. Therefore, the estimated value “2” will be used as a lag length in the analysis.

Table 4. Engle-Granger Cointegration Test Results

Cointegration Equation	Lag Length	ADF Statistics	McKinnon Criteria Value	
			1 %	5 %
Gdp=f (tourism)	2	-3.648962	-3.661661	-2.960411
Tourism=f(gdp)	2	-3. 311773	-3.661661	-2.960411

In Table 4, we estimated cointegration between GDP and TOURISM by employing Engle-Granger Cointegration Method. According to Engle-Granger Cointegration test results, null hyphothesis can’t be rejected, which states a long-term relationship between variables.

Table 5. Granger Causality Test Results

Number of lags	The Direction of Causality	Wald Test	Probability Value
5	TOURISM $\longrightarrow$ GDP	22.16014	0.0005***
5	GDP $\longrightarrow$ TOURISM	2.142391	0.8291

***significance level at 0.01

Error correction model test results indicate that there is a causality to GDP from TOURISM. That is to state that TOURISM variable in the equation which GDP is the dependent variable was statistically significant at the 1% level shows that the cause of the GDP. This situation explained that there was a strong, long and unidirectional causality relationship from the expenditures of foreign tourists in Greece to GDP of Greece.

4. Conclusion

Tourism is one of the most important sectors for Greece. Tourism is both the engine of economic growth and effective for other economic areas. Tourism provides 255 million people employment opportunities and 6 trillion dollar to world’s total revenue. Tourism affects Greek economy in an important way with it’s annual 17 million foreign tourists. On the other hand tourism has many important functions such as creating new job opportunities, reducing the unemployment rate, providing funds to country, decreasing the balance of payments deficit.

In this study, the relationship between tourism and economic growth was tested with time series analysis for Greek economy, which has been coping with crisis and considering tourism as a solution to escape from the negative effects of it. Greek economy has been facing with problems such as budget deficit, high public debt, low competitive capacity in the market and underinvestment by foreign investors. For these reasons, global economic crisis which occured firstly in international financial market and involved real economy, has been affecting Greece severely. Negative situation in other sectors increased the importance of rise and fall in the number of foreign visitors. So this paper will lead similiar economies like Greece in terms of the increasing importance of tourism.

5. References

[1] Akan, Yusuf ve Işık, Cem (2009). Yabancı Ziyaretçi Harcamalarının Ekonomik Büyümeye Etkisi (1970-2007), *Anatolia: Turizm Araştırmaları Dergisi*, 20(2), 197-203.

[2] Association of Greek Turizm Enterprises (SETE), 2015 http://old.sete.gr/EN/TOURISM/The%20Importance%20of%20 Tourism%20for%20Greece/The%20importance%20of%20Tourism%202012/

[3] Brida, Juan Gabriel ve diğerleri (2010). Causality Between Economic Growth and Tourism Expansion: Empirical Evidence From Trentino-Alto Adige, *Tourismos: An International Multidisciplinary Journal of Tourism*, 5(2), 87-98.

[4] Chou, Ming (2013). Does Tourism Development Promote Economic Growth in Transition Countries? *A Panel Data Analysis, Economic Modelling*, 33, 226-232.

[5] Çoban, Orhan ve Özcan, Ceyhun Can (2013). Türkiye'de Turizm Gelirleri-Ekonomik Büyüme İlişkisi: Nedensellik Analizi (1963-2010), *Eskişehir Osmangazi Üniversitesi İİBF Dergisi*, 8(1), 243-261.

[6] Dritsakis, Nikolaos (2004). Tourism as a Long-Run Economic Growth Factor: An Empirical Investigation for Greece Using Causality Analysis, *Tourism Economics*, 10(3), 305-316.

[7] Gökovalı, Ummahan ve Bahar, Ozan (2006). Contribution of Tourism to Economic Growth: A Panel Data Approach, *Anatolia: An International Journal of Tourism and Hospitality Research*, 17(2), 5-13.

