DigitalCommons @WayneState

Journal of Modern Applied Statistical Methods

Volume 17 | Issue 2 Article 12

3-26-2019

A Random Forests Approach to Assess Determinants of Central Bank Independence

Maddalena Cavicchioli
University of Verona, maddalena.cavicchioli@univr.it

Angeliki Papana
Aristotle University of Thessaloniki, angeliki.papana@gmail.com

Ariadni Papana Dagiasis
Cleveland State University, a.papanadagiasis@csuohio.edu

Barbara Pistoresi
University of Modena and Reggio Emilia, barbara.pistoresi@unimore.it

Recommended Citation

Cavicchioli, M., Papana, A., Dagiasis, A. P., & Pistoresi, B. (2018). A random forests approach to assess determinants of central bank independence. *Journal of Modern Applied Statistical Methods, 17*(2), eP2611. doi: 10.22237/jmasm/1553610953

A Random Forests Approach to Assess Determinants of Central Bank Independence

Cover Page Footnote
M. Cavicchioli and B. Pistoresi gratefully acknowledge financial support from PRIN-MIUR Grant 2010J3LZEN-003 and the research grant FAR 2014 (provided by University of Modena and Reggio E., Italy).

Journal of Modern Applied Statistical Methods
November 2018, Vol. 17, No. 2, eP2611
doi: 10.22237/jmasm/1553610953

ISSN 1538 − 9472

A Random Forests Approach to Assess Determinants of Central Bank Independence

Maddalena Cavicchioli
University of Verona
Verona, Italy

Angeliki Papana
Aristotle University of Thessaloniki
Thessaloniki, Greece

Ariadni Papana Dagiasis
Cleveland State University
Cleveland, OH

Barbara Pistoresi
University of Modena and Reggio Emilia
Modena and Reggio Emilia, Italy

A non-parametric efficient statistical method, Random Forests, is implemented for the selection of the determinants of Central Bank Independence (CBI) among a large database of economic, political, and institutional variables for OECD countries. It permits ranking all the determinants based on their importance in respect to the CBI and does not impose *a priori* assumptions on potential nonlinear relationships in the data. Collinearity issues are resolved, because correlated variables can be simultaneously considered.

Keywords: Central bank independence, determinants, collinearity, random forests, minimal depth

Introduction

Large panels of economic and financial data are becoming the starting point for empirical analysis. Data manipulation tools and techniques, developed for small datasets, will become increasingly inadequate. One of these techniques is (linear) regression analysis, and although it is widely applied in economic works, it shows some drawbacks. The selection of the relevant group of variables from a large dataset might bring limitations due to the omitted variables issue and overfitting. It is mainly linear and, when necessary, nonlinear terms and interactions should be modelled *a priori* in the parametric model imposing the functional form. Collinearity problems are addressed by excluding linearly correlated variables. Missing data values are common obstacles that traditional methods cannot handle.

doi: 10.22237/jmasm/1553610953 | Accepted: January 21, 2018; Published: March 26, 2019.
Correspondence: Angeliki Papana, angeliki.papana@gmail.com

An example of empirical research subject to the limitations above is the determinants of Central Bank Independence (CBI). A higher degree of CBI is associated with lower inflation rates, so that society reduces opposition to inflation and public pressure for an independent central bank (Cukierman, 2008, 2013; Alesina & Stella, 2011), and the political economy of monetary policymaking (DeHaan & Eijffinger, 2016). The balance between flexibility and credibility in monetary policymaking determines the equilibrium degree of CBI in a country. The trade-off between costs and benefits in delegating the power to manage paper money to reduce inflationary bias may depend on many aspects of the economy and on its institutional framework (Alesina & Grilli, 1995). This has encouraged the study of the determinants that influence the CBI among a large variety of economic and institutional variables which cause changes in the degree of commitment of the monetary policy (Fernández-Albertos, 2015; D'Amato, Pistoresi, & Salsano, 2009; Farvaque, 2002). The usual way to obtain predictions on CBI is based on the linear regression framework, given that the economic theory does not provide any structural modelling. Furthermore, facing an increasing complexity of the available data, this method faces a lot of problems such as omitted variables and overfitting, as recently pointed out by Brumm (2011).

Machine learning has ways to deal with large databases (Varian, 2014). Examples include boosting, support vector machines, AdaBoost, genetic algorithms (Creamer & Freund, 2010; Emsia & Coskuner, 2016; Gogas, Papadimitriou, Matthaiou, & Chrysanthidou, 2015; Zhou & Lai, 2017; Zhang & Maringer, 2016), and Random Forests (RFs). An RF is an ensemble learner formed by averaging binary tree predictors (Ho, 1995; Breiman, 2001). They are grown non-deterministically, without pruning, using a two-step randomization procedure and thus resulting in reduced bias. Breiman's RF algorithm was developed for classification and regression settings with a variety of applications (Breiman, 2001; Cutler et al., 2007). A Random Survival Forest (RSF) is an extension of Breiman's RF methodology that can be used for building a prediction model in survival analysis (Ishwaran, Kogalur, Blackstone, & Lauer, 2008). In survival settings, the predictor is an ensemble formed by combining the results of many survival trees. The base learner is a survival tree and the ensemble is a cumulative hazard function formed by averaging each tree's Nelson-Aalen's cumulative hazard function. Ishwaran, Kogalur, Gorodeski, Minn, and Lauer (2010) developed a high-dimensional variable selection method based on minimal depth, which avoids directly working with prediction error and relies on a theoretical basis.

