RELAÇÕES INTERNACIONAIS NO MUNDO ATUAL

CENTRO UNIVERSITÁRIO CURITIBA - UNICURITIBA - VOLUME 6 - NÚMERO 39/2023 I e-6267 I JAN-MARÇO I CURITIBA/PARANÁ/BRASIL - PÁGINAS 1 A 12 - ISSN: 2316-2880

PREREQUISITES FOR THE USE OF MACHINE LEARNING FOR BUSINESS VALUATION

Petr Koklev
Saint Petersburg State University – Russia
https://orcid.org/0000-0003-2594-7973
koklevp@gmail.com

ABSTRACT

Goal: The paper examines the fundamental theoretical prerequisites for the use of machine learning in business valuation. **Methods:** The study demonstrates that the use of statistical methods addresses the shortcomings of traditional approaches to valuation, in particular, the income approach and the discounted cash flow method. **Results:** Substantiation is given for the rejection of traditional econometric methods (linear regression, estimated by the least squares method) in favor of more complex nonparametric statistical models. **Conclusion:** Machine learning expands the empirical toolkit of the economist, allows for small datasets, solves the problem of asset valuation complexity, protects against false discoveries, and does not require compliance with Gauss-Markov assumptions. The paper also addresses the black box problem – the difficulty of interpreting models derived from statistical learning.

Keywords: Valuation; Statistical learning; DCF; Relative valuation; Econometrics; Feature importance.

PRÉ-REQUISITOS PARA A UTILIZAÇÃO DA APRENDIZAGEM DE MÁQUINAS PARA A AVALIAÇÃO DE EMPRESAS

RESUMO

Objectivo: O documento examina os pré-requisitos teóricos fundamentais para a utilização da aprendizagem de máquinas na avaliação de empresas. **Métodos:** O estudo demonstra que a utilização de métodos estatísticos aborda as deficiências das abordagens tradicionais de avaliação, em particular, a abordagem do rendimento e o método do fluxo de caixa descontado. **Resultados:** É dada uma fundamentação para a rejeição dos métodos econométricos tradicionais (regressão linear, estimada pelo método dos mínimos quadrados) em favor de modelos estatísticos não paramétricos mais complexos. **Conclusão:** A aprendizagem mecânica expande o conjunto de ferramentas empíricas do economista, permite pequenos conjuntos de dados, resolve o problema da complexidade da avaliação de activos, protege contra falsas descobertas, e não exige o cumprimento das suposições de Gauss-Markov. O documento também aborda o problema da caixa negra - a dificuldade de interpretar modelos derivados da aprendizagem estatística.

Palavras-chave: Avaliação; Aprendizagem estatística; DCF; Avaliação relativa; Econometria; Importância das características.

[Received/Recebido: Setembro 13, 2022; Accepted/Aceito Novembro 16, 2022]

1 INTRODUCTION

The consideration of machine learning (ML) is largely motivated by the many fundamental shortcomings of the discounted cash flow (*DCF*) method. The key problem is considered to be subjectivity in the identification of the main input parameters of *DCF* models: the size of expected cash flows, the discount rate, and the growth rate of cash flows. In other words, any value of the parameter can be justified by the relevant literature. As a result, the use of *DCF* cannot deliver unbiased business valuations (Kovalev & Koklev, 2022).

Understanding the key characteristics of *ML* allows us to realize the purpose of this paper – to argue for the use of *ML* for the problem of valuation.

First, it is necessary to clarify the terminology. The concept of *ML* (hereinafter, to avoid monotony, the following will be used as synonyms for ML: "statistical learning", "nonparametric statistical methods", and "data mining"*)* denotes the following (Gu et al., 2020):