[8] Holzner, Mario (2011). Tourism and Economic Development: The Beach Disease?, *Tourism Management*, 32, 922-933.

[9] Kapiki, Tania (2011), The Impact of Economic Crisis on Tourism and Hospitality: Results From a Study in Greece, *Central European Review of Economics and Finance*, 1(2), 1-14.

[10] Kızılgöl, Özlem ve Erbaykal, Erman (2008). Türkiye'de Turizm Gelirleri ile Ekonomik Büyüme İlişkisi: Bir Nedensellik Analizi, *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 13(2), 351-360.

[11] Oh, Chi-Ok (2005), The Contribution of Tourism Development to Economic Growth in the Korean Economy, *Tourism Management*, 26, 39-44.

[12] Özdemir, Ali Rıza ve Öksüzler, Oktay (2006). Türkiye'de Turizm Bir Ekonomik Büyüme Politikası Aracı Olabilir Mi? Bir Granger Nedensellik Analizi, *Balıkesir Sosyal Bilimler Dergisi*, 107-126.

[13] Pao, J.W., (2004), Recent Developments and Prospects of Macao's Tourism Industry, Monetary Authority of Macao

[14] Polat, Esra ve Günay, Süleyman (2012). Türkiye'de Turizm ve İhracat Gelirlerinin Ekonomik Büyüme Üzerindeki Etkisinin Testi: Eşbütünleşme ve Nedensellik Analizi, *Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 16(2), 204-211.

[15] UN World Tourism Organization, UNWTO, Tourism Highlights 2014 Edition.

[16] Vanegas, M ve Croes, R. (2007). Tourism, Economic Expansion and Poverty in Nicaragua: Investigating Cointegration and Causal Relations, *Staff Paper*, 7-10.

[17] Yavuz, Nilgül Çil (2006). Türkiye'de Turizm Gelirlerinin Ekonomik Büyümeye Etkisinin Testi: Yapısal Kırılma ve Nedensellik Analizi, *Doğuş Üniversitesi Dergisi*, 7(2), 162-171.

Evolution of the regional unemployment in Romania

Corina SCHONAUER (SACALĂ)
PhD Candidate, Cybernetics and Statistics Doctoral School, The Bucharest University of Economics Studies

Abstract

Unemployment is a social and economic phenomenon, with numerous implications at the individual level and at the level of an entire population. At the macroeconomic level, unemployment may be equated with the production value that would be realized by working unused population. At the individual level, unemployment leads to deterioration of living standards, social and psychological problems. În Romania, this phenomenon is detailed as follows: in terms of age, gender distribution, the degree of training, the duration of unemployment, but also in terms of the distribution of the territorial units. This article presents an analysis of unemployment data from the last two censuses and highlights the main changes that took place at the local level. The data presented come from the National Institute of Statistics of Romania.

Keywords: *unemployment, the structure of unemployment, long term unemployment, unemployment for young people*

1. Introduction

Romanian economy transition to a market economy has led to a lot of socio-economic changes that have generated changes in population structure. The decrease in the population of the 40 counties and in Bucharest City is determined by the negative natural increase (in most counties), internal and external migration. Ilfov County is the only county where the population increased by 29.5% between 2002-2011, due to economic development in the north of the capital. The territorial distribution of the population changes between the two censuses. Below is the first 5 and last 5 counties (except Bucharest), ordered by the population:

