1800

Discussion
Papers

2019

# EU Merger Policy Predictability Using Random Forests

Pauline Affeldt

Opinions expressed in this paper are those of the author(s) and do not necessarily reflect views of the institute.

# EU Merger Policy Predictability Using Random Forests[*]

Pauline Affeldt[†]

April 17, 2019

**Abstract**

I study the predictability of the EC's merger decision procedure before and after the 2004 merger policy reform based on a dataset covering all affected markets of mergers with an official decision documented by DG Comp between 1990 and 2014. Using the highly flexible, non-parametric random forest algorithm to predict DG Comp's assessment of competitive concerns in markets affected by a merger, I find that the predictive performance of the random forests is much better than the performance of simple linear models. In particular, the random forests do much better in predicting the rare event of competitive concerns. Secondly, post-reform, DG Comp seems to base its assessment on a more complex interaction of merger and market characteristics than pre-reform. The highly flexible random forest algorithm is able to detect these potentially complex interactions and, therefore, still allows for high prediction precision.

**JEL Classification:** K21; L40

**Keywords:** Merger policy reform, DG Competition, Prediction, Random Forests

## 1. Introduction

European competition policy, i.e. the design and enforcement of competition rules, is designed to ensure fair and equal conditions for businesses. Competition policy covers the monitoring and, where necessary, blocking of anticompetitive agreements (in particular hardcore cartels), abuses by dominant firms, mergers and acquisitions as well as state aid. Among the different areas of the European Commission's (EC) competition policy, this paper focuses on merger policy.

The European Communities Merger Regulation (ECMR), the legal basis for common European merger policy, came into force in 1990. Over the course of the next 25 years, European merger control has seen significant changes, most prominently with the 2004 amendment to the ECMR.

The goal of this reform was to adopt a "more economic approach" to European merger control, i.e. an approach closer to economic principles. With the reform, a new substantive test, the so-called "significant impediment of effective competition" (SIEC) test, as well as the concept of an efficiency defense were introduced, a chief economist was appointed, significant procedural changes were made, and horizontal merger guidelines were issued.

Merger policy can be approached and assessed from various angles. Duso, Gugler, and Szücs (2013) identify three dimensions along which merger policy effectiveness can be evaluated: the predictability, the correctness, and the deterrence effects of a decision. This paper considers only the first part of merger policy effectiveness: its predictability. While ultimately the correctness of a decision is one of the main aspects of effective merger control, the predictability of decisions based on *ex ante* observable merger characteristics is of interest in its own respect.

Prior to the notification of a merger, legal certainty and the predictability of the merger control procedure are important for judges, competition lawyers, and, in particular, for firms' choices of which kind of mergers to propose. In particular, transparent antitrust policy fosters accountability of the antitrust authority and reduces personal bias in decisions taken, as decisions and the reasoning behind them must be made public. In turn, this will lead to a more fair and consistent process, and, consequently, decisions. The fact that the entire assessment process is transparent, with decisions and reasoning made public after a decision has been taken, also increases confidence in the assessment process, thus enhancing the authority's credibility. Lastly, a transparent and predictable process allows firms to better understand the authority's merger review process and, ultimately, predict the outcome of a merger review to a certain extent. Therefore, it should encourage self-compliance: firms should be encouraged to propose pro-competitive mergers and discouraged from proposing anti-competitive mergers (McAfee, 2010). While McAfee (2010) also discusses the costs of transparency in antitrust policy, the desirability of merger control policy with clear, transparent, and traceable rules and proceedings that decrease uncertainty for firms as well as the risk of political influence, has long been stressed (see for example Smith (1958) and Elman (1965)).[1]

One specific goal of the 2004 merger reform was precisely to increase legal certainty and transparency of the merger review process as evidenced by the publication of merger guidelines and the institutional changes made (See for example Gerber (2014)). However, the effect of the reform on the predictability of Directorate General Competition (DG Comp)'s decisions is ambiguous, as the use of a "more economic approach" in the merger review implies a shift from simple general rules, such as concentration thresholds, toward a more in depth case-by-case economic analysis. The question is hence whether the merger reform increased the *ex ante* predictability of decisions based on market and merger characteristics and also how the merger reform changed the decision criteria on which DG Comp bases its merger assessment.

To evaluate the predictability of DG Comp's merger policy, in this paper, I use a dataset con-

---

[1]In particular, McAfee (2010) argues that transparency encourages the use of simple rules that can lead to repeated and consistent mistakes in case assessment, increases the cost of investigations due to additional disclosure costs, might delay innovation in assessment techniques and increases the importance of precedents.

taining almost all mergers notified to the European Commission between 1990 and 2014. This amounts to 25 years of data on European merger control. Unlike most of the existing literature, rather than assessing mergers at the aggregate level, the data is collected at a more fine-grained level, defining an observation as a particular product and geographic market combination concerned by a merger.[2] Importantly, this allows for studying the factors that cause competitive concerns in specific sub-markets, as mergers typically concern several product and geographic markets that can be affected differently by the concentration. The final dataset used in the empirical analysis contains 22,812 product/geographic market level observations belonging to 2,417 DG Comp merger decisions.

The goal of this paper is, firstly, to study the predictability of DG Comp's merger policy and, secondly, to assess how it changed following the 2004 merger reform. Unlike the existing literature studying the determinants of DG Comp's merger intervention decisions and their predictability, I use the non-parametric random forests algorithm by Breiman (2001) to predict DG Comp's assessment of competitive concerns arising in affected markets due to the merger. This machine learning algorithm is designed to maximize predictive performance rather than estimating causal effects and allows for highly flexible, non-linear interactions between covariates. First, I train two random forests, one pre-reform and one post-reform, and compare their predictive performance to the predictive performance of a Linear Probability Model (LPM). Second, I study how the predictions of the pre-reform and post-reform random forests differ and how the merger assessment of DG Comp changed following the 2004 merger reform.

I find that the predictive performance of the random forests is much better than the performance of the LPM models. While all models are able to predict the majority outcome of *no competitive concerns* very well (between 80% and 90% correct predictions), the LPM models do very poorly in predicting the minority outcome of *competitive concerns*, with only between 16% and 44% correct predictions. In particular, the pre-reform LPM model only correctly predicts about 30% of the competitive concern markets in the pre-reform prediction set, the post-reform LPM model only predicts about 31% of the markets with concerns in the post-reform prediction set correctly. Furthermore, based on these predictions as well as the $R^2$, the LPM models would wrongly suggest that the predictability of DG Comp's assessment decreased after the 2004 reform. The random forests however, are able to correctly classify the minority class cases in about 60% of the cases, both pre- and post-reform. Thus, it is not true that the predictability of DG Comp's merger policy decreases post-reform.

Secondly, the random forest trained on pre-reform data only predicts well for the pre-reform period but does significantly worse than the random forest trained on the post-reform data in predicting the minority outcome of *competitive concerns* for the post-reform period. This suggests, in line with the results of Affeldt, Duso, and Szücs (2019), that DG Comp's assessment criteria changed post-reform. Based on the two random forests and correcting for the case mix effect, the policy effect shows a decrease in the concern rate of about 1.5 percentage points post-reform.

---

[2]One exception is Mini (2018), who also uses affected markets as the level of observation.

3

However, this decomposition only considers the average concern rate rather than investigating for which observations the predictions of the pre-reform and the post-reform random forests differ. Studying post-reform cases for which the predictions of the two random forests differ, I find that pre-reform DG Comp seems to have relied more on structural indicators in its merger assessment, while post-reform DG Comp seems to base its assessment of competitive concerns more on a case-by-case analysis and less on simple structural indicators such as market shares or concentration measures. Nevertheless, the highly flexible random forest algorithm still allows for high prediction precision. These results are in line with the findings from Affeldt, Duso, and Szücs (2019), who find the same time-dynamics of DG Comp's decision procedures using causal forests without imposing a different pre- and post-reform model.

The paper is structured as follows. In Section 2, I briefly discuss the institutional background of European merger control and the 2004 reform of the EU merger guidelines. In Section 3, I review studies that empirically investigate the determinants and predictability of merger intervention and present some recent papers that employ machine learning techniques for prediction tasks. I describe the dataset used in the empirical analysis in Section 4 and the random forest algorithm employed for prediction in Section 5. I present the prediction results and discuss the change in the decision rules after the merger reform in Section 6 and conclude in Section 7.

## 2.  Institutional Background

Mergers and merger control are important for firms as well as society and, ultimately, consumers. While mergers can reduce competition and lead to increases in market power, and, consequently, price increases to the detriment of consumers, they can also lead to important efficiency gains due to economies of scale and scope that are partly passed on to consumers, increasing not only producer surplus but also consumer surplus.

In the European Union (EU), regulation of competition has been undertaken by the European Commission since 1975. Specifically, DG Comp is responsible for enforcing EU rules regarding antitrust, mergers, state aid, and liberalization with the goal of protecting consumer surplus.