RF provides a theoretically-justified variable importance measure and a threshold to select and rank predictors. It provides the flexibility to uncover complex data structures, such as dealing with nonlinear effects and interactions among multiple types of variables while allowing for effective predictions due to the law of large numbers and its two-step randomization approach. Moreover, collinearity issues can be handled, since correlated variables can be considered in the analysis. Finally, missing values can be dealt automatically, and the methodology is applicable to high dimensional settings.

The problem of variable selection and prediction of CBI is approached via the RF method, borrowing features from RSF, such as the minimal depth, in order to overcome the drawbacks of linear regression. Particularly, the determinants of CBI are selected among 58 economic and institutional variables (see the Appendix for a detailed description of the variables), for 24 OECD economies, allowing any kind of complexity (e.g., nonlinear relationships).

Methodology

In the case of a regression setting (e.g. problems where the response variable is a quantitative variable), RSF and RF overlap.

RF Methodology

RF trees are formed as described:

Step 1. Draw *B* bootstrap samples from the original sample. Bootstrap samples exclude on average 37% of the data, known as Out-of-Bag (OOB) data.

Step 2. Grow a tree based on the data of each of the bootstrap samples $b = 1, \ldots, B$.

- (a) At each tree node, randomly select a subset of predictor variables on which to split.
- (b) Among all binary splits defined by the predictors selected in (a), find the best split into two subsets (the daughter nodes) according to a suitable splitting criterion.
- (c) Repeat (a), (b) recursively on each daughter node until a stopping criterion is met.

Step 3. Aggregate information from the terminal nodes (nodes with no further split) from the *B* trees to obtain a prediction ensemble (predict new data by aggregating the predictions of all trees, i.e. majority votes for classification, average for regression).

Node splits are a very important step in the algorithm, determined by the setting or type of response. For example, in the case of regression or multivariate analysis, the default rule is the weighted Mean-Squared Error (MSE) (Breiman, Friedman, Stone, & Olshen, 1984, chapter 8.4) or a composite normalized MSE, respectively. For classification analysis, the rule is the Gini index (Breiman et al., 1984, chapter 4.3). For mixed outcomes analysis, a multivariate normalized composite split rule of MSE and Gini index splitting is invoked. For survival analysis, a log-rank splitting rule is implemented (Segal, 1988; Leblanc & Crowley, 1993). For competing risk analysis, a modified weighted log-rank splitting rule, modeled after Gray's test (Gray, 1988), is implemented. In our case, the MSE was used as a splitting criterion.

An estimate of the prediction error can be obtained based on the OOB data as follows:

a. At each bootstrap iteration, predict the OOB data using the tree grown with the bootstrap sample.
b. Aggregate the OOB predictions. Calculate the error rate and call it the OOB estimate of error rate (measured via mean-squared-error for regression, misclassification error for classification, 1-Harrell's concordance index for survival).

Variable Selection

In addition to the good prediction performance of the RF methods (RF and RSF), they are useful tools for variable selection. They provide measures of variable importance (VIMP), calculated for each predictor, so that variables are selected by filtering on the basis of their VIMP. The VIMP of a variable is measured by the change in the prediction error for the forest ensemble, when OOB data for that variable is permuted, while all others are left unchanged. Although there are several RF-based methods utilizing VIMP for variable selection, most of these procedures are limited. Drawbacks of VIMP-like methods are the following: first, they are dependent on the type of prediction error used; and second, a theoretical justification is not available.

In contrast, an alternative way to calculate VIMP, along with a theoretical justification for this new variable selection framework, was introduced by Ishwaran and Kogalur (2007) and Ishwaran et al. (2008). For VIMP calculation, the variable is not permuted. An OOB case is assigned a daughter node randomly whenever a split on this variable is encountered in the in-bag tree. The VIMP of a variable is then the prediction error for the original forest subtracted from the prediction error for the noised-up forest predictor. In both cases, positive VIMP values (and especially largest values) indicate predictive variables. A rigorous approach for variable selection via the concept of the maximal subtree and the statistic minimal depth is presented. Let T be a binary recursively grown tree; that is, if T has M terminal nodes, then T is the function that maps multivariable covariates into those terminal nodes.

Definition 1. For each variable v, call Tv a v-subtree of T if the root node of Tv is split using v. Call Tv a maximal v-subtree if Tv is not a subtree of a larger v-subtree.

Definition 2. A second order maximal (w, v)-subtree is a maximal w-subtree within a maximal v-subtree for a variable v.

Definition 3. Let D_v be the distance from the root node to the root of the closest maximal v-subtree for a given v. Then D_v takes values $\{0,\ldots, \mathrm{D}(T)\}$, where $\mathrm{D}(T)$ is the depth of T (distance from the root farthest terminal node). D_v is called the minimal depth of variable v.

This means that a maximal subtree of a variable v is defined to be the largest subtree whose root node is split using v and no other parent node of the subtree is split using v. Maximal subtrees can be used to quantify the predictiveness of a variable as well as identify variable interactions.