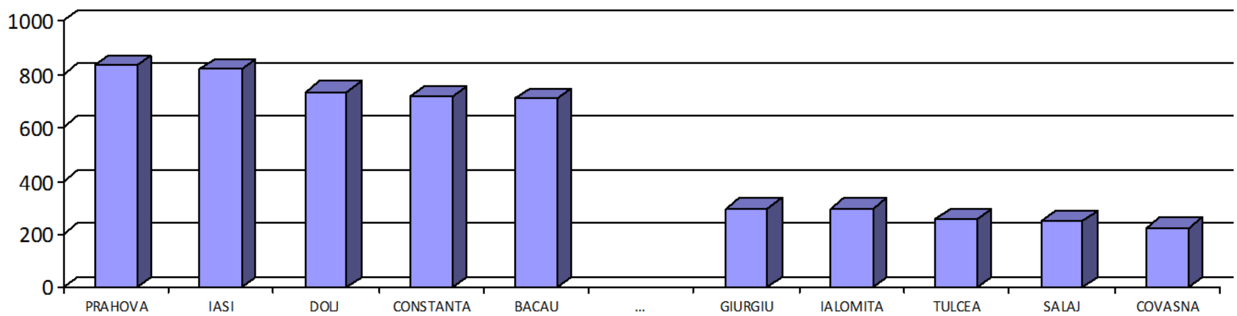


Fig. 1. Distribution of unemployed by regions, at 2002 Census

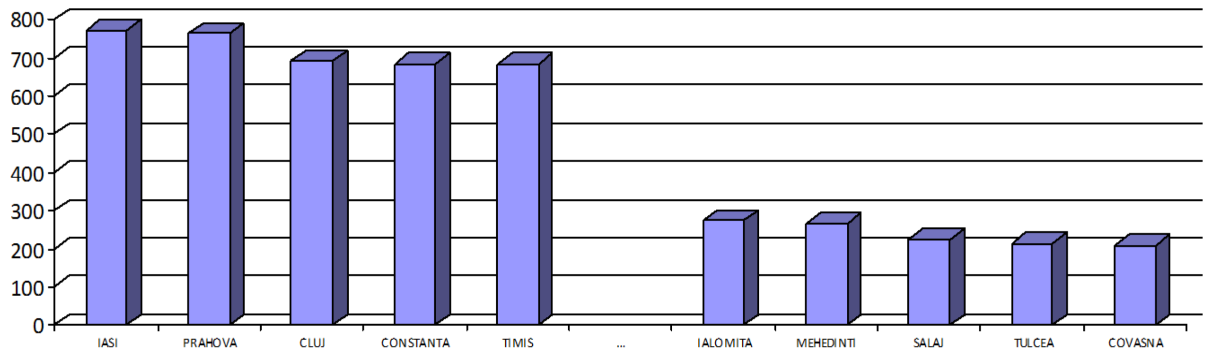


Fig. 2. Distribution of unemployed by regions, at 2002 Census

The different pace of development from one county to another, the creation or existence of large urban centers are the main factors influencing the intensity of internal migration. The most affected counties by the loss of population through relocation were Vaslui, Botoşani, Maramureş, and Olt.

International migration has contributed substantially to the decline in population, particularly since 2007, after Romania joined the European Union. At the territorial level, the sole beneficiary of the positive balance of external migration is Ilfov County. The biggest negative external migration we meet in Bacău, Galaţi, Iaşi, Neamţ, Suceava and Bucharest. The main destination countries for people going on long time are: Italy and Spain.

Also, economic development and changes in the age structure of the population have been changes in the structure of the economically active population (employed and unemployed). Thus, if in 2002 the share of unemployed in the total active population was 11.2%, in 2011 this share drops to 7.3%. Noteworthy is the fact that the share of active population in total population increased from 40.8% to 45.6% from one census to another.

The definition used in the two censuses for the term "unemployed" was the one used by the International Labour Office: persons aged 15-74 years who, in the reference period simultaneously meet the following conditions: have no job and are not carrying out any activity to get income; they are looking for a job, undertaking certain actions during the last four weeks (registering at employment agencies, or private agencies for placement, attempts for starting an activity on own account, publishing notices, asking for a job among friends, relatives, mates, trade unions a.s.o.); they are available to start work within the next two weeks if they immediately find a job.

