Chair for Quantitative Business Administration

Topics for a Master Thesis

1st June 2018

Contents

1 Topics in Climate Economics (Supervisor: Alena Miftakhova)

1.1 Low-Dimensional Representation of Complex Climate Models

Background and Research Question

The key tasks of the integrated climate/economic research are to examine the interaction between the climate system and economy and suggest optimal policies to curb human-induced changes. The corresponding models should provide computationally feasible ways to quantify the mutual influence of the two systems. Yet, it is computationally infeasible to merge state-of-the-art climate models with comprehensive macroeconomic settings. Even the most advanced integrated assessment models have to rely on the accuracy of less complex representations.
The goal of the project is to design an efficient mapping from the indicators of human activities (e.g. greenhouse gas emissions) to the induced changes in the climate system, both on global and regional scale, following Miftakhova et al. (2016). Such a mapping should be verified based on the most recent available observational data. An additional challenge is quantifying the uncertainty around the system response.

Literature

Miftakhova, A.; Judd, K. L.; Lontzek, T. S.; Schmedders, K. 2016. Statistical Approximation of High-Dimensional Climate Models. Available at SSRN:https://ssrn.com/abstract=2887292.

Meinshausen, M.; Raper, S.; Wigley, T. 2011. Emulating coupled atmosphere-ocean and carbon cycle models with a simpler model, MAGICC6Part 1: Model description and calibration. Atmospheric Chemistry and Physics 11 (4), 14171456. http://dx.doi.org/10.5194/acp-11-1417-2011.

IPCC. 2013. Summary for policymakers. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA. www.climatechange2013.org.

1.2 Financial Markets and Climate Models: Empirical Analysis

Background and Research Question

Complex climate models are the most advanced tools to forecast climate for months, years and decades ahead. However, even their short-term forecasts are subject to systematic biases when verified on observational data. If there are any better forecasts available, efficient financial markets should reflect them in the prices of commodities that are sensitive to climate conditions.

Crops futures prices should first of all reflect the expected conditions of corresponding growing seasons.

The objective of the project is to investigate whether financial markets contain any information in addition to the forecasts of climate models. In particular, corn, wheat or soybean markets can serve as an example of a commodity market sensitive to climate conditions.

Literature

Roll, R., 1984. Orange Juice and Weather. The American Economic Review 74 (5), 861–880.

Slater, L. J., Villarini, G., Bradley, A. A., 2016. Evaluation of the skill of North-American Multi-Model Ensemble (NMME) Global Climate Models in predicting average and extreme precipitation and temperature over the continental USA. Climate Dynamics, 1-6.

Kirtman, B. P., Min, D., Infanti, J. M., Kinter, J. L., Paolino, D. A., Zhang, Q., van den Dool, H., Saha, S., Mendez, M. P., Becker, E., Peng, P., Tripp, P., Huang, J., DeWitt, D. G., Tippett, M. K., Barnston, A. G., Li, S., Rosati, A., Schubert, S. D., Rienecker, M., Suarez, M., Li, Z. E., Marshak, J., Lim, Y.-K., Tribbia, J., Pegion, K., Merryfield, W. J., Denis, B., Wood, E. F., 2014. The North American Multimodel Ensemble: Phase-1 Seasonal-to-Interannual Prediction; Phase-2 toward Developing Intraseasonal Prediction. Bulletin of the American Meteorological Society 95 (4), 585–601.

2 Topics in Structural Estimation (Supervisor: Gregor Reich)

2.1 Statistical Efficiency in the Estimation of Dynamic Discrete Choice Models

Background

When applying dynamic economic models, usually the models first have to be *solved* before they can be analyzed, which is often a challenging task of its own from a computational perspective. However, when the goal is to estimate such a model, given data on the state and the corresponding decisions of the agents (for example, the state and age of a car, and the corresponding decision on whether to carry out an expensive repair, or to scrap it instead), the solution of the model can be circumvented under some conditions, by directly inferring the solution from the observed actions in the data. This concept is called *two-step estimation*, and goes back to Hotz and Miller (1994). As a downside, this method is know to have a lower *statistical efficiency*, as the parameter estimates usually face a substantial small sample bias.

