Good Practices in Empirical Corporate Finance and Accounting Research

Marek Gruszczyński
SGH Warsaw School of Economics, Poland
marek.gruszczynski@sgh.waw.pl

ABSTRACT

This paper presents the risks of quantitative research pertaining to corporate finance and accounting. These is followed by a survey-like catalogue of good practices in modelling. All considerations are rooted in financial microeconometrics, the field for examining both practical and theoretical questions of applying econometric techniques in corporate finance and accounting research based on the use of microdata (Gruszczyński 2018a).

Two major parts of the paper include: (1) discussion of the typical drawbacks in applying regression-type models, like causality vs. correlation, selection of explanatory variables and endogeneity, and (2) list of good practices in microeconometric applications to corporate finance and accounting research, based on Faff (2017), Kennedy (2002), Adams (2017), Hyndman (robjhyndman.com/) and author's own experience.

Catalogue of good practices in microeconometric applications to corporate finance and accounting may serve as the checklist for students and researchers.

JEL classification: C50, C58, G30, M40

Keywords: financial microeconometrics; empirical corporate finance; applied accounting; good practices in research

1. INTRODUCTION

Quantitative research in corporate finance and accounting commonly uses microdata on companies, their activities, legal facts, social behavior etc. Typical examples are financial statements, legal registers, statistical reports, always from a possibly large number of companies, like hundreds or thousands. Microdata help to quantify various hypotheses we construct around questions in corporate finance and accounting research.

Well prepared samples of microdata enter the statistical-econometric models and take part in their estimation and verification. The methodology is mainly rooted in microeconomerics. If applied to research problems in corporate finance and accounting it is known as financial microeconometrics (for further discussion see Gruszczyński 2012, 2018a).

Purpose of this paper is to indicate dangers of deficient practices in quantitative research and, most of all, to advocate good practices of quantitative modelling in corporate finance and

accounting. This paper is founded on many thoughts expressed in another author's paper published in Polish (Gruszczynski 2018b).

The discipline of finance has very quantitative background and a number of solid theories. However, as with other social science disciplines, theories are often far from current realities, their endurance is low and they have limited validity across countries and across the financial orders. Similarly, research in corporate finance and accounting frequently yields inconclusive and/or dissimilar results, depending on market, sample, observation period etc. (Gruszczyński 2012). It should be accepted that the research result is limited to "here and now" and is not general like mathematical theorem.

Therefore, it is of vital importance to have some common understanding about research methodology of statistical-econometric origin applied to corporate finance and accounting. The obvious starting point is always the world literature survey where one always finds major stream of the hypotheses in question. The core are papers in leading A-journals, mostly papers of quantitative (statistical) edge. Most papers use databases of microdata on financials.

The structure if this paper is as follows. Section 2 discusses the regression model as prevailing tool for researchers in corporate finance and accounting. Section 3 aims at showing several views on what can be regarded as good practice in applying quantitative (econometric) methodology. Section 4 concludes the paper.

2. REGRESSION MODEL AND ITS DRAWBACKS

Researcher in corporate finance and accounting typically believes in the power of regression model, i.e. the equation where we have a left-hand side variable Y (explained, endogenous) and a number of right-hand side variables X (explanatory, exogenous). Model can have many forms but idea remains the same: there are some determinants (explanatory variables) which explain behavior (variability) of the explained variable. And, when one shows – with some care – that "statistically" model is correct then this constitutes the "proof" of validity of the examined relationship.

Yes, this might be the case but usually it isn't. Why? There is a number of aspects, controls, questions that such model should answer. If we do not follow certain "checklist" of those items, model may seem to be formidable but in fact is not correct or appropriate.

2.1. Causality and selection of variables

First question concerns the type of hypothesis one aims at verifying. Using regression we commonly seek to show that some activity represented by exogenous variable X is (or not) the determinant of behavior of Y variable. This is OK as long as by "determinant" we understand X being associated, correlated with Y. To prove that X is the "cause" of Y we need more than regression model and – even so – the proof is not very much general.