1. The use of a variety of usually nonparametric statistical forecasting methods, capable of incorporating nonlinear relationships between independent variables and interaction effects and approximating process functions of any type and complexity.
2. Application of regularization – a technique that allows punishing complex models to prevent the problem of overtraining (the phenomenon when the constructed model is good at determining the value of companies from the training sample but cannot make quality predictions for companies not included in the training of the model). An ideal company valuation model has to be both complex enough to identify sophisticated patterns of relationships between financial statement items and capitalization and bounded enough to avoid reshaping. The required complexity is created through advanced methods (e.g., the Random Forest algorithm), and boundedness is achieved through a scrupulous selection of optimal values of hyperparameters of the model regularization (tree depth, number of trees in the ensemble, bootstrap sample size). A compromise between model complexity and boundedness is achieved by cross-validation.
3. The use of special algorithms to find the optimal set of hyperparameters (*hyperparameter tuning*) among the many possible specification variants. The *Grid Search* and *Random Search* algorithms are the most common in practice and are well known to users of the popular library for learning ML models – *Scikit-learn* (a library for

[Received/Recebido: Setembro 13, 2022; Accepted/Aceito Novembro 16, 2022]

ML in the *Python* programming language. URL: https://scikit-learn.org/ (accessed July 19, 2022) for *Python*. The quality of a given set of hyperparameters is assessed using a validation set obtained by *k-fold cross-validation*. In addition, testing many different combinations of hyperparameters requires efficient and fast model training. Hence, one of the key elements of statistical learning is the use of efficient algorithms that optimize the cost function, for example, stochastic gradient descent (*SGD)*.

The goal of ML is to maximize prediction accuracy in the absence of sufficient information about the process generating the data. This statement holds true for the most fruitful section of *ML*, *teacher-assisted* learning, when the researcher has information for both the independent variables *x* and the dependent variables *y*. Regression models of company value prediction belong to the section of learning with the teacher. This aspect distinguishes ML from traditional statistical approaches, where one inevitably explicit assumptions about the process are inevitable when specifying a model (Breiman, 2001). Let us elaborate on the differences between statistical training and traditional econometric methods and the main prerequisites for their use.

2 STRENGTHS AND DIFFERENCES FROM TRADITIONAL ECONOMETRIC APPROACHES

Using ML methods, we are interested in modeling the conditional distribution of some dependent variable *y* using several other variables $x = (x_1, \ldots, x_p)$. Variables x are called features, predictors, independent variables, regressors, or covariates. ML focuses on generating a qualitative prediction of *y* using *x* parameters. In the context of the business valuation problem, y is the market capitalization of an enterprise or the multiplier value. The set of *x* features may include, for example, financial reporting data and macroeconomic variables.

It may appear that an appropriate method for capitalization (or multiplier) modeling is a classical linear regression, the parameters of which are estimated by the least squares method (*LSM*). There is all reason to believe, however, that statistical learning is much more appropriate for this task (Varian, 2014). Let us review the main arguments in favor of this statement.

Expansion of the economist's empirical toolkit. A plethora of respected scholars encourages both their peers and practitioners to use *ML* as part of their financial and economic research. H. R. Varian (2014) stresses that economists must use ML to process the ever-increasing volume of data, including multivariate and unstructured

[Received/Recebido: Setembro 13, 2022; Accepted/Aceito Novembro 16, 2022]

data. *ML* offers a toolkit to significantly improve the quality of forecasting and parameter estimation (Mullainathan & Spiess, 2017). Improved predictions are achieved by accounting for non-linearities resulting from inefficiencies in the economy (Coulombe et al., 2022). S. Athey and G. W. Imbens (2017) along with S. Mullainathan and J. Spiess (2017) make the argument that statistical learning is a highly valuable approach for a wide variety of social science problems and tasks. Finally, by extending the economist's empirical toolkit, ML improves interdisciplinary communication between representatives of different sciences (Athey & Imbens, 2017).

Specialization in forecasting. In addition to focusing on the application of advanced statistical methods, ML emphasizes model verification through cross-validation and the use of a separate test sample to obtain an *unbiased* model quality assessment. Thus, the focus shifts from the traditional econometric problem of assessing parameters $\hat{\beta}$ to the result of forecasting $\hat{y}$. For a quality business valuation, the key is to determine the function *h*, which allows for a high-precision forecast of the enterprise's value. It is this prediction and not the estimation of $\hat{\beta}$ coefficients that is of great practical value. Specializing in forecasting problems, *ML* is perfectly suited to achieve the goals of enterprise valuation. The multidimensional nature of statistical learning methods makes them much more flexible than traditional econometric approaches. This flexibility contributes to a better approximation of the unknown function reflecting the company's value-forming process.