The basis for DG Comp's merger policy is the European Communities Merger Regulation (ECMR), which was passed in 1989 and came into force in September 1990.[3] It specifies the scope of intervention and legal competence of the European Commission in merger cases with a "community dimension". The definition of this "community dimension" was broadened by the passing of regulation 1310/97[4] in 1997. In particular, mergers must be notified to the EC if the combined world-wide turnover of the merging parties is sufficiently high, if their combined intra-community turnover is sufficiently high and not too concentrated in one Member State only.[5] This implies that, from 1990 onwards, all major combinations had to be notified and have

---

[3]Council Regulation (EEC) No 4064/89 of 21 December 1989 on the control of concentrations between undertakings [Official Journal L 395 of 30 December 1989].

[4]Council Regulation (EC) No 1310/97 of 30 June 1997 [Official Journal L 180 of 9 July 1997].

[5]In particular, in article 1.2 of regulation 4064/89, a combination is defined to have community dimension by meeting the following conditions: a) the aggregate worldwide turnover of all the undertakings concerned is

4

been scrutinized by DG Comp. Notice that these definitions also include companies that are located, produce, and sell outside of the European Union, as long as their sales to European markets are sufficiently high. Thus, a merger can be subject to the jurisdiction of more than one competition authority. For example, the merger of the two US companies *General Electric* and *Honeywell* was ratified by American authorities, but prohibited by the European Commission in 2001.[6]

Once it is established that a combination is subject to EC jurisdiction, the merging parties must notify the concentration to the EC prior to its implementation. After the reception of the official notification, the EC publishes a note in the Official Journal of the European Communities and third parties can comment on the proposed merger.

After notification to the EC (and the receipt of all necessary information), so-called phase-1 proceedings are initiated. DG Comp then has 25 working days (which can be extended to a maximum of 35 working days) for an initial assessment of the merger. DG Comp can then clear the proposed merger (phase-1 clearance), clear it subject to remedies proposed by the merging parties (phase-1 remedy), or initiate a more in-depth investigation (phase-2 investigation) depending on whether the proposed transaction raises competitive concerns and whether these can be addressed by initial remedies or not. Furthermore, the merging parties can also withdraw the proposed merger during phase-1 (phase-1 withdrawal).

If DG Comp initiates an in-depth phase-2 investigation, it may take up to 90 working days. Based on the conclusions from this in-depth investigation of the effects of the merger, DG Comp can again unconditionally clear the merger (phase-2 clearance), clear the merger subject to commitments by the merging parties (phase-2 remedy), or prohibit the merger (phase-2 prohibition). Again, the merging parties can also withdraw the proposed merger in phase-2 (phase-2 withdrawal). Bergman, Jakobsson, and Razo (2005) argue that withdrawing a merger in phase-2 of the investigation process is virtually equivalent to a prohibition as parties often withdraw a merger before an actual prohibition takes place. Hence, both a prohibition as well as a phase-2 withdrawal suggest that the EC and the notifying parties were unable to find suitable remedies to address the anti-competitive concerns of the proposed transaction.

The ECMR was amended in 2004. This amendment brought significant changes to European merger control with the aim of bringing merger control closer to economic principles: the concept of an efficiency defense was introduced, a chief economist was appointed, the timetable for

---

more than ECU 5 000 million, and b) the aggregate Community-wide turnover of each of at least two of the undertakings concerned is more than ECU 250 million, unless each of the undertakings concerned achieves more than two-thirds of its aggregate Community-wide turnover within one and the same Member State. Regulation 1310/97 assesses a community dimension even if a merger does not meet the original two conditions, provided it satisfies the following four conditions: (a) the combined aggregate worldwide turnover of all the undertakings concerned is more than EUR 2 500 million; (b) in each of at least three Member States, the combined aggregate turnover of all the undertakings concerned is more than EUR 100 million; (c) in each of at least three Member States included for the purpose of point (b), the aggregate turnover of each of at least two of the undertakings concerned is more than EUR 25 million; and (d) the aggregate Community-wide turnover of each of at least two of the undertakings concerned is more than EUR 100 million, unless each of the undertakings concerned achieves more than two-thirds of its aggregate Community-wide turnover within one and the same Member State.

[6]See for example Patterson and Shapiro (2001) for a discussion of the case.

remedies was improved, and horizontal merger guidelines were issued. The reception of the new merger regulation was generally favorable (Lyons, 2004).

One of the most significant changes in the horizontal merger guidelines was the move from the old "Dominance Test" (DT) for market power to a "significant impediment of effective competition test" (SIEC). In the old DT test, a merger was declared as incompatible with the common market if it *"creates or strengthens a dominant position as a result of which effective competition would be significantly impeded"*. This implies that pre-reform, the creation or strengthening of a dominant position was a necessary condition for the prohibition of a merger. Thus, mergers that reduced effective competition without creating a dominant position could not be challenged under the old legislation. It is argued that the dominance test was deficient in cases of collective dominance and tacit collusion and that the "substantial lessening of competition" test employed by the United States' Federal Trade Commission (FTC) would be preferable.

In the revised 2004 Merger Regulation the wording of article 8.3 (prohibition) hence reads:

> *"A concentration which would significantly impede effective competition, in the common market or in a substantial part of it, in particular as a result of the creation or strengthening of a dominant position, shall be declared incompatible with the common market."*[7]

This revision implies that the creation of a dominant position is no longer a necessary condition for intervention and, therefore, aligns the test used by DG Comp more closely with US practice (Bergman, Coate, Jakobsson, and Ulrick, 2010; Szücs, 2012).

## 3. Previous Literature

Mergers are an important research topic in the field of industrial organization. There are large bodies of theoretical and empirical literatures on questions such as firms' incentives to merge and merger policy effectiveness. Duso, Gugler, and Szücs (2013) identify three dimensions along which merger policy effectiveness can be evaluated: the predictability, correctness, and deterrence effects of a decision. A large part of the literature studying the effectiveness of merger control looks at whether the competition authority made the correct decision in a particular case (*ex-post* evaluations of merger policy) (Duso, Neven, and Röller, 2007; Duso, Gugler, and Yurtoglu, 2011; Kwoka, 2013).[8] A correct decision in this context is a decision that achieves the goals set in the legal framework - in the European Union as well as in most other jurisdictions the goal of competition policy is the protection of consumer surplus. A merger that decreases consumer surplus is considered to be anti-competitive. Thus, in order to judge whether a particular decision was correct, one must determine whether a given merger harmed consumer surplus. For example, Duso, Neven, and Röller (2007) use the reaction of competitors' stock market prices to

---

[7]"Dominance" has been defined as *"...a position of economic strength enjoyed by an undertaking which enables it to prevent effective competition being maintained on the relevant market by giving it the power to behave to an appreciable extent independent of its competitors, customers and ultimately of consumers"* by the European Court of Justice in *United Brands* (27/67, E.C.R. 207, para. 65).

[8]Duso (2012) provides a literature review of *ex-post* merger evaluation studies.

evaluate the degree of pro- or anti-competitiveness for a sample of mergers.[9] They then employ a probit model to estimate the frequency of type I (prohibition of a pro-competitive merger) and type II (clearance of an anti-competitive merger) errors in the decisions.

Instead, the present paper focuses only on the first part of merger policy effectiveness: its predictability. The goal of the paper is to understand how DG Comp decides on interventions in merger cases and whether it is possible to predict DG Comp's decision based on *ex ante* merger and market characteristics. However, these predictions do not allow for judging whether DG Comp's decision was correct in the sense that it protected consumer surplus.

Therefore, the present paper is specifically related to two strands of the literature. Firstly, by predicting the intervention decision by DG Comp based on *ex ante* merger characteristics, this paper relates to the literature on the determinants of merger policy intervention by competition authorities and the predictability of decisions, in particular to those papers that empirically review the EU merger policy reform. Secondly, in this paper, I employ machine learning techniques to predict DG Comp's decisions. Hence, on the methodological side, this paper relates to the literature on prediction using machine learning. In the following, I give a brief overview of these two strands of literature.

## 3.1. Policy Predictability

First, to the best of my knowledge, all of the related literature, except the two studies by Bradford, Jackson, and Zytnick (2018) and Mini (2018), investigate the determinants of merger intervention decisions *at the merger level* and for a *sample of merger cases* only. As discussed in detail in Affeldt, Duso, and Szücs (2019), the scope and depth of the present data allow for going beyond the existing literature by allowing for heterogeneity within merger cases by examining the individual product and geographic markets concerned.

Secondly, the existing literature focuses on studying and discussing the relevant determinants of merger intervention decisions, the difference in policy between the EU and the US, or between the EU pre- and post-merger reform. None of the existing studies focus on the predictive performance of the employed empirical models.

Thirdly, all of the existing literature uses parametric models to empirically study the determinants of merger intervention decisions. Instead, I employ flexible, non-parametric machine learning techniques designed to maximize predictive performance to predict DG Comp's decisions. In particular, as I argue below, these machine learning techniques allow for predicting outcomes well, even for the rare outcome. However, as the focus of the existing literature is not on predictive performance, most papers report the percentage of correct predictions but do not distinguish between correct predictions of the majority and minority class. However, at the extreme, if e.g. 90% of the observations are class 1 and the model classifies 100% of the observations as class 1, then 90% of the cases are correctly predicted but the model is essentially useless

---

[9]If the merger is anti-competitive, it will likely lead to rising share prices of competitors, as in an oligopolistic setting an increase in concentration leads to an increase in mark-ups and profits also of competitors.