2. Evolution of unemployed people between 2002 and 2011

At the 2011 census as against the 2002 census there was a reduction of 367,520 unemployed people, which means a reduction in the relative values of over 35%. In one county there is an increase in the number of unemployed: Bucharest, where, from a number of 58,293 unemployed in 2002 ascend to 66,031 unemployed in 2011. In four regions there was a greater decrease than the national level, namely : South East (about 50%), North East and South-Muntenia (over 43%), Central Region (over 40%).

Distribution of unemployed by regions is changing in time. But ranking of regions remain almost the same. Thus, in 2002 on the first place is South Muntenia Region, followed by North East Region, South East Region, Centre Region, North West Region, South West Oltenia Region, West Region and Bucharest-Ilfov Region. The same ranking is also found in 2011 except Bucharest-Ilfov Region, which reached on the 8th place from the 5th, and this because of the increasing number of unemployed in Bucharest City.

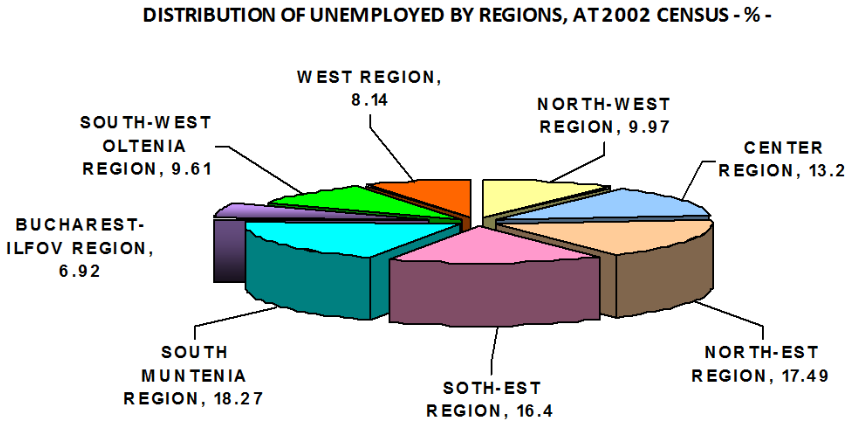


Fig. 3. Distribution of unemployed by regions, at 2002 Census

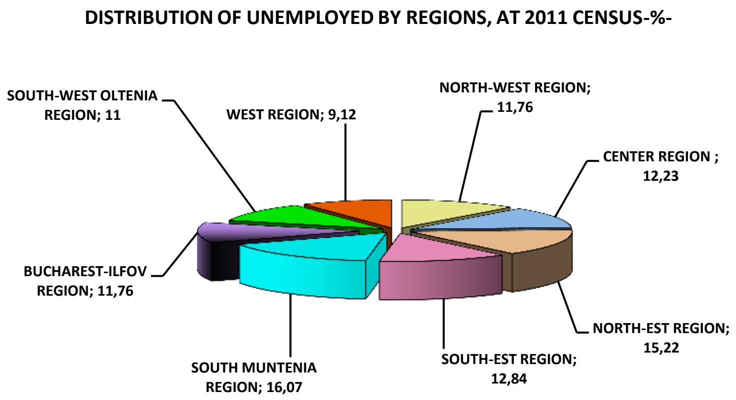


Fig. 4. Distribution of unemployed by regions, at 2011 Census

Analyzing the values of unemployment from the two censuses, at county level, we can see changes in their distribution by counties, which lead to changes in the ranking of counties, in terms of this indicator.

In the North West region, in Bistriţa Năsăud County, unemployed decreased to more than half and in Sălaj and Cluj with more than a third. Satu Mare County, however, registers an increase in the number of unemployed by 2,200 people, which makes that their share in this region will increase from 8.48% to 13.96%. In Central Region, counties with the largest decreases in the number of unemployed are: Covasna (over 60%), Braşov (over 40%), Alba (over 45%).