Solution to the problem of valuation complexity. Several properties of ML methods make them excellent candidates for modeling processes with unknown or indeterminate forms, to which a company's value creation process should definitely be attributed. The core property is diversity. Even within one of the many dissimilar methods, a researcher can test an infinite number of different specifications. For instance, a single hyperparameter used in many *ML* methods, the *learning rate*, can assume an infinite number of values from among positive real numbers. Additionally, the selection of the best models through cross-validation allows for controlling the overtraining problem and avoiding false discoveries that result from uncontrolled testing of many different model specification variants (Gu et al., 2020).

Possibility of application for small data sets. Despite their initial specialization, analytical methods developed for big data are particularly effective when dealing with small data sets. They are markedly superior to traditional methods, such as the *LSM* for regression problems and *logistic regression* for classification problems, which in their original form are unable to account for the non-linear impact of independent

[Received/Recebido: Setembro 13, 2022; Accepted/Aceito Novembro 16, 2022]

variables on the dependent variable. The ability to incorporate the multidimensional nature of the data is the main reason for the superiority of nonparametric AI methods.

The *LSM* still allows accounting for the non-linear impact of factors on company value and the effect of interaction. In practice, this requires expanding the matrix of plan *X* by pairwise products of predictors and by exponentiation of features. The rows of matrix *X* are made up of observations of individual companies. An observation consists of a vector, the components of which are the values of items in the company's financial statements. With this approach, however, the final number of independent variables becomes exorbitant. For example, after including the pairwise products of independent variables and taking them to the second power, we get 56,615 features for a total of 335 financial statement items. 335 features is the approximate number of financial statement items based on the eight most recent quarterly reports of the three major reporting forms as independent variables. With such a large number of features, it is virtually impossible to extract the signal from a small number of observations and avoid overtraining. In the case of raising to the third power and including pairwise products for powers below three, the total number of columns in matrix *X* will exceed six million. Such a large number of degrees of freedom will considerably slow down the process of determining the coefficients and will also unavoidably result in overtraining. The resulting model will thus be unsuitable for business valuation.

Interpretability. As noted above, a disadvantage of traditional methods is the tendency to overtrain. An *LSM* model tends to make extremely inaccurate predictions outside of the training sample. Interpretability ceases to be an advantage of linear regression. The use of Tikhonov regularization (*Ridge*) (Tikhonov, 1963), *Lasso* regression (Tibshirani, 1996), as well as their combination – *Elastic Net* regularization improves the quality of forecasting while retaining the possibility of a simple interpretation of the estimated coefficients. Given all the disadvantages, some experts believe that the use of the *LSM* in social sciences has to be minimized and can be replaced by two types of models. The first type is ensembles of the most advanced ML techniques for the purpose of high-fidelity prediction. The second type is linear models with penalties (*Ridge, Lasso,* or *Elastic Net*) to achieve high interpretability (Hindman, 2015). It can be demonstrated that aside from obtaining sparse models with few nonzero coefficients, traditional statistical tests can also be used for regression with a penalty, including the construction of confidence intervals and obtaining p-values (Lockhart et al., 2014; Park & Casella, 2008). The *LSM* does not allow making sparse models. Each independent variable included in the specification will receive a nonzero

[Received/Recebido: Setembro 13, 2022; Accepted/Aceito Novembro 16, 2022]

coefficient. With a large number of features, this makes it much more difficult to interpret the estimated model.

Correlation of features. The explanatory variables used in business valuation, financial statement items, are often similar to each other and strongly correlated. Examples of correlated variables are revenue and gross profit. A more extreme example is the balance sheet total as of the last quarterly report date and the balance sheet total as of the penultimate quarterly report date. A correlation heatmap for a sample of some financial statement items based on data from US stock market companies is presented in Figure 1. A consequence of unrestricted multicollinearity is a high standard error value for the estimated coefficients, which devoids the *LSM* of the main reason to use it – interpretability of the assessed indicators. Furthermore, the *LSM* stops working when the number of variables exceeds the number of observations (the $X^T X$ matrix becomes irreversible). The use of *ML* allows one to successfully work with data that is pathological for traditional methods and obtain viable models.