7

as it is unable to correctly predict any of the rare class 2 cases.

Bergman and co-authors study European merger control in a series of papers. In the first study, Bergman, Jakobsson, and Razo (2005) employ a logit model for a sample of 96 EU merger cases to estimate the likelihood of going to phase-2 or prohibition decisions as a function of market-relevant and political variables. They report an overall rate of correctly classified cases of going to phase-2 of 79% - 91% and of the decision to prohibit a merger of 84% - 86%. The authors also discuss that while this overall rate of correctly classified cases is relatively high, only between 20% and 27% of the prohibitions are correctly classified while all non-prohibitions are correctly classified. When grouping prohibitions and phase-2 approvals with commitments together, the overall level of predictability lies between 69% and 93% depending on the specification. For one of the models, the authors report that they are now able to classify 100% of the approvals and 75% of the prohibitions/commitments correctly. However, note that the regressions are based only on a sample of 96 merger cases, out of which 17 are prohibitions, 30 are other phase-2 cases, and 49 are other cases. In terms of unbalancedness of the data, only the specification not regrouping prohibitions and approvals with commitments is comparable to the rate of competitive concerns in the dataset that I use. In this specification, the percentage of correctly predicted prohibitions is below 30%, which is much lower than the percentage of correct predictions for the rare class I achieve with my model (see Section 6).

Bergman, Coate, Jakobsson, and Ulrick (2010) examine instead similarities between EU and US merger decisions using a sample of horizontal phase-2 mergers between 1990-2004 for both the EU (109 cases) and the US (166 cases). They estimate a probit model for each regime to evaluate enforcement policy, where the dependent variable is an indicator for intervention (one for prohibition, approval subject to substantial remedies, or withdrawal by the parties at least one month into the phase-2 investigation). The authors report 83% - 84% and 87% - 91% correct predictions for the EU model and the US model, respectively, depending on the specification. While they state that in both regimes "challenges are predicted more accurately than are closed investigations" (Bergman, Coate, Jakobsson, and Ulrick, 2010, p.321), they do not report the percentage of correct predictions for the different outcomes.

In the most recent study of the series, Bergman, Coate, Mai, and Ulrick (2016) update the dataset of Bergman, Coate, Jakobsson, and Ulrick (2010) by adding observations both to the EU as well as the US dataset for the time period following the 2004 EU merger policy reform. The final dataset used in the analysis contains a sample of 151 EU phase-2 cases and 260 US cases, covering the 1993-2013 period. Separate logit models on an intervention indicator variable are estimated EU cases (distinguishing pre- and post-reform) and US cases. While the authors report about 82% correct predictions for the EU models and above 90% correct predictions for the US models, they do not discuss these results in the paper nor do they report the percentage of correct predictions for the intervention and no-intervention cases separately.

Szücs (2012) investigates the convergence between US and EU merger policy following the 2004 EU merger policy reform using a sample of 309 EU and 286 US merger cases decided by DG Comp and the FTC, respectively, between 1991 and 2008. For each of the pre-reform EU, post-

reform EU and US merger samples, he estimates a logit model on the decision to intervene and then uses the estimated models to predict the probability of intervention for each merger case from the point of view of both competition authorities (similar to Bergman, Coate, Jakobsson, and Ulrick (2010)). While the US model classifies 86% of the cases correctly, the percentage of correctly classified cases is above 90% for the EU model both pre- and post-reform. The author does not report the percentage of correct predictions for the intervention and no-intervention outcomes separately.

Duso, Gugler, and Szücs (2013) evaluate European merger policy effectiveness along three dimensions: the predictability, correctness, and deterrence effects of a decision. Regarding predictability of European merger policy, Duso, Gugler, and Szücs (2013) estimate two probit models (one pre-reform, one post-reform) for a sample of 368 EU merger cases where the intervention decision of DG Comp (remedies or prohibition) is a function of *ex ante* observable merger characteristics. Model fit is discussed in terms of pseudo $R^2$ and the percentage of correctly classified observations (71% pre-reform and 76% post-reform). Once again, the percentage of correctly classified cases is not reported for the intervention and no-intervention outcomes separately.

Mai (2016) studies the effect of the EU merger policy reform on the probability of a merger being challenged by DG Comp based on a sample of 341 phase-1 and phase-2 horizontal mergers between 1990 and 2012. The probability of a challenge in a probit model pooling pre- and post-reform cases is driven by the market shares of the merging parties, entry barriers, and some other factors. Mai (2016) also estimates separate pre- and post-reform models and applies the methodology used by Bergman, Coate, Jakobsson, and Ulrick (2010), Szücs (2012) and Bergman, Coate, Mai, and Ulrick (2016) by predicting the probability of a challenge for pre-reform mergers using the post-reform model and vice versa. The author reports an overall rate of correct predictions above 80% and up to 90% depending on the specification, where the post-reform models generally perform better than the pre-reform models. As in the papers discussed previously, correct predictions for the different classes are not reported.

Bradford, Jackson, and Zytnick (2018) empirically investigate whether European merger control is used for protectionism and find no evidence that DG Comp intervened more frequently or extensively in transactions involving non-EU or US-based firms. Differently from the previous literature and similar to the present dataset, they collect information on all merger cases decided by DG Comp between 1990 and 2014. However, their analysis is still conducted at the level of the merger rather than the concerned product and geographic market. Furthermore, they do not collect information on the structural parameters of market shares, concentration, likelihood of entry, and foreclosure from the case documents. For the linear probability models estimating the probability of challenge, the authors only report $R^2$ as a measure of model fit and do not discuss the predictive performance of the model at all.

The paper with the most closely related dataset to the one used here is the one by Mini (2018). Like this paper, and unlike all other existing studies, Mini (2018) also collects information on the universe of EU merger decisions from the publicly available case documents between 1990 and 2013, recording each market concerned by the transaction as a separate observation.

Thus, for each merger, he records potentially many observations and collects similar merger and market level characteristics from the case documents, like those included in the dataset used here. He then estimates probit models at this concerned market level for horizontal overlap markets, interacting all explanatory variables with a post-reform indicator variable. In the first model, the main variables of interest are the merging parties' market shares and the change in market shares, in the second model, the main variables of interest are post-merger Herfindahl-Hirschman-Index (HHI)[10] as well as the change in HHI due to the merger. For these models, the author reports about 92% correctly predicted observations; however also here predictions are not distinguished by class.

Thus, while most of the existing literature reports the percentage of correctly predicted observations together with the $R^2$ as an indicator of model fit, Bergman, Jakobsson, and Razo (2005) is the only paper that discusses the lower predictive performance of the model for the rare class. Most papers report percentages of correctly predicted observations above 80% and up to 90%. However, once again, if the data contains for example 90% class 1 observations and only 10% class 2 observations and the model classifies 100% of the observations as class 1, then 90% of the cases are correctly predicted but the model is not able to correctly predict any of the rare class 2 cases.

## 3.2. Prediction using Machine Learning

Unlike the existing literature on the effects of the EU merger policy reform, in this paper I employ non-parametric machine learning techniques to predict the intervention decision by DG Comp and to evaluate how the decisions changed post EU merger reform. On the methodological side, this paper therefore relates to the economics literature employing machine learning techniques for prediction tasks. While this list is by no means exhaustive, I mention a few papers here. The topics these papers study are very different from my application, but all studies try to predict an outcome based on observables using machine learning techniques. Not all of them use the random forest algorithm, but they all focus on prediction rather than causal inference employing machine learning techniques that allow for more complex interactions between covariates than do parametric models.

According to Kleinberg, Ludwig, Mullainathan, and Obermeyer (2015) an increasing number of empirical studies consider prediction policy problems. In particular, the authors mention prediction problems in education (e.g. value added of teachers), labor market policy (e.g. unemployment spell length), social policy (e.g. predicting highest risk youth for targeting interventions), and finance (e.g. identifying credit-worthiness of borrowers). Kleinberg, Ludwig, Mullainathan, and Obermeyer (2015) further include an illustrative healthcare application in which they use a sample of Medicare beneficiaries to predict the pay-off of hip or knee replacement surgery. The question the authors ask is whether one can predict which surgeries are futile based on patient characteristics. In order to do so they predict mortality in the 1-12 months after

---

[10]The HHI is defined as the sum of squared market shares of all firms active in the market.

hip or knee replacement using Lasso based on patient demographics, co-morbidities, symptoms, injuries, acute conditions, and health care utilization. The predictive performance of the model is however not discussed.

Chalfin, Danieli, Hillis, Jelveh, Luca, Ludwig, and Mullainathan (2016) study the selection of the most productive labor input, which is also a prediction policy problem. In particular, they illustrate the use of machine learning techniques in this respect with two applications: they predict worker productivity to improve police hiring practices (lowering police use of force or misbehavior) and teacher tenure decisions (improving teacher value added) using stochastic gradient boosting in the first and regression with lasso penalty in the second application. Explanatory variables include socio-demographic attributes of workers (i.e. police officers or teachers), students, schools and surveys capturing e.g. prior behavior. However, the authors do not discuss the predictive performance of the models in detail.