Their survey paper on causality in empirical corporate finance research Atanasov and Black (2016) begin with following paragraph: *"Much corporate finance research is concerned with causation – does a change in some input cause a change in some output? Does corporate governance affect firm performance? Does capital structure affect firm investments? How do corporate acquisitions affect the value of the acquirer, or the acquirer and target together? Without a causal link, we lack a strong basis for recommending that firms change their behavior or that governments adopt specific reforms."*

In order to prove the validity of this type of questions for data representing a particular time and space setting it is necessary to use techniques other than regression. They are based on "treatment effects" approach, i.e. on understanding that there is a treatment variable like "new

governance order in company" and the effect variable like "company's financial performance". For examining causal relationships most appropriate are methods of counterfactual analysis, such like matching techniques (including propensity score matching), difference-in-differences methods, regression discontinuity approach and other.

These methodological approaches are appealing, however not easy to correctly apply. Atanasov and Black (2016) suggest comprehensive list of conditions to be met in order to accept analyses based on counterfactual approach. Authors have studied 13 thousand articles from 22 leading journals of economics, finance, law and management in 2001–2011. Among them, 863 papers focused on the relationship between corporate governance and companies' performance and out of this number 74 papers were using *shock-based research* approach (with 40 different shocks/ treatments). Authors concentrate on examining three very good papers and on showing that – even in such cases – there are doubts as to all stages of research process, especially to techniques aimed at proving causality effects.

Summing up: techniques for properly verify causality effects are not trouble-free solutions to questions which are not resolved by regression analysis.

Knowing that, is it worth to use regressions in corporate finance and accounting research? The answer is the obvious "yes". Regression analyses are valuable, especially when there is no other choice. Their outcomes are often close to ascertaining causality, particularly when we use panel regression techniques. Common regression (correlation) technique based only on observational data has no ability to evidence causality. However, it has valuable interpretative value, especially when the sample includes companies properly fit to each other.

Major question in setting up the regression analysis is the choice of explanatory variables. Solutions are more or less the same for linear regression as for more complicated nonlinear models. For example, in logit or probit model the explanatory variables enter the linear combination of regressors in the same way as in linear models. Selection of explanatory variables (regressors) to the models shall be based on theory, relevance and the experience of other researchers in the same field. In the framework of big-data and more operational approach, the process of selection of regressors may be more technical and (therefore) atheoretical.

2.2. Endogeneity

Now comes weightier element at the finale of discussing regression analysis and its drawbacks for corporate finance research. It is the problem of endogeneity, today more than fashionable, at least in asking questions about the quality of research reviewed. What it is about? Technically, this is the situation when the explanatory (exogenous, regressor) variable in regression model is correlated with its error term, i.e. correlated with another regressor which has not been included in the model. Well, this is just technical explanation that "most corporate financial decisions are determined endogenously in a complex network of relationships" (Li 2016).

The following example is taken from Li (2016). Assume that the performance of a company (endogenous variable) is explained by CEO "power" and by other explanatory (exogenous) variables. CEO power is expressed by means of index showing the "salary distance" between CEO and the no. 2 person in the company [i.e. (CEO compensation – no. 2 person compensation)/ (CEO compensation)]. This index is called *GAP*. Financial performance of the company is represented by Tobin's Q. How *GAP* can be endogenous in the model explaining company's performance (Q)? There are two possibilities: "either causality runs from Q to the *GAP*, or causality runs both ways. A random shock that enters the regression model through the error term affects Q. Because Q affects the *GAP*, *GAP* will be correlated with the error term, generating a biased coefficient on the *GAP*. The second situation is that the *GAP* and Q have no direct effect on each other, but they are spuriously correlated through some third variable. If we do not

explicitly control for the third variable, the error term will absorb the effect of this variable. Thus, the error term will be correlated with the GAP, causing biased and inconsistent estimates".

So, in the regression model we have the left-hand side variable Q and the right-hand side variable *GAP*. To address the possible endogeneity of *GAP* one can use various techniques. Li (2016) uses instrumental variables approach, fixed effects model (firm and year effects), lagged dependent variables and control variables. All this resulted in significant change of the regression coefficient by *GAP* variable: from positive when the endogeneity effect is not addressed to negative when the techniques of accounting for endogeneity are applied. In another words: more power of the CEO is associated with poorer company's performance.