A more recent development in the literature on estimating dynamic discrete choice models (DDCM) is the incorporation of state variables that are *serially correlated*, and at the same time unobserved by the econometrician who estimates the model. An important example for such as

state variable is learning: While it takes place in the head of the decision maker and is therefore often not part of any measurement nor data, it is (hopefully) serially correlated as there is usually progress in learning over time. To compute the likelihood function for such a model, a large "average" has to be computed, taking into account all possible values of the unobserved state variables (e.g. learning states), which corresponds to—mathematically speaking—computing a *high-dimensional integral*. For both estimation approaches, recent methods exist to incorporate such variables, namely Arcidiacono and Miller (2011) for the two-step estimators, and Reich (2016) for the maximum likelihood estimation.

Research Question

To date, nothing is known about the statistical efficiency of two-step estimators for DDCMs with serially correlated unobserved states in practice. Therefore, in this project, you pick a (simple) DDCM, featuring serially correlated unobserved states, and estimate it both using (i) the maximum likelihood estimation approach with recursive likelihood function integration by Reich (2016), and the two-step *expectation–maximization* approach by Arcidiacono and Miller (2011). Using *simulated datasets* and thus having control over the data generating process, you can extensively analyze and compare the two methods for statistical as well as computational efficiency.

Literature

Arcidiacono, P., Miller, R. A. (2011). Conditional Choice Probability Estimation of Dynamic Discrete Choice Models with Unobserved Heterogeneity. Econometrica: Journal of the Econometric Society, 79(6), 1823-1867.

Hotz, V. J., Miller, R. A. (1993). Conditional Choice Probabilities and the Estimation of Dynamic Models. The Review of Economic Studies, 60(3), 497.

Reich, G. (2016). Divide and Conquer: Recursive Likelihood Function Integration for Hidden Markov Models with Continuous Latent Variables. Revise and Re-Submit at Operations Research.

2.2 Estimation of Agent-Based Models

Background

Agent-based computational economics is a simulation-based approach to explain aggregate dynamics from actions on the micro-level, e.g. explaining stylized facts of financial markets by modeling the behavior of individual investors and their environment. For example, the model of Kirman (1991) generates herding behavior through simple agent interaction that replicates some standard stylized facts such as volatility clustering and fat tails of the return distribution. However, agent-based models usually have many parameters, some of which have no real-world equivalent and are thus hard to set. Given data on the particular market that is modeled, one

can choose the parameters such that the real data is "well predicted" by the model in some sense; several approaches exist using minimization of a *moment function* (see, for example, Gilli and Winker, 2003, or Franke, 2009), to match the statistical moments of real data (e.g. stock returns) to the ones from the simulated data.

However, the fundamental problem of most existing methods face is that the evaluation of the moment function involves simulating the model, and thus is subject to *noise*: if the moment function is evaluated several times for identical parameter values, it will always return slightly different values, since the model simulation always produces different paths. As a consequence, the objective function has many local extrema—both, minima and maxima—that are purely artificial, and make it difficult for the optimization solver to identify the true, "natural" solution to the moment fitting problem.

Research Question

In this project, you will first review the literature on more recent heuristic approaches using simulation techniques to find parameter values that produce a good fit to the data; this includes studying a master's thesis on this topic previously written at the chair (Foster, 2014). Second, a different approach to the estimation of simulation models is implemented as a prototype, and then applied to the models of Kirman (1991) and Franke (2009); finally, the estimation results and their robustness are compared to the findings of the original papers.

Literature

Foster, J. (2014). Structural estimation using global optimization techniques. Master's thesis, University of Zurich.

Franke, R. (2009). Applying the method of simulated moments to estimate a small agent-based asset pricing model. Journal of Empirical Finance, 16(5), 804-815.

Gilli, M., Winker, P. (2003). A global optimization heuristic for estimating agent based models. Computational Statistics & Data Analysis, 42(3), 299-312.

Kirman, A. (1991). Epidemics of opinion and speculative bubbles in financial markets. In Money and financial markets (pp. 354-368). Blackwell.

2.3 Parallel Global Optimization for Estimation Problems

Background

The process of *estimating a economic model*, i.e. fitting its parameter such that the data predicted from the model fits the real data as good as possible, is tightly linked to (mathematical) optimization: Either, a function called the *likelihood function*, which maps the model parameters to the probability of observing the real data given the model, is *maximized*, or, *moment function*, which maps the model parameters to the statistical moments of the data predicted

by the model and compared them to the moments of the real data, is *minimized*. However, in "real" problems, these functions often exhibit difficult numerical properties, most prominently multiple solutions, and non-differentiability, which are hard to overcome by traditional optimization solvers.