Björkegren and Grissen (2018) use machine learning techniques to predict loan repayment for post-paid mobile subscriptions in a developing country context. In particular, they use random forests and logistic regression with a model selection procedure to predict the probability of loan default using behavioural patterns derived from raw mobile operator transaction records. Model performance is mainly evaluated based on the area under the receiver operating characteristic curve (AUC). For the random forests, the AUC lies between 0.62 and 0.71 depending on the model, which is much lower than the AUCs I achieve with the random forests (see Section 5).

Kleinberg, Lakkaraju, Leskovec, Ludwig, and Mullainathan (2018) and Ribers and Ullrich (2018) go one step further by not only predicting outcomes but also studying whether the machine learning algorithm makes better decisions than humans. Kleinberg, Lakkaraju, Leskovec, Ludwig, and Mullainathan (2018) analyze the problem of predicting risk of defendants' committing a crime in the context of judges' bail decisions using gradient boosted decision trees and judging whether machine prediction can improve judges' bail decisions. The authors highlight the importance of unobservable characteristics, selected labels, and omitted pay-offs to properly compare machine prediction and human decisions. In particular, the selective labels problem arises because crime outcomes can be observed only for those defendants who were released by the judges. Hence, crime rates of those defendants who were kept in bail have to be predicted. However, this is problematic as judges might have based their decisions on unobservables, so the crime rates of released defendants might not be a good proxy for crime rates of detained defendants with similar observable characteristics. To deal with this problem the authors use different econometric strategies including quasi-random assignment of cases to judges.

Ribers and Ullrich (2018) train a random forest on a high-dimensional administrative dataset from Denmark to predict bacterial causes of urinary tract infections. Unlike many machine learning papers tackling prediction centred policy problems, the authors have the advantage of observing labels, i.e. they have patient test outcomes independent of physician prescription choices. This implies that they can actually evaluate whether physicians or the machine made the right prediction. They show that machine predicted bacterial risk is highly correlated with the actual presence of bacterial urinary tract infection. They model the prescription decision as a

11

trade-off between the social cost of prescribing (as it increases resistance) and the curative effect on the patient in case she truly has a bacterial infection. They find that antibiotic prescribing can be lowered by up to 10% with no reduction in the number of treated patients suffering from a bacterial infection based on a machine learning assisted prescription rule that allows to redistribute prescriptions from low risk to high risk patients.

I cannot make this additional step of evaluating whether DG Comp or the machine learning algorithm took the right merger intervention decision. In order to judge the correctness of a decision, I would need a measure of whether a proposed merger is anti- or pro-competitive, i.e. whether it decreases or increases consumer surplus. For example Duso, Neven, and Röller (2007) and Duso, Gugler, and Szücs (2013) use the reaction of competitors' stock market prices for a sample of mergers to determine whether a proposed merger is pro- or anti-competitive in order to evaluate whether DG Comp took the right decision.

## 4. Data

The initial merger database contains almost the entire population of DG Comp's merger decisions, both in the dimension of time and with regard to the scope of the decisions encompassed. The data were obtained from the publicly accessible cases published by DG Comp on the EC's webpage.[11] We start data collection with the very first year of common European merger control, 1990, and include all years up to 2014. This amounts to data on the first 25 years of European merger control. Rather than taking a particular merger case as the level of observation, we collected data at a more fine-grained level and defined an observation as a particular product and geographic market combination affected by a merger. For further details on the merger database as well as the data collection procedure, see the data documentation Affeldt, Duso, and Szücs (2018).

For the analysis in this study, I do not use all observations contained in the merger database. First, I drop cases that were referred back to member states as well as phase-1 withdrawals.[12] This leads to a dataset containing 5,109 DG Comp merger decisions, where each decision occupies a number of rows equal to the number of product/geographic markets affected in the specific transaction. This dataset contains a total of 30,995 market level observations and is used in the analysis in Affeldt, Duso, and Szücs (2019). Appendix A.1 contains summary statistics for this comprehensive dataset.

Secondly, I further reduce the dataset, as I only keep observations of product/geographic markets for which information on the merging parties' joint market share is available and the calculation of the post-merger HHI is possible. The reason is that both the merging parties' joint market share as well as HHI are important criteria taken into account by DG Comp when

---

[11]The types of notified mergers, decisions taken and reports for each of the EC's decisions can be downloaded from: `http://ec.europa.eu/competition/mergers/cases/` and `http://ec.europa.eu/competition/mergers/legislation/simplified_procedure.html`.

[12]There is only two phase-1 withdrawals contained in the dataset.

assessing a merger proposal and, thus, I want to include them in the set of explanatory variables. This leads to a final dataset used in the empirical analysis containing 22,812 product/geographic market level observations belonging to 2,417 DG Comp merger decisions. All summary statistics and analyses presented in the following are based on this dataset.

As market share information is most frequently missing for phase-1 clearance cases, keeping only observations for which market share information is available leads to a selection issue. If mergers for which market share information is missing are systematically different from mergers for which market share information is available, then the estimated correlation between market share and concentration and any intervention decision also captures this unobserved difference and results would be biased. However, the selection issue is studied in Affeldt, Duso, and Szücs (2019), where separate OLS regressions are estimated based on the entire sample, based on the sample without available market share information, based on the sample with available market share information, and based on the sample with available market share information including the merging parties' joint market share as well as post-merger HHI as additional explanatory variables. The estimation results show that DG Comp's decision determinants are rather similar across all sub samples.[13] Furthermore, in this paper, I am not seeking to estimate a causal relationship between for example market shares and an intervention decision. I am just interested in a correct prediction of the intervention decision, which does not require unbiased or causal coefficients on the explanatory variables (see Section 5).

The dataset contains, first, information on the name of the acquirer and the target firm as well as the countries of the merging parties, the dates of the notification to the EC and the final decision[14] as well as the type of decision eventually taken by DG Comp (clearance, remedy, and prohibition) or whether the proposing parties withdrew the notification. The data also allows for distinguishing between a phase-1 and a phase-2 decision.

Table 1 reports the number of phase-1 clearances, phase-1 remedies, phase-2 clearances, phase-2 remedies, as well as prohibitions distinguishing the pre-reform and post-reform periods. Compared to the full merger database, keeping only observations that contain market share information leads to a slight over-representation of phase-2 cases.[15] Note also that, while the full merger database contains 19 prohibitions and 5 phase-2 withdrawals, for one of the prohibitions and all of the phase-2 withdrawals market share information is not available. Thus, my estimation dataset does not contain any phase-2 withdrawals.

The dataset contains a number of *ex ante* characteristics of the merger, some of them at the merger level, some of them at the level of the product/geographic markets affected. While Table 2 contains summary statistics of the variables contained in the dataset that vary at the market

---

[13]Mini (2018) claims in his paper that sample selection is not an issue because he uses the universe of horizontal merger cases in his estimation. However, I do not agree with this statement, as he also uses the merging parties' market share as explanatory variable in his estimation and therefore only records cases "provided that the EC disclosed data on merging parties' shares" (Mini, 2018, p.5).

[14]Note that the notification of a merger and the decision do not necessarily take place in the same year.

[15]See Affeldt, Duso, and Szücs (2018) for comparison: In the full merger database about 90% of the cases are phase-1 clearances.

Table 1: Type of Decisions, 1990-2014

| Type of decision | Pre-reform | | Post-reform | |
|---|---|---|---|---|
| | frequency | percent | frequency | percent |
| Phase-1 clearance | 1,108 | 84.90 | 916 | 82.37 |
| Phase-1 remedy | 99 | 7.59 | 128 | 11.51 |
| Phase-2 clearance | 21 | 1.61 | 26 | 2.34 |
| Phase-2 remedy | 63 | 4.83 | 38 | 3.42 |
| Prohibition | 14 | 1.07 | 4 | 0.36 |
| Total | 1,305 | 100.00 | 1,112 | 100.00 |

level, Table 3 contains summary statistics for the merger level variables. As I predict DG Comp's assessment separately for the pre- and post-reform periods in Section 6, I also report summary statistics distinguishing pre- and post-reform cases.

The first variable in Table 3 (*Intervention*) is a dummy variable that indicates whether DG Comp intervened in a particular merger case. I define this variable to be equal to one if DG Comp prohibited the merger, cleared the merger subject to remedies in phase-1, or cleared the merger subject to remedies in phase-2.[16] The corresponding variable at the product/geographic market level is the first variable in Table 2, (*Concern*), which is an indicator variable equal to one if the merger raised competitive concerns in a specific product/geographic market according to DG Comp. This is the case in about 14% of the markets pre-reform and 13% of the markets post-reform. As this variable indicates in which particular markets the merger is likely to be problematic, this is the dependent variable of the empirical analysis presented in Section 6. Thus, instead of estimating the overall probability of an intervention, I estimate the probability that competitive concerns are found in a market affected by the merger. The higher the fraction of concerned markets in which competitive concerns are found, the higher the likelihood that DG Comp will intervene in a merger case.[17]

The other variables contained in both tables are used as covariates in the empirical analysis and describe the merger as well as how the merger affects the concerned markets according to DG Comp's *ex ante* assessment. However, note that of course all of these variables are based on what the official decision documents state, so to some extent they might reflect the assessment or subjective view (or mistakes) of DG Comp. This issue is present in most papers in this literature. One exception is Duso, Gugler, and Szücs (2013), who include only truly *ex ante* observable merger characteristics (such as for example the nationality of the merging parties or whether the concentration is a full merger) as explanatory variables in the probit estimation for

---

[16]In principal, I would also treat a phase-2 withdrawal as an intervention by DG Comp as in Affeldt, Duso, and Szücs (2019). Given that the phase-2 withdrawals fall out of my estimation dataset due to lack of market share information, the treatment of withdrawals by the merging parties in phase-2 is not an issue here.