Theorem 1. Let D_v be the minimal depth of v and $\pi_{v,j}(t)$ be the probability that v is selected as candidate variable for splitting a node t of depth j, assuming no maximal v-subtree exists at depth less than j. Let $\theta_{v,j}(t)$ be the probability that v splits a node t of depth j given that v is a candidate variable for splitting t and no maximal v-subtree exists at depth less than j. Then for depth $d \in \{0,\ldots, \mathrm{D}(T) - 1\}$, it holds

$$\mathrm{P}\left\{D_v = d \mid l_0,\ldots,l_{\mathrm{D}(T)-1}\right\} = \left[\prod_{j=0}^{d-1}\left(1-\pi_{v,j}\theta_{v,j}\right)^{l_j}\right]\left[1-\left(1-\pi_{v,d}\theta_{v,d}\right)^{l_d}\right], \qquad (1)$$

where l_d is the number of nodes at depth d.

The minimal depth of a maximal subtree or first order depth (for simplicity, depth) equals the shortest distance from the root node to the parent node of the maximal subtree. The second order depth is the distance from the root node to the second closest maximal subtree of that variable. The depth measures the predictiveness of a variable. The smaller the minimal depth, the more impact a variable has on prediction. Additionally, the mean of the minimal depth distribution is used as the threshold value for deciding whether a variable's minimal depth value is small enough for the variable to be classified as strong.

Mean Minimal Depth Threshold Rule

Choose a variable v if its forest-averaged minimal depth, $D_v = d_v$, is less than or equal to the mean minimal depth of D_v when v is a noisy variable. The advantage of working with the methodology of maximal subtrees and their statistics is that they are dimensionless, they are free to any type of prediction error, and they apply to all forests settings (survival, classification, and regression). It was extended to high-dimensional data. The minimal depth distribution in equation (1) was studied under various scenarios and its fairly robust threshold value for identifying strong variables imposes an automated variable selection method.

Random Survival Forests-Like Strategy

Although a regression setting is examined, this approach could easily be adopted in alternative settings. The following steps are implemented with the R software package randomSurvivalForest:

a. Grow a random forest to yield regression using a weighted mean squared error splitting rule
b. Implement a random forest variable selection using a tree minimal depth methodology to rank all variables in terms of importance in the model
c. Find the threshold value as the mean of the minimal depth distribution for selecting the most predictive variables in the RF model; that is variables with minimal depth smaller than the calculated threshold

Data and Economic Framework

Theoretical models on CBI suggest a large set of characteristics of a nation's economy that cause changes in the degree of independence of the monetary policy, e.g. inflation rate, size of the economy, GDP per capita, and various measures of efficiency of institutions. Moreover, the inflationary bias has specific features in open economies due to the interdependence in the stabilization monetary policy. In this context, international business cycle synchronization, the degree of openness, and exchange regime are also important (D’Amato et al., 2009). The political economy approaches suggested this institutional innovation should be understood, not only in terms of economic efficiency, but also in terms of political convenience, adding other potential determinants of CBI. Politically heterogenous contexts, such as systems of checks and balance, federal systems, or coalition or multiparty governments have incentive to delegate to an independent monetary authority (see Fernández-Albertos, 2015). This wide literature has encouraged the study of the determinants that influence CBI among the variety of economic, social, and institutional variables that cause changes in the degree of commitment of the monetary policy.

Consider the predictors of CBI for 24 OECD economies. In the considered period, the countries are: Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Italy, Japan, Korea, Luxembourg, the Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, the UK, and the US, over the period 1980-2003. CBI is measured by the legal independence index of Cukierman (1992) and updated by Polillo and Guillén (2005) until 2003. The 58 predictors are economic, political and institutional determinants of the CBI.

The economic variables are taken from the International Monetary Fund (2009) and the World Bank Group (2008) databases and include: two proxies for the world-wide component in the business cycle (i.e. the correlation between the GDP growth in each country and the world or US GDP growth); a dummy EMU variable, taking the value one in the 1998-2003 period for the countries that joined the European Monetary Union after complying with the convergence criteria provided for by the Maastricht Treaty; different measures of inflation (i.e. current/past inflation and past average inflation); and various measures of development (i.e., log real GDP total and real GDP growth rate). For the world GDP growth, a weighted average of the growth rates of the economies in the sample was used, with weights equal to the GDP levels in each country. The average GDP correlation 1960-79 was used to analyze the CBI 1980-91 to reduce the endogeneity problem, and the average GDP correlation 1980-92 to study the CBI 1992-2003. Moreover, past average inflation of 1960-79 and 1980-92 was used to analyze the CBI 1980-91 and CBI 1992-2003, respectively.

The political and institutional variables are from the DPI2006 database of World Bank by Keefer (2007). This database is divided into five different groups: 1) Chief Executive variables containing variables that are relevant to characterize the executive power (e.g., whether countries are presidential or parliamentary; number of years in office of the chief executive); 2) Party variables in the legislature including those variables relevant to the parties that make up the legislative power (e.g., government fractionalization, the number of government/opposition seats;

average age of parties); 3) Electoral rules that are relevant to the electoral rules (e.g., mean district magnitude; if plurality system; proportional representation); 4) Stability and checks and balances, containing relevant variables to the stability of the political system (e.g., longest/shorter tenure of a veto players) 5) Federalism, including variables relevant to the state form (e.g., if there are autonomous regions or municipal governments locally elected and whether it is a federal state).