With over 40% decreases the number of people who said they were unemployed in the North East. In Botoşani, Vaslui, Neamţ the number of unemployed halves. But in Bacău county the decline does not even reach 25%.

In South-East Region, the biggest drop is found in Buzău (-26,088 persons). At the opposite end is Vrancea County, where the number of unemployed increased from a census to another, which leads to doubling their share in the total unemployed in the region (from 6.42% to 12.81%).

In South Muntenia Region, the number of unemployed people drops with over 82,000 and about 35% of them are from Prahova county. However, this county remains in first place, in terms of the number of people unemployed among the top counties in the region, with a slight drop of it's share from 31.68% to 29.38%. The smallest decrease in the number of unemployed is found in Dâmboviţa (only 17,75%), leading to an increased share in this region from 14.12% to 20.43%, thus climbing a place in ranking of this region.

The only region where the number of unemployed increased, the Bucharest-Ilfov nearly doubles its share in the total unemployed in the country (from 6.92% to 11.76%). The trend in the region is given almost exclusively from the Bucharest City, which holds over 80% of the unemployed in the region.

In South-West Oltenia Region, Olt County move from the second place to the 4th place in the ranking of this region, registering a fall of over 35% in the number of unemployed. Dolj County continues to be first in the region, living here almost 30% of the unemployed in the region.

Hunedoara County manages, by about 50% decrease in the number of unemployed, to move from first to second place among Western region. Only 10% decrease in the number of unemployed between the two censuses, raises Timiş on the first place in the region, with a share of almost 33%. To note is the situation in Arad, where the number of unemployed increased by 2,000 people.

3. Unemployment by duration between 2002 and 2011

Distribution of unemployed by duration of unemployment, according to the 2011 census, changes from that of 2002 census. We can remark the decline in the share of long term unemployment (over 27 months) from 33.2% to 20.7%. The counties where there is the greatest drops in the long-term unemployed are Prahova (over 19,000 people) and Buzău (over 15,000 people).

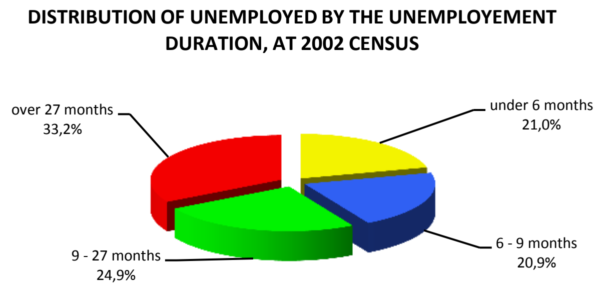


Fig. 5. Distribution of unemployed by the unemployement duration, at 2002 Census

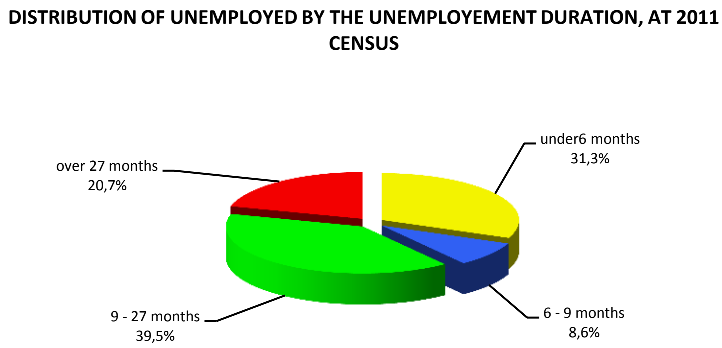


Fig. 6. Distribution of unemployed by the unemployement duration, at 2011 Census

4. Unemployed by age group between 2002 and 2011

Population distribution by age shows that the aging process is emphasized. The employed population is concentrated in age groups 30-39 years and 40-49 years. Instead, in the category of unemployed, young people under 25 have a significant share: 31.44% at 2002 census to 32.19% 2011 census. The share of employed persons of 50 years and over has increased due to raising the retirement age. But, the share of unemployed in the age group 50-59 years increased from 1.26% to 4.61%. In 2002 about 37% of total number of unemployed were persons searching for their first job and in 2011 their share dropped to 35.4%.