Question of attending to endogeneity in a model is never simple. Investigating the relationship between managerial ownership and corporate performance Coles, Lemmon and Meschke (2012) found that "in the ownership–performance context, the use of proxy variables, fixed effects, and instrumental variables does not generally provide a reliable solution to the endogeneity problem. One prominent reason is that nonlinear regression specifications are likely to be appropriate. Another is that some remedies do not address endogeneity arising from the joint time-series (within-firm) variation of performance, size, and contract design"

Thus, it will be continuously the challenge of how to deal with endogeneity in the specific model. In the context of empirical corporate finance it is worth to refer to an excellent paper by Roberts and Whited (2013). They indicate two groups of techniques as remedies for endogeneity. First are those which take into account the source of variability of exogenous variable: instrumental variables approach, estimators differences-in-differences and also the regression discontinuity design. Second group of techniques makes use of the endogeneity in the modeling itself: e.g. use of panel data or matching estimation.

Today we know that addressing endogeneity in regression-type models may give inconclusive results, depending on the research setup, the researcher's creativity etc. Nevertheless, it is advocated to always consider this question. Atanasov and Black (2016) comment on that: « We share neither the perspective of some researchers, whose view can be caricatured as "endogeneity is everywhere, one can never solve it, so let's stop worrying about it"; nor the "endogeneity police", whose attitude is that "if causal inference isn't (nearly) perfect, a research design is (nearly) worthless"; nor that of authors who know they have an endogeneity problem, but say little or nothing about it in their paper, hoping the referee won't notice, or else use a weak instrument to address endogeneity and hope the referee won't object. Our anecdotal sense is that paper acceptance and rejection decisions often turn on which position – endogeneity is everywhere, endogeneity police, or our middle ground – best describes the referee and the editor ».

The deficiencies of regression analysis indicated above, such as inference about causality or the endogeneity problem and the proposed remedies are not the common knowledge among econometricians, unfortunately. Angrist and Pischke (2017) comment on that in a survey of undergraduate syllabi of econometrics in the leading US universities.

3. GOOD PRACTICES

Statistical-econometric methodology in contemporary corporate finance and accounting research is dominated by techniques based on the use of microdata. In typically proposed regression-type models the techniques of microeconometrics and advanced data analysis are prevailing. Variety of research questions and proposals how to solve them might certainly entertain some standard or good practice rules or suggestions to follow.

Good practices in microeconometric modeling are proposed on the basis of author's own experience and on the survey articles of similar type. Important inspiration is the paper by Renée Adams (2017) for researchers in corporate governance topics. We have used also papers of

Kennedy (2002) on ten commandments of applied econometrics as well as the page of Rob Hyndman (robjhyndman.com/). Invaluable was also the pitching template by Robert Faff (2017).
　　And now the good practices.
1. Dedicate reasonable share of your time for this[1]:
　– formulate major features of your research question in one sentence,
　– identify sources (papers) of the scientific mainstream being modern foundation of your topic,
　– write in one paragraph the motivation for your intended research, indicating the "puzzles" you are going to solve; avoid questions which are wrong, remember that "an approximate answer to the right question is worth a great deal more than a precise answer to the wrong question",
　– identify fundamental idea that "drives intellectual content" of your research topic, along with major research hypothesis,
　– determine the key explained variable in your model as well as the principal explanatory variables; consider the possibilities of endogeneity and the viable remedies,
　– carefully explore the availability and quality of your data,
　– discuss major steps in your research schedule and the choice of (quantitative) methodology,
　– clarify again the originality of the intended research; how will you argue that it is really novel?,
　– answer the question: "So what?"; why your topic is reasonable? "how will major decisions/behaviour/activity etc be influenced by the outcome of this research?".
2. Once you decide to apply the econometric/ microeconometric model, remember about the following[2]:
　– propose your model on the basis of relevant theory (if possible), make use of the results of other researchers, and apply a lot of common sense,
　– if you employ microdata do not forget that they have low level of aggregation; due to this the classical linear relations are rarely applied; for example, this is the reason why goodness-of-fit measures, like R-squared, have low values; possible substantial heterogeneity of units (companies) should be taken into account,
　– samples of microdata in corporate finance and accounting research are typically not random; do not avoid questions about sample biases; most research uses samples of large public companies, commonly the best on the market; remember that the investigation results pertain only to companies in the sample,
　– use the rule of KISS: *Keep It Sensibly Simple* (but not *Keep It Simple, Stupid*); this rule of simplicity may mean many things (see: Kennedy 2002); we advocate the following: do not include too many explanatory variables into your model (e.g. if you have five profitability ratios use only 1 or 2); your explanatory variables shall have some merit; avoid adding new variables unless you are sure that they belong to the model (otherwise you quickly switch to data-minig exercise),
　– in most models, also non-linear ones, the explanatory variables enter in the form of linear combination (i.e. as the sum of products of each variable by its parameter); it is advisable that in such linear combination we shall minimize effects of multicollinearty, i.e. high mutual correlation of explanatory variables; this might be done e.g. by selecting to the model the variables that are not very much correlated with each other,
　– make sure that the estimation result can be sensibly interpreted: that the signs of parameter estimates are as expected / fit to theory (sometimes: are the same as sings of simple correlations[3]), that the variables which are important are also statistically