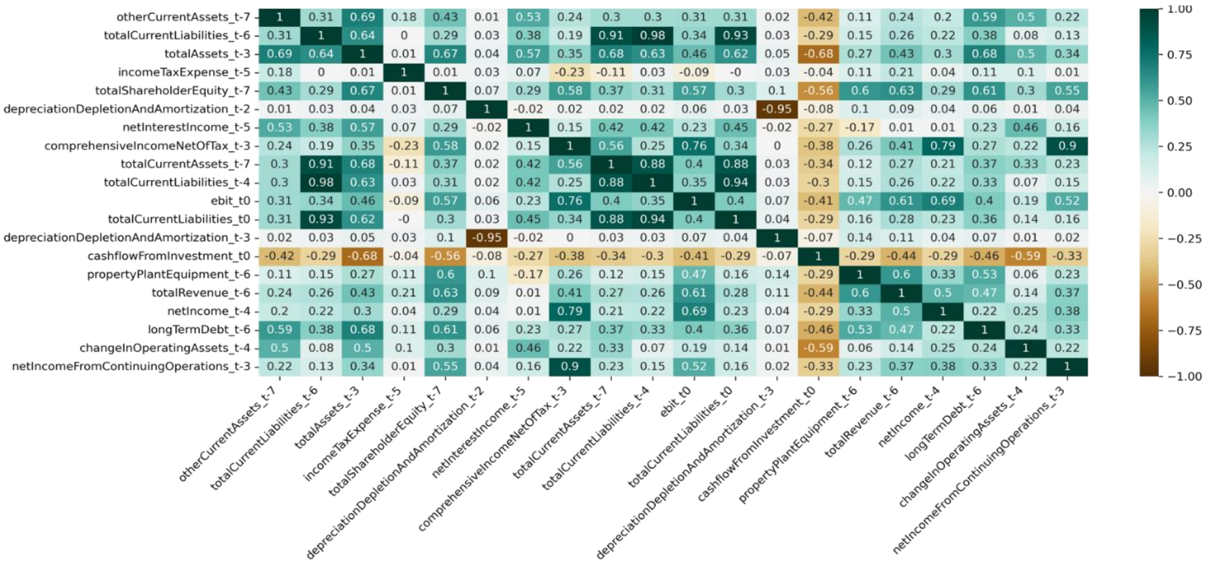

Figure 1. Correlation heatmap for twenty random financial statement items

As follows from the data in Figure 1, many predictors have an extremely high absolute value of the Pearson correlation coefficient, which negatively affects the interpretation of the coefficients estimated by the LSM.

Protection from false discoveries. Another key component of ML is the use of *cross-validation* and a separate test set that can only be used once to verify model quality. This aspect prevents an unscrupulous researcher from testing four million regression model specifications, selecting one of them that supports any desired hypothesis, and publishing a paper with a false result (de Prado, 2017; Sala-i-Martin, 1997). The use

[Received/Recebido: Setembro 13, 2022; Accepted/Aceito Novembro 16, 2022]

of cross-validation and test sets makes the results and conclusions robust. Thus, statistical learning offers tools not only for high-quality forecasting but also for obtaining reproducible results.

Violation of Gauss-Markov assumptions. The process of company value creation cannot be a simple linear combination of reporting items. Clearly, the use of financial reporting documents leaves out many other extremely relevant variables that determine the capitalization or value of the company. The coefficients obtained through the *LSM* can only be unbiased if the model specification includes all the independent variables relevant to the considered phenomenon. The presence of assumptions and the introduction of restrictions on the functional form do increase the interpretability of the results. However, this option is fraught with a major risk of the inaccurate model specification (Jung et al., 2018). This way, the interpretability of an *LSM* model may be false and meaningless, describing a nonexistent reality.

None of the possible economic data sets can comprise all the variables relevant to a phenomenon. In traditional statistical analysis, key features are almost always left out because of the incorrect model specification (de Prado, 2018b, 2020b). Given the complexity of financial systems, even if the researcher somehow managed to determine all the independent variables relevant to the process, they would still be unable to determine the exact functional form of the phenomenon (de Prado, 2018a).