In Centre Region, only in Alba and in Braşov, the share of unemployed aged up to 25 years register a decrease below 28%. În the other counties the share of the young unemployed has a tendency to increase exceeding 32%. În Brasov, in 2011, it finds the largest increase in the share of unemployed in the age group 55 years and above (from 1.36% to 8.23%).

Bihor County is one of the counties where the percentage of young people under 25 years looking for their first job in the 2011 census reaches below 70%. In this region, only Cluj between the two censuses recorded an increase of over 5% for the unemployed of 55 years and above.

In Botosani county, in the North-East Region, at the 2011 census, young unemployed people under 25 years represent nearly 44% of the total number of unemployed. Here, the share of those who are looking for their first job, in this segment age, increased by 5% between the two censuses (from 79.8% in 2002 to 84.3% in 2011). In this region, the biggest share of unemployed in the age group 55 years and above was found in Vaslui (5.83%).

South East Region is characterized by high weights reached by the unemployed in the age group 55 years and over, in 2011, especially in the counties of Constanţa (8.73%) and Tulcea (9.68%). With the lowest share of elderly unemployed, Buzău County stands out for the high percentage of young unemployed: 43.21% at 2011 census, of which about 85% are seeking first job.

In Giurgiu County, the share of unemployed people aged up to 25 years attain to 40.3% in 2011, value that is are ranked third in the top of counties with the highest proportion of young unemployed people. In Prahova and Teleorman, which register higher shares of the unemployed over 55 years in South Muntenia Region, it highlights a downward trend in the share of young unemployed.

With a share of 25.6% of young people unemployed in 2011, the Bucharest ranks first among the counties with the fewest unemployed aged up to 25 years. Also, opportunities in the capital and lifestyle are leading to early engagement of young people, so that in 2011, only 68, 2% of the 16,915 young unemployed, were searching for their first job.

Valcea county, from South-Est Oltenia Region, stands out as the county with the lowest percentage of unemployed people aged 55 years and older in the total unemployed in the county (0.5% in 2002 and 2.96% in 2011).

West Region is distinguished by small weights of young unemployed in the total number of unemployed in the county. Arad County registers a big enough increase of share of unemployed persons aged 55 years and over, reaching 8.17% in 2011, followed by Timiş with 7.5%.

Conclusions

Demographic changes, economic and technological progress have a permanent influence on the structure of the labor force in different territorial units in Romania. By comparing the last two censuses is observed qualitative and quantitative changes of the components of the active population, including that of the unemployed. The unemployment rate in 2011 was 7.3 per 100 active persons, 4.5% less than in 2002. Socio-economic development of the various different territorial units of the country has led to changes in the concentration of unemployment in some areas. The North-East, South-East and South focuses most unemployed people. Here they are found most unemployed with unemployment lasting more than 27 months.

Once Romania joins the European Union should adopt a number of measures to help encourage people to remain in work or find a new job, including: the promotion of a life-cycle approach to work, encouraging lifelong learning, improving support to those seeking a job, as well as ensuring equal opportunities in all territorial units.

References:

[1] *"Aspects of demographic changes at Population and Housing Census 2011"*, National Institute of Statistics of Romania, Bucharest, 83, 2014.
[2] *"Romanian Statistical Yearbook"*, National Institute of Statistics of Romania, Bucharest, 730, 2002- 2014.
[3] *"Romanian labour force: employment and unemployment"*, National Institute of Statistics of Romania, Bucharest, 288, 2014.
[4] - site www.insse.ro - RPL 2002.
- site www.insse.ro - RPL 2011.