[17]See also Affeldt, Duso, and Szücs (2019). In the regressions explaining an intervention decisions at the merger level, the fraction of affected markets in which the merger leads to competitive concerns according to DG Comp positively affects the probability of intervention.

14

Table 2: Summary Statistics Variables at Market Level

| | Pre-reform | | | Post-reform | | |
|---|---|---|---|---|---|---|
| | mean | sd | obs | mean | sd | obs |
| Concern | 0.14 | 0.34 | 8,531 | 0.13 | 0.33 | 14,281 |
| Vertical merger | 0.18 | 0.39 | 8,531 | 0.35 | 0.48 | 14,281 |
| Conglomerate merger | 0.05 | 0.21 | 8,531 | 0.00 | 0.07 | 14,281 |
| National market | 0.66 | 0.47 | 8,531 | 0.64 | 0.48 | 14,281 |
| EU wide market | 0.24 | 0.43 | 8,531 | 0.22 | 0.41 | 14,281 |
| Worldwide market | 0.07 | 0.25 | 8,531 | 0.13 | 0.33 | 14,281 |
| Left open market | 0.03 | 0.17 | 8,531 | 0.01 | 0.10 | 14,281 |
| Entry barriers | 0.12 | 0.33 | 8,531 | 0.16 | 0.37 | 14,281 |
| Risk of foreclosure | 0.05 | 0.21 | 8,531 | 0.02 | 0.14 | 14,281 |
| Number of competitors | 1.85 | 2.23 | 8,531 | 2.04 | 2.58 | 14,281 |
| No competitor information | 0.45 | 0.50 | 8,531 | 0.46 | 0.50 | 14,281 |
| Joint market share | 30.49 | 22.31 | 8,531 | 33.63 | 24.27 | 14,281 |
| Post-merger HHI (low) | 1,890.51 | 2,112.91 | 8,531 | 2,301.39 | 2,495.94 | 14,281 |
| Post-merger HHI (high) | 5,658.73 | 2,252.89 | 8,531 | 5,627.19 | 2,249.99 | 14,281 |

intervention. In the working paper version (Duso, Gugler, and Szücs, 2012), a second model, the so-called "investigation model", is estimated; it contains results from the merger investigation as additional explanatory variables. The predictive power of this second model, measured by pseudo $R^2$ as well as the percentage of correctly classified observations, increases significantly compared to the first.[18] Therefore, the predictability of DG Comp's intervention decisions would likely decrease if I based my estimations only on unambiguously objective merger and market characteristics contained in the decision documents (such as, for example, whether the merger is a full merger or not). However, there is a trade-off between basing the estimation only on *ex ante* observable merger characteristics and omitting important determinants of whether a merger raises competitive concerns, such as market shares or entry barriers. Furthermore, to the extent that the legal system, in which DG Comp is operating, provides a consistency check (e.g. on how market shares should be calculated or in which types of markets entry barriers are high), I believe that DG Comp's merger assessment should be consistent.

The nature of the merger is described by a number of indicator variables. The dataset contains an indicator variable distinguishing full mergers from acquisition of shares as well as an indicator variable for joint ventures; see Table 3. At the market level, the dataset contains information on whether a product/geographic market is vertically affected by the merger. Vertically affected markets are markets where one or more of the merging parties operate in a market that is upstream or downstream of a market in which another merging party is active and any of

---

[18]In particular, the pseudo $R^2$ increases from 0.19 (pre-reform) and 0.25 (post-reform) to 0.68 and 0.59 pre- and post-reform, respectively. Also the percentage of correctly classified observations increased from 71% (pre-reform) and 76% (post-reform) to 90% for both the pre- and post-reform models.

their individual or combined market shares at either level is 25% or more.[19] The dataset further includes an indicator variable that is one if the merger is conglomerate in nature in the particular concerned market[20], see Table 2.

Table 3: Summary Statistics Variables at Merger Level

|  | Pre-reform | | | Post-reform | | |
|  | mean | sd | obs | mean | sd | obs |
| --- | --- | --- | --- | --- | --- | --- |
| Intervention | 0.13 | 0.34 | 1,305 | 0.15 | 0.36 | 1,112 |
| Full merger | 0.60 | 0.49 | 1,305 | 0.69 | 0.46 | 1,112 |
| Joint Venture | 0.35 | 0.48 | 1,305 | 0.19 | 0.39 | 1,112 |
| Number of concerned markets | 7.47 | 11.53 | 1,305 | 14.50 | 23.42 | 1,112 |
| EU acquirer | 0.69 | 0.46 | 1,305 | 0.64 | 0.48 | 1,112 |
| EU target | 0.74 | 0.44 | 1,305 | 0.70 | 0.46 | 1,112 |
| Indicator for July/August | 0.18 | 0.38 | 1,305 | 0.17 | 0.37 | 1,112 |
| Indicator for December | 0.06 | 0.23 | 1,305 | 0.05 | 0.21 | 1,112 |

Furthermore, the dataset contains information on the geographic market definition adopted in each market by DG Comp, distinguishing geographic markets that are defined as being national, EU wide, or worldwide. Lastly, the geographic market definition can also be left open.

Further indicator variables record whether DG Comp considered barriers to entry to exist and whether DG Comp raised concerns that the merger would foreclose other firms in a particular market. The database also contains a count of the number of competitors in the concerned market and an indicator variable equal to one if no information on competitors is available. Merging parties face about two competitors on average; however, information on competitors is missing in about 50% of the markets - these are mainly mergers that were cleared in phase-1.

As said before, I only use observations for which information on the market shares of the merging parties could be collected from DG Comp's competitive assessment in the decision document. Thus, data availability is constrained by the extent of DG Comp's analysis. Market share information is collected at the level of the relevant product/geographic market combination. Since the publicly available case documents generally report only market share ranges, the dataset contains the midpoint of the reported market share interval.[21] Therefore, I cannot avoid

---

[19]Commission Regulation (EC) No 802/2004 of 7 April 2004 implementing Council Regulation (EC) No 139/2004 on the control of concentration between undertakings [Official Journal L 133 of 30 April 2004]. The market share threshold has been raised to 30% at the end of 2013. See Commission Implementing Regulation (EU) No 1269/2013 of December 2013 [Official Journal L 336 of 14 December 2013].

[20]Conglomerate mergers are *"mergers between companies that are active in closely related markets (e.g. mergers involving suppliers of complementary products or products that belong to the same product range)."* Guidelines on the assessment of non-horizontal mergers under the Council Regulation on the control of concentrations between undertakings (2008/C 265/07), paragraph 5 [Official Journal of 18 October 2008].

[21]Since DG Comp generally reports only a range of market shares in the publicly available documents, the market shares are defined to be equal to the central value of the interval. If, for example, the market share range indicated is [0-10] percent, a market share of 5 percent is recorded. If however the interval given in the decision is only 5 percentage points wide, the conservative lower market share bound is reported. If for example the market share interval is [15-20] percent, 15 percent market share is reported.

16

that market shares contain measurement error leading to endogeneity bias. This is an issue that this study shares with the existing literature. To my knowledge, Mini (2018) is the only one who constructs expected market shares and expected concentration measures rather than using the midpoints of the market share ranges reported by DG Comp. He highlights the issue of measurement error in market shares and HHI and explicitly accounts for it in estimation. In the dataset we only recorded the midpoints of the market shares ranges. Thus, I cannot follow Mini (2018)'s approach. To my understanding it is also unclear in which direction the bias goes.[22] However, as we always recorded the lower bound of the market share range whenever the market share intervals given in the decision were only 5 percentage points wide, I expect that the market shares are underestimated. This measurement error leads to an attenuation bias in the estimated relation between market shares and intervention decision even if the size of the measurement error is not correlated with market share itself. Therefore, I expect that the relation between market shares and intervention decision is underestimated. As stated previously in relation to the selection issue though, I am interested in prediction and not in estimating causal effects of market shares or HHI on intervention decisions. Therefore, I do not require unbiased coefficients (see Section 5).