The use of ML eliminates the need for unrealistic assumptions about the process that generates the data. Statistical learning automatically accounts for the relevant effects and nonlinear relationships. Numerous empirical problems, including company valuation, do not have enough theoretical grounds for any specific functional form. The use of *ML* requires only a list of possibly relevant factors that can be useful for business valuation (Jacobsen et al., 2016). In this case, the researcher does not need to specify a particular functional form when specifying the model.

The normality of the *LSM* model errors, another Gauss-Markov assumption, is also out of the question. The Jarque-Bera normality test statistics (Jarque & Bera, 1980) of the linear model to predict capitalization based on financial statements for US stock market companies data amounts to 440.65 and the corresponding p-value equals zero.

3 LIMITATIONS OF ML

Having great potential for enterprise valuation, ML still has some disadvantages and limitations. The predictions obtained by applying the models are measurements. These

[Received/Recebido: Setembro 13, 2022; Accepted/Aceito Novembro 16, 2022]

measurements alone are not indicative of the fundamental mechanisms that shape the market capitalization of a company. In addition, the process of assigning a value can itself seem nontransparent. Even specialists sometimes find it rather difficult to use plain language to describe the logic behind the formation of a particular prediction. The difficulty of communication is one of the barriers to applying ML to some financial problems. The compromise between accuracy and explainability is a well-known phenomenon in the literature. Improvement of forecast quality with complex, nonlinear models is inevitably coupled with a decrease in interpretability (Breiman, 2001). Compared to econometric publications, a paper on the application of ML will have fewer theoretical results accompanied by formal statistical tests. Fortunately, significant progress has already been made in applying statistical tests to models derived from *Data Mining* approaches (Farrell et al., 2021; Wager & Athey, 2018).

Contrary to the popular belief that ML is a black box suitable only for prediction, there are tools that allow proposing and testing hypotheses, including in financial research (Lommers et al., 2021). Whether or not AI methods are a black box depends entirely on the subject applying them and their background, knowledge, and skills (de Prado, 2020a).

ML algorithms detect patterns in multidimensional space. These patterns relate features (company characteristics) to the outcome (cost). The nature of these relationships can be incredibly complex. By using methods for assessing the importance of features, it is possible to determine which covariates are most important to the model. As wittily noted by M. L. de Prado (2018c), ML will not analytically derive the gravitational attraction formula of Newton's classical gravitation theory but it will determine that mass and distance are the key factors.

Special attention is now being paid to the problem of the interpretability of results. Modern methods of assessing *feature importance* solve the problem of interpretability to a large extent (Carvalho et al., 2016). *Partial Dependence Plots* demonstrate the marginal effect of features on the predicted variable. *Permutation Importance* is another simple and intuitive way to determine the importance of features. Recently, one of the most popular methods has come to be SHAP, which estimates the local contribution of the value of the analyzed feature to the final forecast of the model. Given recent progress in the field of interpretive AI, the term black box does not seem appropriate anymore. By now, the use of feature importance methods is considered a good rule of thumb and is a necessary component of any study using *ML* methods.

There are several barriers to the broad adoption of ML in finance. This is largely

[Received/Recebido: Setembro 13, 2022; Accepted/Aceito Novembro 16, 2022]

explained by inertia. Economic scholars have a well-established toolkit for forecasting and are skeptical of new and unfamiliar approaches. Some mistakenly believe that *ML* is blindly combing through data, although ML is the exact discipline that has developed many ways to combat data overfitting (overtraining). *ML a*ddresses this problem explicitly and on a fundamental level.

There is also an opinion that ML is a substitute for real knowledge of the subject (economics). This thesis appears to be far-fetched. Understanding *domain knowledge* is a prerequisite for developing high-quality statistical learning models (Munkhdalai et al., 2019). Formulation of the hypothesis, the choice of the phenomenon under study, i.e. the object of application of AI methods, its studied characteristics, the selection of relevant predictors, the construction of a training data set, the identification of the most appropriate preliminary data processing methods – all this requires deep domain knowledge.