The political and institutional variables used in the empirical analysis are listed below. A detailed definition of each variable can be found in the DPI2006 by Keefer (2007). The same definitions hold here when commenting on the results. 1) Chief Executive variables: system, yrsoff, finittrm, yrcurnt, multpl, allhouse; 2) Party variables in the legislature: herfgov, govfrac, numgov, numvote, gov1seat, gov1vote, gov2seats, gov2vote, gov3seats, gov3vote, govothst, herfopp, oppfrac, numopp, oppvote, opp1seat, opp1vote, opp2seat, opp2vote opp3seat, opp3vote, oppothst, herftot, frac, oppmajh, maj, partyage, exelec, execspec, govspec, coalspec; 3) Electoral rules: liec, mdmh, plurality, pr, housesys; 4) Stability and checks and balance: tenlong, tenshort, tenshortlax, checks, stabs, stabns; 5) Federalism: auton, federal.

To select the determinants of CBI, the RF is applied. There are few suggestions from the economic theory to properly model the underlying relationships. Here, minimal depth is used to assess a variable's predictiveness. A built-in threshold, which is independent of *a priori* tuning of parameters, is provided for variable selection. Computations were implemented using the freely available R software package randomForestSRC (RStudio, 2015; Ishwaran & Kogalur, 2007).

Results

Reported in Table 1 are the ranked minimal depth (depth) of the selected variables based on the RF analysis of the CBI dataset. The model size turns out to be 10, for depth threshold equal to 6.08. Another selected variable is time, which indicates the temporal variation of some macro-economic variables and it is not commented on further. Variable importance (vimp) is also provided (as introduced above), confirming depth ordering. The cumulative contribution of the selected variables is computed from the normalized variable importance (vimpnorm) (see Grömping, 2009). The variable emu, which mainly reflects a change in the institutional design of monetary policy, is a major determinant (41% of the total variation in CBI), past average inflation, business cycles synchronization, and the degree of development explain 16%, 19%, and 4% of the CBI variation, respectively, while political variables account for the remaining 20% of the CBI variation.

The marginal effect of each selected variable on CBI is examined in Figures 1 and 2. The vertical axis displays the predicted CBI, while each predictor is plotted on the horizontal axis. The dummy emu suggests that participation in the Euro (emu equal to 1 in Figure 1) implies a greater CBI: it encourages the individual countries to change the institutional design of the monetary policy in view of greater price stability.

Table 1. Minimal depth and variable importance (denoted by vimp, vimpnorm when normalized) obtained from RSF analysis of CBI dataset; the latter includes 59 variables for 24 OECD countries and spans from 1980 to 2003, as described in a previous section; the selection of 10 variables comes from a depth threshold equal to 6.08

Series	depth	vimp	vimpnorm
emu	1.4440	0.0280	0.4080
averageinfl	3.0120	0.0110	0.1580
world cycle	3.3400	0.0070	0.1000
usa cycle	3.6230	0.0060	0.0920
mdmh	4.2590	0.0050	0.0690
numgov	4.7050	0.0040	0.0510
logrealgdp	5.3920	0.0020	0.0350
gov1seat	5.4820	0.0030	0.0360
numopp	5.5540	0.0020	0.0260
opp1seat	6.0010	0.0020	0.0230