The market share information allows the calculation of the merging parties' combined market shares and the construction of a post-merger HHI. Table 2 also includes summary statistics for the market share related variables. The average merging parties' joint market share is slightly above 30%, with average post-merger HHI between 1,891 and 5,659 depending on the period and the calculation method. In the database, we include two different HHI measures. The variable *Post-merger HHI (low)* is a lower bound of the post-merger HHI: it is calculated as the square of the merging parties joint markets share plus the sum of squared market shares of competitors whenever information on competitors' market shares is available. This assumes that competitors are very small, whenever market share information of competitors is not available but market shares do not add up to 100%. The variable *Post-merger HHI (high)*, on the other hand, is an upper bound for the post-merger HHI: it adds the square of all missing market shares (100% minus all available market share information) to *Post-merger HHI (low)*. This hence treats all missing market share information as one missing competitor. In the empirical analysis, I use the *Post-merger HHI (high)* variable in order to be conservative as this measure should be an upper bound for market concentration.[23] If anything, I will overestimate market concentration and

---

[22]According to Mini (2018), the midpoint approach yields the correct expected value if the random variable has a symmetric marginal distribution which is centered on the midpoint. However, if for a merger case the sum of the upper bounds of all market share intervals (including competitors) is larger than 100%, the midpoint approach is no longer correct: according to Mini (2018) the domains of the marginal distributions of market shares are no longer necessarily the whole range from lower to upper market share bound and the distributions are even no longer necessarily symmetric. And even in cases where the midpoint of the market share range is the expected market share, the change in HHI as well as the contribution of the merging parties to the post-merger HHI would be underestimated.

[23]To the extent that I underestimate the merging parties' market shares, any HHI measure will also be underestimated. However, this is especially the case for the *Post-merger HHI (low)* measure as it assumes that any remaining competitors, for which no information is available, are very small. This is therefore an additional reason to use the *Post-merger HHI (high)* variable.

Electronic copy available at: https://ssrn.com/abstract=3379536

how problematic a given merger will be in a particular market. If a merger is unlikely to raise competitive concerns in a market for which concentration is measured by the upper bound of HHI, it will also not raise competitive concerns if the lower bound of HHI is used instead.

Table 4: Industry Groups, 1990-2014

| Industry group | Pre-reform | | Post-reform | |
|---|---|---|---|---|
| | obs | cases | obs | cases |
| Accomodation and food service | 56 | 11 | 66 | 10 |
| Agriculture, forestry, fishing, mining | 256 | 41 | 550 | 40 |
| Arts, other services, households as employers | 53 | 7 | 188 | 12 |
| Electricity, gas, steam | 181 | 47 | 617 | 56 |
| Financial service activities | 255 | 54 | 341 | 38 |
| Information and communication | 312 | 56 | 570 | 55 |
| Insurance and pensions | 299 | 61 | 303 | 42 |
| Manufacturing (coke, petroleum, chemicals) | 1,234 | 128 | 1,998 | 118 |
| Manufacturing (computer, electronics, optical products) | 394 | 56 | 881 | 82 |
| Manufacturing (food, beverages, tobacco) | 737 | 75 | 824 | 75 |
| Manufacturing (furnitures , other manufacturing) | 127 | 6 | 463 | 19 |
| Manufacturing (machinery and equipment) | 248 | 43 | 393 | 41 |
| Manufacturing (metals and metallic products) | 284 | 62 | 484 | 52 |
| Manufacturing (motor vehicles, trailers, transport equipment) | 582 | 96 | 496 | 65 |
| Manufacturing (pharmaceuticals) | 681 | 28 | 1,090 | 42 |
| Manufacturing (rubber, plastic, non-metallic) | 219 | 45 | 517 | 31 |
| Manufacturing (textiles, clothes, leather) | 26 | 4 | 87 | 5 |
| Manufacturing (wood, paper, printing) | 303 | 50 | 554 | 39 |
| Public administration, education, human health, social work | 28 | 8 | 62 | 6 |
| Real estate, professional activities, administrative service activities | 234 | 49 | 562 | 30 |
| Repair, installation of machinery and equipment | 649 | 91 | 101 | 15 |
| Telecommuications | 244 | 64 | 414 | 49 |
| Transporting and storage | 407 | 76 | 1,702 | 73 |
| Water supply, waste management, construction | 134 | 28 | 166 | 24 |
| Wholesale and retail trade | 588 | 119 | 852 | 93 |
| Total | 8,531 | 1,305 | 14,281 | 1,112 |

Lastly, the data include information on the main industry in which a merger took place. The industry is identified by NACE codes, which is the industry classification system used by the European Union to classify different economic activities. For the empirical analysis, I grouped the industries into 25 groups as shown in Table 4, where some NACE codes are grouped together while primarily the manufacturing industry has been further divided into smaller subgroups. For 280 pre-reform and 205 post-reform observations, the industry classification was missing. To avoid losing these observations in the analysis, I returned to the decision documents and manually assigned these mergers to the 25 industry groups.

I additionally define a few further variables. The variables *EU acquirer* and *EU target* are indicator variables that are equal to one if at least one of the acquiring or targeted firms is located within the EU, respectively. I include these variables in order to test for differential treatment of

18

mergers by DG Comp depending on the nationality.[24] In order to test whether the holiday season matters for the likelihood of intervention by DG Comp due to resource constraints during these periods, I define two further indicator variables for July/August and December respectively.

Comparing the summary statistics pre- and post-reform, the post-reform mergers seem to be slightly more often full mergers and less often joint ventures than pre-reform. Additionally, they concern more markets post-reform and more product/geographic markets are affected vertically post-reform. Lastly, post-reform, DG Comp defined more geographic markets to be worldwide in nature and identified entry barriers in more markets than pre-reform.

## 5. Prediction using Random Forests

In this paper, I use the random forest algorithm by Breiman (2001) implemented in the randomForest package in R (Liaw and Wiener, 2002) to predict whether DG Comp found competitive concerns in the markets affected by a merger and to evaluate how DG Comp's assessment and its predictability changed post EU merger reform. I chose the random forest algorithm over alternatives, such as logistic regression or lasso, because it is able to uncover highly flexible, non-linear functions in a high-dimensional feature space without over-fitting.

Random forests are one example of supervised machine learning techniques that typically use a set of features $X$ to predict an outcome $Y$. Thus, the goal is to construct $\hat{Y}(x)$, which is an estimator of $\mathbf{E}[Y|X = x]$, rather than estimating the causal effect of $X$ on $Y$ (Athey, 2018). Therefore, the aim is to reach a low error in the prediction $\hat{Y}$, which does not require the coefficients to be unbiased or causal. Our usual econometric tools, such as regression techniques, are made for causal inference, i.e. for obtaining unbiased estimates of the causal effect of covariates $X$ on outcome $Y$. As prediction error is a function of not only bias but also variance, these tools do not yield the most accurate prediction $\hat{Y}$ (Chalfin, Danieli, Hillis, Jelveh, Luca, Ludwig, and Mullainathan, 2016). The tools from machine learning, on the contrary, are designed to do exactly this: they adaptively use the data to decide on how to trade off bias and variance to maximize prediction performance while allowing for a rich set of covariates $X$ and functional forms (so higher order interaction terms or trees that allow for a high degree of interactivity between covariates). While the analyst has to provide the list of covariates $X$, the functional form is at least in part determined as a function of the data.

The random forest algorithm by Breiman (2001) uses regression or classification trees to predict an outcome $Y$. In a standard CART tree (Classification and Regression Tree), the goal is to predict individual outcomes $Y_i$ using the mean outcome $Y$ of observations that are "close" in $X$-space. To determine which observations are "close", the algorithm starts to recursively split the covariate space (binary splits) until it is partitioned into a set of so-called leaves $L$ that contain only a few observations. The outcome $Y_i$ for observation $i$ is then predicted by identifying the leaf containing observation $i$ based on its characteristics $X_i$ and setting the prediction to the mean

---

[24]See, for example, Bradford, Jackson, and Zytnick (2018), who empirically investigate whether EU merger control is used for protectionism.

outcome within that leaf:

$$\hat{Y}(x) = \frac{1}{|\{i : X_i \in L(x)\}|} \sum_{\{i : X_i \in L(x)\}} Y_i \qquad (1)$$

The algorithm automatically decides on the splitting variables and split points. This is done based on an in sample goodness-of-fit criterion (so essentially how close the predicted outcomes are to the actual outcomes). For regression trees (continuous outcome variable $Y$), the goodness-of-fit criterion used is the mean squared error; for classification trees (categorical outcome variable $Y$), the goodness-of-fit criterion is a measure of classification error based on the empirical classification probabilities in the leaves.[25] The algorithm then splits on the covariate at the cut-off value that leads to the greatest improvement in the goodness-of-fit criterion. Once the best split at a given node in the tree is found, the splitting process is repeated in each of the resulting two regions. For CART trees, the splitting process is usually stopped when a specified minimum node size is reached - by default this is a node size of 5 for regression and 1 for classification trees. The tree is then pruned based on some cost-complexity trade-off measure in order to avoid over-fitting (See Hastie, Tibshirani, and Friedman (2008) chapter 9 for further details).