The real significant disadvantage seems to be the impossibility of applying ML to implement the *VBM* (value-based management) concept in practice. For *VBM* tasks, one of the key requirements for models is the criteria of understandability and explainability (Volkov, 2004). At this stage, we should not expect the top management of corporations to have the proper training to understand the principles of *ML*. As previously stated, understandability and explainability depend on the qualifications of the subject interpreting the model. At present, ML methods are unfit to communicate the ideas of value-based company management. This, however, is not an intrinsic limitation of ML methods but rather a limitation of the manager's qualification. The situation is always changing, and there is reason to expect that due to an increase in the share of personnel skilled in data analysis, methods based on statistical modeling will come in demand within the concept of value-based analysis (*VBA*). It is difficult to trace the connection, the mechanism that resulted in a particular value of the enterprise. As a result, it is difficult to justify depriving a manager of a premium because of a decrease in the value of the company established through an accurate but non-transparent model. ML can be utilized as an auxiliary *VBA* instrument, yet the communication itself needs to be conducted with a different approach. This kind of discussion of the problem of choosing the valuation model for the purpose of practical implementation of *VBA* can be found in the works of D. L. Volkov (2004). *Residual Income Models* are believed to be the most suitable for VBA ideas (Edwards & Bell, 965; Ohlson, 1995).

[Received/Recebido: Setembro 13, 2022; Accepted/Aceito Novembro 16, 2022]

4 CONCLUSION

The use of statistical methods as part of a mass valuation overcomes the fundamental disadvantages of the DCF method that make the *DCF* unfit for obtaining an unbiased business valuation. Specializing in the tasks of forecasting $\hat{y}$ instead of determining and interpreting $\hat{\beta}$ parameters, ML methods are ideal for the valuation of a company, allowing one to simulate the complex process of enterprise value creation due to their ability to incorporate a complex and non-linear relationship between financial reporting data and market capitalization. The existing methods of assessing the importance of indicators help to solve the black box problem and obtain interpretable models.

The existence of numerous studies documenting successful applications of ML to a wide variety of economic and financial problems suggests that *ML* has great potential for business valuation as well (Ksenofontova et al., 2017).

REFERENCES

Athey, S. & Imbens, G.W. (2017). The state of applied econometrics: Causality and policy evaluation. *Journal of Economic Perspectives, 31*(2), 3-32. http://dx.doi.org/10.1257/jep.31.2.3

Breiman, L. (2001). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical Science, 16*(3), 199-231. http://dx.doi.org/10.1214/ss/1009213726

Carvalho, D.V., Pereira, E.M., & Cardoso, J.S. (2019). Machine learning interpretability: A survey on methods and metrics. *Electronics, 8*(8), 832. http://dx.doi.org/10.3390/electronics8080832

Coulombe, P.G., Leroux, M., Stevanovic, D., & Surprenant, S. (2022). How is machine learning useful for macroeconomic forecasting? *Journal of Applied Econometrics, 37*(5), 920-964. https://doi.org/10.1002/jae.2910

Edwards, E.O., & Bell, P.W. (1965). *The theory and measurement of business income*. Berkeley, Los Angeles: University of California Press.

Farrell, M.H., Liang, T., & Misra, S. (2021). Deep neural networks for estimation and inference. *Econometrica, 89*(1), 181-213. http://dx.doi.org/10.3982/ECTA16901

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies, 33*(5), 2223-2273. http://dx.doi.org/10.1093/rfs/hhaa009

Relações Internacionais do Mundo Atual Unicuritiba.
[Received/Recebido: Setembro 13, 2022; Accepted/Aceito Novembro 16, 2022]

Hindman, M. (2015). Building better models: Prediction, replication, and machine learning in the social sciences. *The Annals of the American Academy of Political and Social Science, 659*(1), 48-62. http://dx.doi.org/10.1177/0002716215570279

Jacobsen, J.P., Levin, L.M., Tausanovitch, Z. (2016). Comparing standard regression modeling to ensemble modeling: How data mining software can improve economists' predictions. *Eastern Economic Journal, 42*(3), 387-398. http://dx.doi.org/10.1057/eej.2015.8

Jarque, C.M., & Bera, A.K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters, 6*(3), 255-259. https://doi.org/10.1016/0165-1765%2880%2990024-5

Jung, J.-K., Patnam, M., & Ter-Martirosyan, A. (2018). An algorithmic crystal ball: Forecasts-based on machine learning. International Monetary Fund Working Papers 18(230). http://dx.doi.org/10.5089/9781484380635.001