---

[1]　Following Faff (2017) and Faff et al. (2017); also Kennedy (2002) and Hyndman (robjhyndman.com/).
[2]　Following Kennedy (2002), Gruszczyński (2012).
[3]　Gruszczyński (2012, p. 82).

significant[4]; do not forget that small values of estimates are not the symptom of their "lesser validity" (all is decided by statistical tests),

–   do not forget that in order to show the causality we need to use special techniques (see section 2 above); otherwise the only interpretation shall be using terms: "correlation", "association",

–   carefully use the potential of data-mining; strategy of modelling in econometrics is rooted in theory as the inspiration for specifying model equations; once we employ data-mining in order to find model that "best fits to the data" then the result may not be correct, especially when appropriate theory exists; however, the data-mining exercises may reveal regularities that can "be seen in the data", and this may constitute a good hint for further modelling; sometimes the modelling target is not about revealing relationships between explanatory and explained variables but about predicting (*ex post*) the values of explained variable, in such case the data-mining techniques are also appropriate.

3. Quantitative research in corporate finance and accounting as reported in published papers, especially in good quality journals, is carefully scrutinized by reviewers. Here is the excerpt from the set of "good practices" formulated by prof. R. Adams (2017)[5]:

–   remember about your assumptions; most papers are rejected because of critical reviews concerning the choice of instrumental variables or treatment effects approach – without discussing assumptions necessary for model identification; key element is also the proper description of institutional set-up concerning the investigated issue of corporate finance or accounting (e.g. the legal framework),

–   clarity and transparency of the text is important to the same extent as research result itself; follow the literature and avoid preparing the article already written by someone else; in the same time, do not believe everything which has been published, question what you read; do not forget about visualization of your thoughts (a must when you use difference-in-differences and regression discontinuity techniques),

–   if you intend to demonstrate the causal effects, it is necessary to show relevant strategy of identifying those effects; this is not simple; most papers ignore such discussion, other employ incorrect identification techniques; e.g. popular technique of difference-in-differences is usually not very appropriate (e.g. in corporate governance topics it is difficult to find a set-up corresponding to a proper medical-type experiment); also, matching techniques are usually not applied correctly; obviously, neglecting discussion about causality is not advisable; if identification of causal effects is not feasible then the best solution is to apply regression and correlation techniques – with deep analysis of possible biases in the results (for this, prof. Adams advocates paper by Miller, 2013),

–   good dataset is crucial for the results; for researchers in corporate finance and accounting the everyday headache are various deficiencies like holes in the data, limited availability of data for "soft" variables, not unified data for quantitative variables, data incomparable between companies, etc.; on the other hand, the availability of more data is not always positive for modelling (prof. Adams points out that in some corporate governance research the cross-sectional data fare much better than panel data),

–   it is advisable to consider all "classic" econometric topics like: correction for heteroscedasticity, use of fixed effects (if possible), OLS as the benchmark for more advanced techniques, etc.; however, the question of statistical significance of variables is not very important since for large datasets it is almost always assured – more important is interpretation and feasibility of the result,

–   important is the replicability of your outcomes; more and more often the journals require to supply your dataset along with the paper; sometimes, your dataset is placed online.