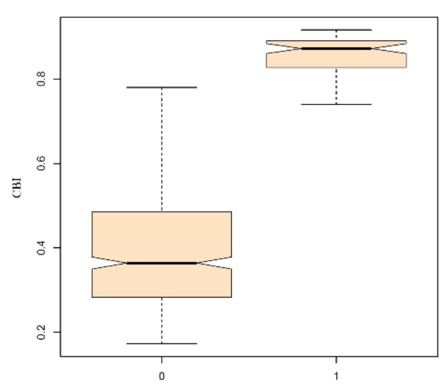

Figure 1. Marginal effect plot of the dummy variable emu on the CBI index

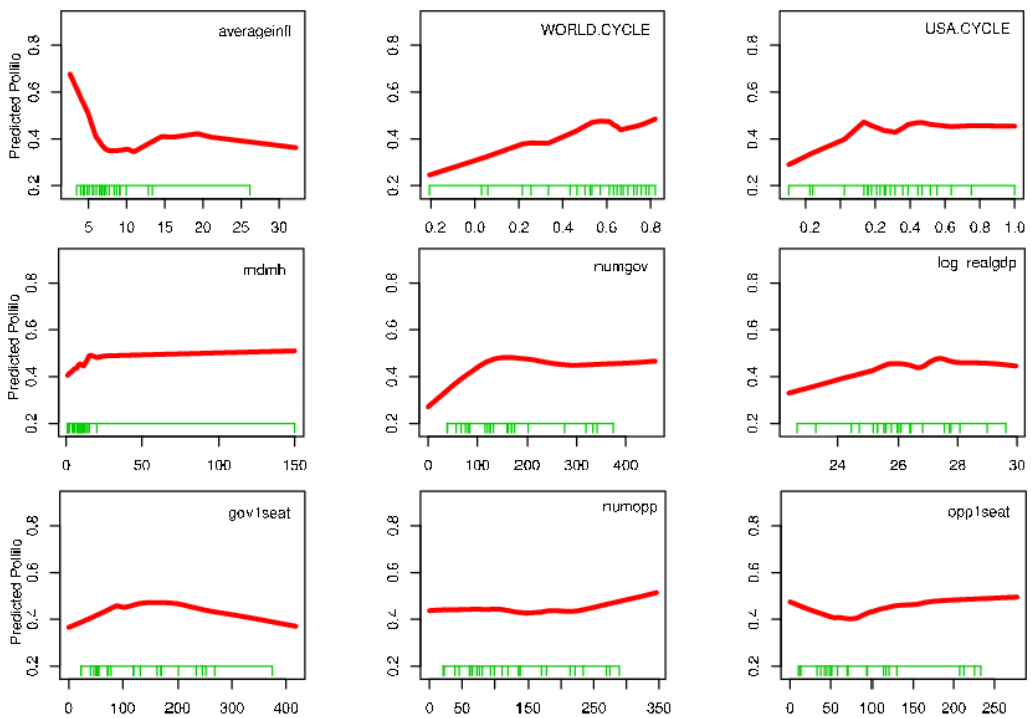

Figure 2. Marginal effect plots of the predictor variables avergeinfl, WORLD CYCLE, USA CYCLE, logrealgdp, mdmh, numgov, numopp, gov1seat, opp1seat on the predicted CBI index shown by the curves; each country's mean predictor value is displayed with vertical line marks at the horizontal axis

---

The predicted smoothed curves in Figure 2 suggest the prevalent positive or negative behavior of the selected variables towards CBI given the mean predictor values of each country (ticks at the bottom). Nonlinearity of the variable past average inflation (denoted by averageinfl) is recognized; in fact, the relation with CBI stays negative when the level of inflation is up to 6-7%, and changes sign afterwards. However, there exists a dominant negative relation suggested by the accumulation of points (countries) to the left side of the averageinfl graph in Figure 2. This fact supports the idea stressed by Cukierman (1992) that inflation leads to the evolution of automatic accommodative mechanisms such as indexation of contracts in the labor and capital markets to the general price level. Society reduces opposition to inflation and public pressure for an independent central bank.

Then, an almost linear and positive behavior can be detected for the two variables regarding synchronization of business cycles (WORLD CYCLE and USA CYCLE). The larger the size of the common component in the business cycle in the countries, the larger the CBI. To understand this result, consider that governments expect their economies to be in the same state of the world (boom or slumps) as foreign economies. Governments in each country have a strategic incentive to commit monetary policy in order to free ride on the stabilization provided abroad and gain credibility at home. Hence, the larger the correlation among shocks, the larger the incentives to commitment, i.e. the larger the CBI (see D'Amato & Martina, 2005).

From the point of view of the inflationary bias approach to monetary policy, the impact of per capita GDP is not clear-cut. The real GDP per capita (logrealgdp) is found to be linearly and positively linked to CBI. On the one hand, a higher level of per capita income level entails a lower degree of (real and financial) market failures in the economy, a more efficient fiscal system, and therefore a lower incentive to create inflation for the central banker. On the other hand, economic agents in high-income countries might be better hedged against inflation; hence, their inflation aversion may be lower (Campillo & Miron, 1997). Opposite effects on the inflationary bias in monetary policy entail opposite effects on the incentives to precommit monetary policy. The real GDP per capita is considered to be an indicator of a general measure of development. In Romer (1993), a larger per capita GDP has a negative impact on inflation, i.e. lower inflationary bias. The reduced inflationary bias lowers the incentive to commit with negative impact on the level of independence of the central bank. Lane (1997) and Campillo and Miron (1997) obtained a positive sign for the log per capita GDP on average inflation. Hence, the present outcome is not consistent with the commitment interpretation of the results in Romer, but it is in line with Campillo and Miron's argument.

Also, the relationship between political instability and the level of dependence is not clear-cut in the commitment literature. The high variability of the political environment may imply a lower ability to achieve commitment of monetary policy through delegation to an independent institution. However, a larger political instability may increase the benefits of commitment. From an empirical point of view, the relation between political instability and CBI is ambiguous and mainly depends on the variable used to proxy instability. For example, Cukierman (1992) predicted and empirically verified a high level of party-political instability induces a larger level of independence, whereas the political instability regime has a negative effect on CBI. A partial list of similar studies, in which different measures of political instability and several indices of CBI are used, includes de Haan and van't Hag (1995), Habibi and Bagheri (1997), and Farvaque (2002). Broadly speaking, the literature on the political economy approach to CBI suggests a politically heterogeneous context (federal systems, strong systems of checks and balance, coalition and multiparty governments) pushes for the adoption of independent central banks (see Fernández-Albertos, 2015).

Among the political variables, the following are found to be relevant: the mean district magnitude (mdmh), the number of government and opposition seats (numgov and numopp), and the number of seats of the largest government party and of the largest opposition party (gov1seats and opp1seats). An increase in district magnitude induces a higher CBI, given the accumulation of country-points in the first bit of the curve. The rationale behind this outcome is that an increase in district magnitude tends to increase the number of parties and party system fragmentation (Rae, 1995); so, the larger the heterogeneity in policy preferences, the larger the CBI.