A random forest is essentially an ensemble of regression or classification trees, where the predictions are averaged across trees.[26] A random forest introduces two layers of randomness compared to a single classification tree. First, each individual tree in the forest is grown using a random sample with replacement from the training set. In each tree, one-third of the data is not used for training and can be used for testing (out-of-bag error). Secondly, differently from CART trees, splitting a node in a tree is done based on only a random subset of the covariates $X$ rather than the full set of covariates and each tree is grown to the largest extent possible without pruning. The idea behind random forests is to reduce variance and produce more robust predictions compared to a single tree. Using a different bootstrap sample of the data to grow each tree in the forest as well as splitting each node based on only a subset of the covariates reduces the correlation between the trees in the forest and, therefore, the variance of the predictions (See Breiman (2001) and Hastie, Tibshirani, and Friedman (2008) chapter 15 for further details). Random forests are robust to overfitting and while one must choose the number of trees, the number of variables to be considered for splitting at each node, as well as the node size[27], it is argued that results do not usually seem to be very sensitive to the choice of these parameters (Liaw and Wiener, 2002). However, I use 5-fold cross-validation to tune these parameters (see below).

In order to predict the competitive concerns found by DG Comp and to evaluate how the assessment and its predictability changed post EU merger reform, I train two random forests: one based on pre-reform data and one based on post-reform data. For each of the two time periods,

---

[25]The randomForest R package uses the Gini index as measure for node impurity. Specifically, if $\hat{p}_{mk}$ is the proportion of class $k$ observations in node $m$, the Gini index for node $m$ is given by $\sum_{k=1}^{K} \hat{p}_{mk}(1 - \hat{p}_{mk})$.

[26]For classification problems, the random forest obtains a class vote from each tree and then classifies based on majority vote.

[27]The node size is the minimum size of final nodes within a tree.

I randomly split the data into one training set (80% of the observations) used for building the random forest and one hold-out set (20% of the observations) used for out-of-sample prediction. After training the two random forests on the two training datasets, I first use the pre-reform and post-reform prediction sets to evaluate the predictive performance of the random forests and then, secondly, to study how DG Comp's assessment differs pre- and post-reform.

The random forests are trained to predict competitive concerns in a particular market affected by the merger. Thus, instead of estimating the overall probability of an intervention *at the merger level*, I estimate the likelihood that competitive concerns are found in a particular market affected by the merger. However, this directly relates to the intervention decision of DG Comp, as the probability of intervention increases with the fraction of affected markets in which competitive concerns are found.

However, training the model at the market level, rather than the merger level, has an important implication. It is unlikely that the observations of the different markets affected by a merger are independent. On the contrary, it is likely that there are unobservables at the affected market level that are correlated across the different markets concerned by a given merger and that influence whether DGComp identified competitive concerns. Furthermore, also the observable characteristics are correlated across affected markets: Firstly, because some observable characteristics only vary at the merger level (and, hence, take exactly the same value across markets). Secondly, because concerned markets are often very similar markets where, for example, if entry barriers are found to be high in one market, they are likely to be high in another as well. If the data is then simply randomly split into training and hold-out set, for a given merger some concerned markets might end up in the training set and some might end up in the hold-out set used for prediction. Given that the outcome of competitive concerns is highly correlated across affected markets and that the observable characteristics of these observations are very similar or identical in some cases, the predictive performance of this random forest would be overstated if it was assessed based on this hold-out set. The predictive performance should instead be assessed based on a completely independent prediction set.

Consequently, I split the data into training and hold-out sets keeping all markets affected by a given merger together in one set to avoid leakage. I split the data randomly using the R package caTools, which allows for balancing the class distribution of the dependent variable in the training and hold-out sets (Tuszynski, 2014). This is particularly relevant in this application, as competitive concerns arise in less than 15% of the concerned markets and both the training as well as the hold-out set need to contain observations of both classes. An overview of the means both of the dependent variable *Concern* as well as of the variables used as predictors in the random forests is given in Table 5. As the table shows the concern rate of DG Comp is similar for the respective training and prediction sets. However, the table also shows that the training and prediction sets differ in many of the observables, given that the size of the dataset is not extremely large and that I require all markets affected by a merger to be allocated to the same subset of the data. Compared to a balanced split, where the training and prediction sets are similar in all dimensions, I impose a stricter performance test on my prediction models as

21

they also need to be able to classify a quite different dataset well.

Table 5: Mean Observables Training and Prediction Sets

|  | Pre-reform | | | Post-reform | | |
|  | Training mean | Prediction mean | Diff. p-value | Training mean | Prediction mean | Diff. p-value |
| --- | --- | --- | --- | --- | --- | --- |
| Concern | 0.14 | 0.14 | 0.99 | 0.13 | 0.13 | 0.99 |
| Entry barriers | 0.13 | 0.09 | 0.00 | 0.16 | 0.14 | 0.00 |
| Risk of foreclosure | 0.05 | 0.01 | 0.00 | 0.02 | 0.04 | 0.00 |
| Full merger | 0.73 | 0.67 | 0.00 | 0.77 | 0.77 | 0.58 |
| Joint Venture | 0.23 | 0.26 | 0.00 | 0.11 | 0.09 | 0.02 |
| Conglomerate merger | 0.05 | 0.03 | 0.01 | 0.00 | 0.01 | 0.01 |
| Vertical merger | 0.19 | 0.18 | 0.44 | 0.34 | 0.35 | 0.58 |
| National market | 0.68 | 0.59 | 0.00 | 0.66 | 0.59 | 0.00 |
| EU wide market | 0.23 | 0.29 | 0.00 | 0.20 | 0.28 | 0.00 |
| Worldwide market | 0.06 | 0.07 | 0.15 | 0.13 | 0.12 | 0.34 |
| Number of competitors | 1.81 | 2.04 | 0.00 | 1.93 | 2.47 | 0.00 |
| No competitor information | 0.45 | 0.42 | 0.02 | 0.47 | 0.40 | 0.00 |
| EU acquirer | 0.69 | 0.69 | 0.93 | 0.62 | 0.60 | 0.02 |
| EU target | 0.71 | 0.77 | 0.00 | 0.66 | 0.69 | 0.02 |
| Indicator for July/August | 0.16 | 0.17 | 0.28 | 0.15 | 0.18 | 0.00 |
| Indicator for December | 0.15 | 0.13 | 0.06 | 0.07 | 0.13 | 0.00 |
| Joint market share | 31.25 | 27.44 | 0.00 | 34.62 | 29.70 | 0.00 |
| Post-merger HHI (high) | 5,637.12 | 5,745.20 | 0.08 | 5,693.24 | 5,363.11 | 0.00 |
| Observations | 6,825.00 | 1,706.00 | . | 11,424.00 | 2,857.00 | . |

When growing a random forest, one needs to set some tuning parameters. First, I chose to grow forests containing 1,000 trees each, even though preliminary random forests showed that the overall out-of-bag error rate already stabilizes at about 200 to 300 trees. However, since the stability of the variable importance measures (see Section 6) depends on the number of trees in the random forest (Liaw and Wiener, 2002), I use 1,000 trees in each of the random forests.

In order to choose both the number of covariates that are considered when splitting a node as well as the node size, I perform 5-fold cross-validation over a tuning grid both for the pre-reform as well as the post-reform random forest. I first split the respective training set (pre-reform and post-reform) into 5 folds, once again keeping all markets concerned by a particular merger together in a fold. 5-fold cross-validation then implies that a random forest is grown using four of the folds as training set and the fifth fold for evaluating model performance, permuting the folds used in training and evaluation. Hence, five different random forests are grown for each set of parameters on the tuning grid, where I tune over the node size and the number of variables considered at each split.[28] For the parameter tuning stage, I use the train function implemented

---

[28]For the tuning grid, I use node sizes of 1, 3, 5, 10, 15, 20, 25, 30, 35 and 40, as well as number of covariates considered at each split of 2, 4, 5, 6, 7, 8, 9, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28 and 30. For classification, the default node size is one, while the default number of covariates that are considered for splitting a node is the square root of the number of covariates included in the random forest. As the default in this application would be to consider

22

in the caret package in R (Kuhn, 2008).[29]

Based on the results of this tuning stage, I choose a node size of 15 and to consider 12 variables at each split for the pre-reform random forest. The post-reform random forest has a node size of 20 but considers only 7 variables at each split. The results of the parameter tuning using 5-fold cross-validation show however that the model performance is relatively robust over the tuning range of node sizes and covariates considered at each split (see Appendix A.2).[30] The final models of the pre- and post-reform random forests are estimated using the full respective training set.

Lastly, I use the "classweight" option available in the randomForest package. The reason is that the random forest algorithm is constructed so as to minimize the overall error rate. Hence, it will tend to concentrate more on prediction accuracy of the majority class, which can result in poor prediction accuracy of the minority class (Chen, Liaw, and Breiman, 2004). In my dataset, competitive concerns arise in only about 14% of the affected markets pre-reform and about 13% of the affected markets post-reform. When I trained random forests without class weights, the out-of-bag error rate for the minority class (markets affected by the merger in which DG Comp found competitive concerns) was above 30%, while the out-of-bag error rate for the majority class was close to zero. However, a prediction model predicting competitive concerns of mergers should be able to filter out problematic cases well, rather than predicting no competitive concerns more or less by default just because in more than 80% of the observations in the training data, there are no competitive concerns. The classweight option of the randomForest package is based on the idea of cost sensitive learning: as the randomForest algorithm is biased towards the majority class, we give the minority class larger weight, i.e. we put a heavier penalty on misclassifying the minority class. According to Chen, Liaw, and Breiman (2004), these weights are used in two places when training the random forests: firstly, the class weights are used to weight the goodness-of-fit criterion when deciding on splitting variables and split points within the trees; secondly, the class prediction within each terminal node is done by weighted majority vote. I use class weights of 1 : 6 in training the random forests, as this corresponds roughly to the proportion of concern to no concern markets. Using the optimal parameters from the tuning

---

six covariates at each split, I chose a grid that is finer around the default value.