Kovalev, V.V., & Koklev, P.S. (2022). Nedostatki dokhodnogo podkhoda dlia otsenki biznesa ["Disadvantages of the income approach to business valuation]. *Economic Sciences, 9(*214), 49-54. https://doi.org/10.14451/1.214.49

Ksenofontova, T.Y., Bezdudnaya, A.G., & Kadyrova, O.V. (2017). Basic problems of interregional differentiation in Russia and innovative and reproduction prerequisites to overcome them. *International Journal of Applied Business and Economic Research, 15*(8), 1-10.

Lockhart, R., Taylor, J., Tibshirani, R.J., & Tibshirani, R. (2014). A significance test for the lasso. *Annals of Statistics, 42*(2), 413-468. http://dx.doi.org/10.1214/13-AOS1175

Lommers, K., El Harzli, O., & Kim, J. (2021). Confronting machine learning with financial research. *The Journal of Financial Data Science, 3*(3), 67-96. http://dx.doi.org/10.3905/jfds.2021.1.068

Mullainathan, S., & Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives, 31*(2), 87-106. http://dx.doi.org/10.1257/jep.31.2.87

Munkhdalai, L., Munkhdalai, T., Namsrai, O.-E., Lee, J.Y., & Ryu, K.H. (2019). An empirical comparison of machine-learning methods on bank client credit assessments. *Sustainability, 11*(3), 699. http://dx.doi.org/10.3390/su11030699

Ohlson, J.A. (1995). Earnings, book values, and dividends in equity valuation. *Contemporary Accounting Research, 11*(2), 661-687. https://doi.org/10.1111/J.1911-3846.1995.TB00461.X

[Received/Recebido: Setembro 13, 2022; Accepted/Aceito Novembro 16, 2022]

Park, T., & Casella, G. (2008). The Bayesian lasso. *Journal of the American Statistical Association, 103*(482), 681-686. http://dx.doi.org/10.1198/016214508000000337

de Prado, M.L. (2017). Finance as an industrial science. *Journal of Portfolio Management, 43*(4), 1-7.

de Prado, M.L. (2018a). The 10 reasons most machine learning funds fail. *The Journal of Portfolio Management, 44*(6), 120-133. http://dx.doi.org/10.3905/jpm.2018.44.6.120

de Prado, M.L. (2018b). The myth and reality of financial machine learning (presentation slides). *SSRN Electronic Journal.* http://dx.doi.org/10.2139/ssrn.3120557

de Prado, M.L. (2018c). Ten financial applications of machine learning (seminar slides). *SSRN Electronic Journal.* http://dx.doi.org/10.2139/ssrn.3197726

de Prado, M.L. (2020a). *Machine learning for asset managers*. Cambridge: Cambridge University Press.

de Prado, M.L. (2020b). Three machine learning solutions to the bias-variance dilemma (seminar slides). *SSRN Electronic Journal*. http://dx.doi.org/10.2139/ssrn.3588594

Sala-i-Martin, X.X. (1997). I just ran four million regressions. *American Economic Review, 87*(2), 178-183.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological), 58*(1), 267-288. https://doi.org/10.1111/J.2517-6161.1996.TB02080.X

Tikhonov, A.N. (1963). O reshenii nekorrektno postavlennykh zadach i metode reguliarizatsii ["On solving incorrectly set problems and the regularization method]. *Proceedings of the USSR Academy of Sciences, 151*(3), 501-504.

Varian, H.R. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives, 28*(2), 3-28. http://dx.doi.org/10.1257/jep.28.2.3

Volkov, D.L. (2004). Upravlenie stoimostiu kompanii: Problema vybora adekvatnoi modeli otsenki ["Managing the value of the company: The problem of choosing an adequate valuation model]. *Vestnik of Saint Petersburg University. Management, 4*(32), 79-98.

Wager, S., & Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association, 113*(523), 1228-1242. http://dx.doi.org/10.1080/01621459.2017.1319839

Relações Internacionais do Mundo Atual Unicuritiba.
[Received/Recebido: Setembro 13, 2022; Accepted/Aceito Novembro 16, 2022]