Similar results are obtained for the government and the opposition parties. The larger the number of government seats, makes the policy making more difficult and induces a greater incentive to delegate to an independent central. A similar argument holds for the number of opposition seats. Moreover, the larger the first government party, the smaller the CBI, while the larger the first opposition party, the greater the CBI. This is even clearer when taking into account the relative weight, in terms of number of seats, for these two parties, focusing on the nonlinear behavior in Figure 2. In fact, up to about 100 seats, the effect on CBI is positive (resp. negative) for gov1seat (resp. opp1seat) while it is negative (resp. positive) for a number larger than 100. In general, the larger the fragmentation of the government or party system, the larger the CBI index.

Conclusion

The random forests (RFs) method was implemented to identify the main determinants of the Central Bank Independence (CBI) index from a large database of institutional, political, and economic variables. To the best of our knowledge, RF has not been previously used for the identification of CBI determinants, although it has been utilized in finance (e.g., Creamer & Freund, 2010; De Luca, Rivieccio, & Zuccolotto, 2010; Booth, Gerding, & McGroarty, 2015; Ward, 2017). RF has been utilized to overcome limitations such as omitted variables, collinearity, overfitting, and linear functional form of the regression.

Considering multicollinearity in regression analysis directly bears on whether the coefficients of the model are uniquely identified. This can be problematic from an inferential view: if two variables are correlated, increases in the first variable may be offset by decreases in the second one (and vise versa), so the combined effect is to negate each other. On the other hand, a regularization is performed in

RF through the number of variables sampled at each split. The larger the number of features to choose from, the better splits one can obtain. That also makes each tree more highly correlated with the others, somewhat moderating the diversifying effect of estimating multiple trees in the first place. Importantly, no part of the RF is harmed by highly collinear variables. If two variables provide the same child node purity, one of them may be picked without diminishing the quality of the result. The original paper of Breiman (2001) discusses those issues in detail; however, further studies focus on those advantages of RF, such as Kimes (2006), Dormann et al. (2013), and Kane, Price, Scotch, and Rabinowitz (2014).

Variable selection is efficiently performed, and new implications are derived with respect to the empirical literature on CBI. The analysis shows that the economic variables account for 80% of the variation in CBI, while the ones reflecting party system's fragmentation explain the remaining 20%. Two-thirds of the explained variation due to the economic group is attributed to external constraints, that is, international business cycle and the participation in the European Monetary Union (EMU). Moreover, half of the selected predictors interact nonlinearly with the CBI index, for example average inflation and the number of party seats. Such an empirical strategy turns out to be particularly important when a clear structural model is not available to the researcher, as in this specific economic problem.

Acknowledgements

The authors are grateful to the Editor of *JMASM* and two anonymous referees for their very useful suggestions and remarks which improved the final version of the paper.

References

Alesina, A., & Grilli, V. (1995). The European central bank: Reshaping monetary politics in Europe. In T. Persson & G. E. Tabellini (Eds.), *Monetary and fiscal policy* (Vol. 1) (pp. 49-85). Cambridge MA: MIT Press.

Alesina, A., & Stella, A. (2011). The politics of monetary policy. In B. M. Friedman & M. Woodford (Eds.), *Handbook of monetary economics* (Vol. 3) (pp. 1001-1054). Amsterdam, Netherlands: North-Holland. doi: 10.1016/B978-0-444-53454-5.00006-2

Booth, A., Gerding, E., & McGroarty, F. (2015). Performance-weighted ensembles of random forests for predicting price impact. *Quantitative Finance, 15*(11), 1823-1835. doi: 10.1080/14697688.2014.983539

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5-32. doi: 10.1023/a:1010933404324

Breiman, L., Friedman. J., Stone, C., & Olshen, R. (1984). *Classification and regression trees*. Boca Raton, FL: Chapman and Hall/CRC.

Brumm, H. J. (2011). Inflation and central bank independence: Two-way causality? *Economics Letters, 111*(3), 220-222. doi: 10.1016/j.econlet.2011.02.005

Campillo, M., & Miron, J. A. (1997). Why does inflation differ across countries? In C. D. Romer & D. H. Romer (Eds.), *Reducing inflation: Motivation and strategy* (pp. 335-362). Chicago, IL: University of Chicago Press.

Creamer, G., & Freund, Y. (2010). Using boosting for financial analysis and performance prediction: Application to S&P 500 companies, Latin American ADRs and banks. *Computational Economics, 36*(2), 133-151. doi: 10.1007/s10614-010-9205-3

Cukierman, A. (1992) *Central bank strategy, credibility, and independence: Theory and evidence*. Cambridge, MA: MIT Press.