---

[4]   Gruszczyński (2012, p. 67).

[5]   Adams (2017); prof. Adams reports her hints and suggestions in the form of "Adams' alphabet" – from A to Z; we show here only few items.

Finally, it is worth to mention Winston Churchill and his confidential memo, dated August 9, 1940, entitled *Brevity* (to be found in internet). It is devoted to eliminating unnecessary jargon and concentrating on the intended message: "discipline on setting out the real points concisely will prove an aid to clearer thinking".

4. CONCLUSION

Corporate finance and accounting research calls for applying modern statistical-econometric methodology, with the extensive use of sets of microdata on companies. For this purpose it is advisable to use several "good practices" in order to avoid unnecessary efforts leading to incorrect results. The research projects in corporate finance and accounting applying the techniques of microeconometrics are exposed to a number of risks, most of them connected with uncertainties about relevant methodological approach.

The financial microeconometrics topics emerge as the field for examining both practical and theoretical questions about applying econometric techniques in corporate finance and accounting research based on the use of microdata. This paper offers a short survey on recommendations concerning quantitative research in corporate finance and accounting. Thoughts of Faff (2017), Kennedy (2002), Adams (2017) and Hyndman (robjhyndman.com/) are supplemented by the author's own experience. The catalogue of good practices is by no means exhaustive although it may serve as the starting point for considerations about new research projects in corporate finance and accounting.

References

Atanasov V., Black B. (2016). Shock-Based Causal Inference in Corporate Finance and Accounting Research. *Critical Finance Review*. 2016, 5: 207–304.

Adams R.B. (2017). The ABCs of empirical corporate (governance) research, *Corporate Governance International Review.* 2017; 25: 461–464.

Angrist J.D., Pischke J.-S. (2017). Undergraduate Econometrics Instruction: Through Our Classes, Darkly. *Journal of Economic Perspectives*. 31 (2): 125–144.

Coles J.L., Lemmon M.L., Meschke J.F. (2012). Structural models and endogeneity in corporate finance: The link between managerial ownership and corporate performance. *Journal of Financial Economics*. 103 (2012); 149–168.

Faff R., (2017). *Pitching Research®*. Version 15 (or next). Available in SSRN: http://ssrn.com/abstract=2462059.

Faff R. et al. (2017). *Increasing the Discoverability of Non-English Language Research Papers: A Reverse-Engineering Application of the Pitching Research Template*; version May 2017; Available in SSRN: https://ssrn.com/abstract=2948707.

Gruszczyński M. (2012). *Empiryczne finanse przedsiębiorstw. Mikroekonometria finansowa*. Warszawa: Difin.

Gruszczyński M. (2018a). Financial Microeconometrics as Research Methodology in Corporate Finance and Accounting, in: T. Dudycz, G. Osbert-Pociecha, B. Brycz (ed.) *Efficiency in Business and Economics*, *Springer Proceedings in Business and Economics*, 2018, p. 71–80.

Gruszczyński M. (2018b). Badania ilościowe w finansach przedsiębiorstw – wyzwania metodyczne. *Finanse, Rynki Finansowe, Ubezpieczenia*, nr 1/2018 (91): 23–34.

Kennedy P.E. (2002). Sinning in the Basement: What are the Rules? The Ten Commandments of Applied Econometrics. *Journal of Economic Surveys*. 16: 569–589.

Li F. (2016). Endogeneity in CEO power: A survey and experiment. *Investment Analysts Journal*. 45:3, 149–162.

Miller, M.K. (2013). The uses and abuses of matching in political science. *Working Paper*, George Washington University. Available as https://sites.google.com/site/mkmtwo/research.

Roberts M.R., Whited T.M. (2013). Endogeneity in Empirical Corporate Finance, in: G. Constantinides, M. Harris, R. Stulz (ed.). *Handbook of the Economics of Finance*, vol 2, Amsterdam, North Holland.