More recently, a class of methods called *heuristics* has been established, solving optimization problems based on less rigorous concepts, therefore often not relying on particular mathematical properties such as differentiability or convexity. One class of heuristics to find the globally optimal solution are the *population based heuristics*, with its most prominent member being the *genetic algorithms*: Here, the solution is not obtain from following a trajectory of candidate solutions obtained from purely local improvements—such as prototypically done in the steepest decent methods; rather, a "population" of candidate solutions is maintained, and each individual is evolved based on the "knowledge" from the whole population. By means of a *fitness comparison*, good candidates are favored over bad ones, and the population as a whole is evolutionarily evolved. This procedure can overcome the global solution problem, by making sure that solutions that are locally, but not globally optimal will eventually be replaced by global solution candidates; the application of heuristics in econometrics to overcome the globality problem has triggered some substantial interest in the past, see e.g. Dorsey and Mayer (2012).

Another recent development in computational sciences is the fact that single compute units are not becoming faster anymore; rather, the *number* of compute units, often the *cores* of a processor, is increasing instead. Therefore, computing-intensive applications—statistical estimation being one example!—are more and more required to make use of parallelism.

Research Question

In this project, you will first review the literature on parallel population based heuristics, in particular regarding the limited degree of parallelism arising from small population sizes, and the load-imbalancing problems if the evaluation time of the objective function varies. Second, you will conceptually derive an *asynchronous* population based method, implement a prototype, and compare the results to the standard algorithms from the literature.

Literature

Dorsey, R. E., Mayer, W. J. (2012). Genetic Algorithms for Estimation Problems With Multiple Optima, Nondifferentiability, and Other Irregular Features. Journal of Business & Economic Statistics.

2.4 Estimation of Poorly- or Non-Identified Parameters

Background

When estimating the parameters of economic models, such as dynamic discrete choice models (DDCM), care has to be taken whether the parameters are *identified* at all, i.e. whether the model is set up such that given the data, it is possible to recover the parameters using an

estimation technique such as maximum likelihood estimation. It is well known that for many models, important parameters such as the time discount factor cannot be recovered together with the other parameters of the model. For example, Rust (1987)—pioneering the DDCM estimation literature—notes that "I was not able to precisely estimate the discount factor β. Changing β to .98 or .999999 produced negligible changes in the likelihood function and parameter estimates of [the other parameters]. The reason for this insensitivity is that β is highly collinear with [one of the other parameters]" which holds true for many other models as well.

Research Question

In this project, you will first compare different approaches to estimate models for *all* values of a particular *critical parameters*, and argue under which circumstances the "naïv" approaches fail. Second, you will derive a novel approach, implement and apply a prototype, and compare your results both qualitatively and quantitatively.

Literature

Rust, J. (1987). Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher. Econometrica: Journal of the Econometric Society, 55(5), 999-1033.

3 Topics in Portfolio Optimization (Supervisor: Robert Erbe)

3.1 Applied Topics in Asset Management

We have established a fruitful collaboration with a Swiss-based asset manager that allows master students to work on real-world portfolio optimization problems. The general research topics are typically quite involved and require solid mathematical knowledge along with good programming skills. The necessary datasets as well as supplementary R code will usually be provided to the student—thus having already worked with R often makes things a bit easier but it is no prerequisite. We invite ambitious students to proactively approach us by sending us their CV as well as their latest university transcript, highlighting their areas of interest.

The following topics have been assigned to students in the past (non-exhaustive list):

- Minimum Variance Portfolio with Market Impact Costs
- Estimation of High-Dimensional Covariance Matrices
- Portfolios from Sorts – Bayesian Methods for Portfolio Optimization

4 Topics in Blockchain Mechanism Design (Supervisor: José Parra Moyano)

4.1 Blockchain Mechanism Design

Background

Smart contracts are computer protocols that facilitate, verify, and enforce contracts. Blockchain Technology empowers the implementation of smart contracts that can substantially improve existing inter- and intra-corporate structures, making them more efficient and avoiding the conduction of parallel tasks. These new rules can open a new field of study in the Information Systems literature.