Cukierman, A. (2008). Central bank independence and monetary policymaking institutions – Past, present and future. *European Journal of Political Economy, 24*(4), 722-736. doi: 10.1016/j.ejpoleco.2008.07.007

Cukierman, A. (2013). Monetary policy and institutions before, during, and after the global financial crisis. *Journal of Financial Stability, 9*(3), 373-384. doi: 10.1016/j.jfs.2013.02.002

Cutler, D. R., Edwards, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J. (2007). Random forests for classification in ecology. *Ecology, 88*(11), 2783-2792. doi: 10.1890/07-0539.1

D'Amato, M., & Martina, R. (2005). Credibility and commitment of monetary policy in open economies. *European Journal of Political Economy, 21*(4), 872-902. doi: 10.1016/j.ejpoleco.2004.11.006

D'Amato, M., Pistoresi, B., & Salsano, F. (2009). On the determinants of central bank independence in open economies. *International Journal of Finance & Economics, 14*(2), 107-119. doi: 10.1002/ijfe.360

DeHaan, J., & Eijffinger, S. C. (2019). The politics of central bank independence. In R. Congleton, B. Grofman, & S. Voigt (Eds.), *Oxford handbook* *of public choice* (Vol. 2) (pp. 499-519). Oxford, UK: Oxford University Press. doi: 10.1093/oxfordhb/9780190469771.013.23

de Haan, J., & van't Hag, G. J. (1995). Variation in central bank independence across countries: some provisional empirical evidence. *Public Choice, 85*(3-4), 335-351. doi: 10.1007/bf01048203

De Luca, G., Rivieccio, G., & Zuccolotto, P. (2010). Combining random forest and copula functions: A heuristic approach for selecting assets from a financial crisis perspective. *Intelligent Systems in Accounting, Finance and Management, 17*(2), 91-109. doi: 10.1002/isaf.315

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, Gabriel, … Lautenbach, S. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography, 36*(1), 27-46. doi: 10.1111/j.1600-0587.2012.07348.x

Emsia, E., & Coskuner, C. (2016). Economic growth prediction using optimized support vector machines. *Computational Economics, 48*(3), 453-462. doi: 10.1007/s10614-015-9528-1

Farvaque, E. (2002). Political determinants of central bank independence. *Economics Letters, 77*(1), 131-135. doi: 10.1016/s0165-1765(02)00120-9

Fernández-Albertos, J. (2015). The politics of central bank independence. *Annual Review of Political Science, 18*, 217-237. doi: 10.1146/annurev-polisci-071112-221121

Gogas, P., Papadimitriou, T., Matthaiou, M., & Chrysanthidou, E. (2015). Yield curve and recession forecasting in a machine learning framework. *Computational Economics, 45*(4), 635-645. doi: 10.1007/s10614-014-9432-0

Gray, R. J. (1988). A class of *K*-sample tests for comparing the cumulative incidence of a competing risk. *The Annals of Statistics, 16*(3), 1141-1154. doi: 10.1214/aos/1176350951

Grömping, U. (2009). Variable importance assessment in regression: Linear regression versus random forest. *The American Statistician, 63*(4), 308-319. doi: 10.1198/tast.2009.08199

Habibi, N., & Bagheri, F. (1997). *Political institutions and central bank independence: A cross country analysis* (Technical report no. 9703). Ankara, Turkey: Department of Economics, Bilkent University.

Ho, T. K. (1995). Random decision forests. *Proceedings of the Third International Conference on Document Analysis and Recognition, 1*, 278-282. doi: 10.1109/icdar.1995.598994

International Monetary Fund. (2009). *World economic and financial surveys: World economic outlook database*. Retrieved from https://www.imf.org/external/pubs/ft/weo/2009/02/weodata/index.aspx

Ishwaran, H., & Kogalur, U. B. (2007). Random survival forests for R. *R News, 7*(2), 25-31. Retrieved from https://cran.r-project.org/doc/Rnews/Rnews_2007-2.pdf

Ishwaran, H., Kogalur, U. B., Blackstone, E. H., & Lauer, M. S. (2008). Random survival forests. *The Annals of Applied Statistics, 2*(3), 841-860. doi: 10.1214/08-aoas169

Ishwaran, H., Kogalur, U. B., Gorodeski, E. Z., Minn, A. J., & Lauer, M. S. (2010). High-dimensional variable selection for survival data. *Journal of the American Statistical Association, 105*(489), 205-217. doi: 10.1198/jasa.2009.tm08622

Kane, M. J., Price, N., Scotch, M., & Rabinowitz, P. (2014). Comparison of ARIMA and random forest time series models for prediction of avian influenza H5N1 outbreaks. *BMC Bioinformatics, 15*(1), 276. doi: 10.1186/1471-2105-15-276

Keefer, P. (2007, December). DPI2006, Database of political institutions: Changes and variable definitions. Washington, DC: The World Bank Development Research Group. Retrieved from https://siteresources.worldbank.org/INTRES/Resources/469232-1107449512766/dpi2006_vote_share_variable_definitions.pdf

Kimes, R. V. (2006). *Quantifying the effects of correlated covariates on variable importance estimates from random forests* (Unpublished Master's thesis). Virginia Commonwealth University, Richmond, VA. Retrieved from https://scholarscompass.vcu.edu/etd/1433/

Lane, P. R. (1997). Inflation in open economies. *Journal of International Economics, 42*(3-4), 327-347. doi: 10.1016/s0022-1996(96)01442-0

Leblanc, M., & Crowley, J. (1993). Survival trees by goodness of split. *Journal of the American Statistical Association, 88*(422), 457-467. doi: 10.1080/01621459.1993.10476296

Polillo, S., & Guillén, M. F. (2005). Globalization pressures and the state: The worldwide spread of central bank independence. *American Journal of Sociology, 110*(6), 1764-1802. doi: 10.1086/428685

Rae, D. W. (1995). Using district magnitude to regulate political party competition. *Journal of Economic Perspectives, 9*(1), 65-75. doi: 10.1257/jep.9.1.65