Research Question

The research question is to analyze how the business logic of an existing corporate structures can be improved by means of implementing smart contracts in a Blockchain. Specifically, the objective is to find out how "exporting" an existing business model to a Blockchain structure can improve the efficiency of the business logic, avoid failures and inefficiencies. Answering these questions can serve as a basis for the development of the Applied Information System Literature.

Literature

Fung, B and Halaburda, H. (2016). Central Bank Digital Currencies: A Framework for Assessing Why and How, Curreny Department Bank of Canada.

Barrdear, J. and Kumhof, M. (2016). Staff Working Paper No. 605. The macroeconomics of central bank issued digital currencies

5 Topics in Optimization Models for Portfolio Selection (Supervisor: János Mayer)

5.1 Portfolio Selection with Equal Risk Contribution according to Conditional Value–at–Risk

Background and Reserach Question

The financial portfolio selection literature reflects an increasing popularity of the Equal Risk Contribution (ERC) approach. Generally ERC means a portfolio selection technique, which is based on the decomposition of the total portfolio risk into an equal risk contribution of each of the component positions, see Maillard, Roncalli and Teiletche (2010). ERC has been usually implemented with the standard deviation chosen as risk measure. In the recent literature this

approach is extended by employing Conditional Value–at–Risk (CVaR) as the underlying risk measure, see Boudt, Carl and Peterson (2013) and Cesarone and Colucci (2015). For CVaR see, e.g., Rockafellar and Uryasev (2002).

The main constituents of this project are as follows: Carrying out a literature review and discussing both CVaR and ERC from the theoretical point of view, as well as summarizing the practical aspects of ERC. Subsequently comparative numerical studies are being carried out with monthly return data sets according to two aspects. On the one hand, the portfolios resulting from the CVaR–ERC approach, with– and without the assumption of normally distributed returns, are to be compared. On the other hand, an out–of–sample comparison has to be carried out, by comparing the portfolios from the CVaR–ERC approach with the equally weighted portfolio and with the portfolios obtained from the global minimum variance model and from the global minimum CVaR model. The portfolio performance measures used for this purpose should include besides the Sharpe-Ratio also some additional portfolio performance measures, not yet utilized in Cesarone and Colucci (2015). In all of the portfolio selection models short sales are to be excluded.

Literature

Boudt, K., Carl, P. and Peterson, B. (2013). Portfolio Optimization with Conditional Value–at–Risk Budgets. Journal of Risk, 15(3), 39–68.

Cesarone, F. and Colucci, S. (2015). Minimum Risk vs. Capital and Risk Diversification Strategies for Portfolio Construction. Available at SSRN: https://ssrn.com/abstract=2552455

Maillard, S., Roncalli, T., and Teiletche, J. (2010). On the properties of equally–weighted risk contributions portfolios. Journal of Portfolio Management 36(4), 60–70.

Rockafellar, R.T. and Uryasev, S. (2002). Conditional value–at–risk for general loss distributions. Journal of Banking & Finance 26, 1443–1471.

6 Topics in Machine Learning (Supervisor: Vanessa Kummer)

6.1 Testing the Robustness of the Multiple, Iterative Imputation Approach proposed by Egger, Kummer, Meusel, and Schmedders (2017)

Background

Missing data arise in almost all statistical analyses. A common approach is to delete the missing values. The problem that comes along with listwise deletion is the loss of power and sample size, as well as potential bias in the data that comes along with listwise deletion. A better

approach is therefore to fill in (impute) the missing values. An often used imputation method is mean imputation. The advantage of mean imputation is that it is easy to implement and that it does not change the mean. The drawback is that the relationships with other variables are ignored. Fortunately, there are exist other methods.

Research Question

In order to impute a data set having missing values in all variables, one can make use of unsupervised learning. In our working paper (Egger, Kummer, Meusel, and Schmedders (2017)), we describe an approach that combines multiple state-of-the-art imputation methods. So far, this method has only be applied on real-world data. It would be interesting to see though, how robust the method performs on simulated data.

Literature

Egger, A., Kummer, V., Meusel, M., & Schmedders, K. *Increasing the Value of Search Subscriptions for Housing Market Analyses.* Working Paper.