Romer, D. (1993). Openness and inflation: Theory and evidence. *The Quarterly Journal of Economics, 108*(4), 869-903. doi: 10.2307/2118453

RStudio. (2015). RStudio [computer software]. Boston, MA: RStudio. Retrieved from http://www.rstudio.com/

Segal, M. R. (1988). Regression trees for censored data. *Biometrics, 44*(1), 35-47. doi: 10.2307/2531894

Varian, H. R. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives, 28*(2), 3-27. doi: 10.1257/jep.28.2.3

Ward, F. (2017). Spotting the danger zone: Forecasting financial crises with classification tree ensembles and many predictors. *Journal of Applied Econometrics, 32*(2), 359-378. doi: 10.1002/jae.2525

World Bank Group. (2008). *World development indicators*. Retrieved from https://databank.worldbank.org/data/databases

Zhang, J., & Maringer, D. (2016). Using a genetic algorithm to improve recurrent reinforcement learning for equity trading. *Computational Economics, 47*(4), 551-567. doi: 10.1007/s10614-015-9490-y

Zhou, L., & Lai, K. K. (2017). AdaBoost models for corporate bankruptcy prediction with missing data. *Computational Economics, 50*(1), 1-26. doi: 10.1007/s10614-016-9581-4

Appendix: Political and institutional variables from the DPI2006 database of World Bank by Keefer (2006)

1) Chief Executive Variables:

system: it indicates the type of the political system. Its values are: 0 if the system is presidential, 1 if the president is assembly elected, 2 if it is parliamentary

yrsoffc: it indicates how many years the chief executive has been in office

finittrm: it's a dummy that indicates if there is finite term in the office (1) or if there is not (0)

yrcurnt: it indicates the years left in the current term

multpl: it indicates when there are formal restraints on an executive's term (NA if not), if he can serve additional terms following the current one (1) or not (0)

allhouse: it indicates if the party of the executive controls (1) or not (0) one of the relevant houses

2) Party variables in the legislature:

herfgov: it is the Herfindahl index of the government, i.e., the sum of the squared seat shares of all parties in the government

govfrac: it is the probability that two deputies picked at random from among the government parties will be of different parties

numgov: it indicates the number of total parliament seats held by government parties

numvote: it indicates the vote share of government parties

gov1seat: it indicates the seats of the first government party

gov1vote: it indicates the vote share of the first government party

gov2seat: it indicates the seats of the second government party

gov2vote: it indicates the vote share of the second government party

gov3seat: it indicates the seats of the third government party

gov3vote: it indicates the vote share of the third government party

govothst: it indicates the seats of the other government parties

herfopp: it is the Herfindahl index of the opposition, calculated in the same manner as the Herfindahl government

oppfrac: it is the probability that two deputies picked at random from among the opposition parties will be of different parties

numopp: it indicates the number of opposition seats

oppvote: it is the vote share of opposition parties

opp1seat: it indicates the seats of the first opposition party

opp1vote: it indicates the vote share of the first opposition party

opp2seat: it indicates the seats of the second opposition party
opp2vote: it indicates the vote share of the second opposition party
opp3seat: it indicates the seats of the third opposition party
opp3vote: it indicates the vote share of the third opposition party
oppothst: it indicates the seats of the other opposition parties
herftot: it is the Herfindahl total index, calculated in the same manner as the Herfindahl of the government and Herfindahl of the opposition
frac: it is the probability that two deputies picked at random from the legislature will be of different parties
oppmajh: it is a dummy, which is 1 if the opposition party has an absolute majority in House
maj: it is the margin of majority, i.e., the fraction of seats held by the government, calculated by dividing the number of government seats by total (government plus opposition plus non-aligned) seats
partyage: it is the average of the ages of the first government party, second government party and first opposition party or the subset of these for which age of party is known
exelec: this variable indicates if there is (1) or not (0) an executive election in the current year
execspec: it is a dummy which is 1 if there is executive party special interest
govspec: it is a dummy which is 1 if there is first government party special interest
coalspec: it is a dummy which is 1 if there are any coalition parties' special interest

3) Electoral rules:

liec: it is an index of electoral competitiveness: it goes from 1 to 7, and an increasing value corresponds to a decreasing vote share of the largest party
mdmh: it represents the weighted average of the number of representatives elected by each constituency size
pluralty: it has a value equal to 1 if the legislative election winner takes the majority of the seats, and it's 0 otherwise
pr: it has a value equal to 1 if there is proportional representation in legislative elections, 0 otherwise
housesys: it deals with electoral rules: is equal to 1 if the majority of seats are assigned with plurality rules, and it's 0 if they are assigned with proportional rules

4) Stability and checks and balance:

tenlong: it measures the tenure of veto player with the longest tenure
tenshort and tenshortlax: they measure the tenure (years) of the veto player with the shortest tenure; their difference depends on the numbers of veto players.

checks: it indicates the number of veto players
stabs: it counts the percent of veto players who drop from the government in any given year

5) Federalism:

auton: it is a dummy which is 1 if there are autonomous regions
federal: it takes value 0 if neither local executive nor local legislature are locally elected, 1 if the executive is appointed, but the legislature elected, 2 if they are both locally elected.