[29]Beside the tuning grid, one also needs to decide based on which summary metric the optimal model should be selected. For classification forests, such as in the present application, "accuracy" and "kappa" can be used. The default is to use accuracy, which is simply the fraction of all observations that are correctly classified. However, this metric does not distinguish the classes. In a highly imbalanced dataset with, for example, 80% class 1 observations and 20% class 2 observations, a classifier that would simply predict class 1 for all observations would not be very useful despite achieving accuracy of 80%. Kappa is similar to accuracy but is normalized at the baseline of random chance. Therefore, it compares the overall accuracy with expected random chance accuracy, where a Kappa of zero implies that the classifier performs no better than random classification. Given that the majority of mergers do not raise competitive concerns, I use Kappa to select the optimal model in the tuning stage.

[30]In particular, for the default values of a node size of one and six covariates considered at each split, the Kappa for the pre-reform random forest is about 0.5 compared to 0.52 for the optimal values chosen based on the cross-validation results. For the post-reform random forest, the Kappa for the default values is about 0.34 compared to 0.35 for the values chosen based on cross-validation. Furthermore, the two random forests using the default values for the node size and the number of covariates considered at each split also have very similar out-of-bag error rates as the random forests using the values chosen based on the cross-validation results.

stage, I however grew also pre- and post-reform random forests with class weights of $1:4$, $1:5$, and $1:7$ as a robustness check. These random forests have very similar out-of-bag error rates as the ones using class weights of $1:6$.

Figures 4 and 5 in Appendix A.3 plot the out-of-bag error rate of the final random forests, distinguishing overall, for class 1 (Concern), and for class 2 (No concern) over the number of trees in the forest. For the pre-reform random forest, the overall out-of-bag error stabilizes at about 9% already at about 200-300 trees. The overall out-of-bag error rate of the post-reform random forest stabilizes at around 15% even earlier. The plots also show that using class weights in the random forests allows for achieving a very low out-of-bag error rate for the minority class (about 8% in the pre-reform random forest, about 5% in the post-reform random forest).

The two random forests produce a prediction of whether competitive concerns arise in a particular market affected by the merger. This prediction is based on the class vote from each tree and then classifying based on majority vote: if more than 50% of the trees in the forest predict competitive concerns for a particular observation, the random forest will predict competitive concerns. However, it could also be that DG Comp's classification threshold is different, maybe DG Comp would even predict competitive concerns if the probability of concerns arising was lower than 50%. Any classification threshold between 0 and 1 leads to a predicted outcome that, when compared to the true outcome, can be classified as correctly predicted positives (true positive), correctly predicted negatives (true negatives), falsely predicted positives (false positives), and falsely predicted negatives (false negatives). A common metric measuring prediction precision is the area under the receiver operating characteristic curve (AUC). The receiver operating characteristic curve (ROC curve) plots the model's true positive rate versus the false positive rate as the classification threshold varies over $[0,1]$ (see Appendix A.4). For the pre-reform random forest, the AUC is 0.9711, for the post-reform random forests, the AUC is 0.9512, which is very high compared to, for example, an AUC of 0.707 reported by Kleinberg, Lakkaraju, Leskovec, Ludwig, and Mullainathan (2018) or an AUC between 0.62-0.71 reported by Björkegren and Grissen (2018). An AUC of 0.5 represents random prediction, while an AUC of 1 means perfect prediction.

## 6. Estimation Results

In this section, I present and discuss the results based on the two random forests trained as explained in the previous section. Specifically, I first discuss the predictive performance of the random forests compared to the benchmark of a Linear Probability Model (LPM) estimated on the same covariates as the random forests. Secondly, I look into how the assessment of competitive concerns by DG Comp changed post-merger reform and investigate for which type of mergers and affected markets the predictions between the pre-reform and post-reform random forests differ.

24

## 6.1. Predictive Performance

In this section, I discuss the predictive performance of the random forests trained on the pre-reform and post-reform data. In order to assess the ability of the random forests to correctly predict DG Comp's competitive concerns, I compare the results to predictions of DG Comp's competitive concerns based on pre- and post-reform LPM models.[31]

Figure 1 shows the variable importance for the random forest trained on the pre-reform and the random forest trained on the post-reform training datasets, respectively. The plot shows the weighted total decrease in node impurity from splitting on the respective covariate, where node impurity is measured by the Gini index (see Section 5).

Figure 1: Variable Importance Plot for Pre- and Post-Reform Random Forests
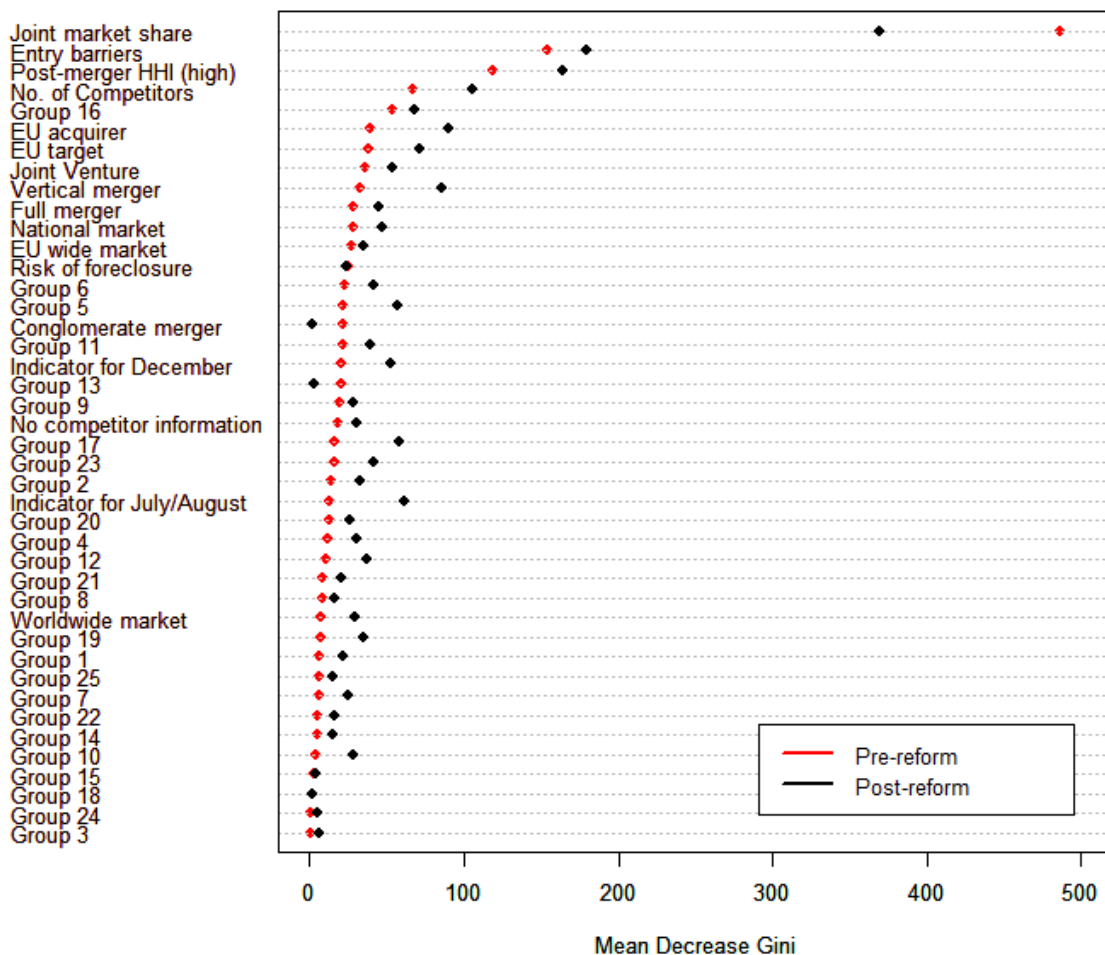


Although I am not primarily interested in the relative importance of the different determinants of a decision but in correct predictions, the variable importance plot essentially confirms the re-

---

[31]Instead of an LPM model, one could as well compare the predictive performance to a logit or probit model for competitive concerns raised by DG Comp.

sults of Affeldt, Duso, and Szücs (2019). While, in the pre-reform random forest, the joint market share of the merging parties is by far the most important covariate, its importance measured by the mean decrease in the Gini index decreases in the post-reform random forest. Furthermore, the variable importance plot shows that, while most of the same covariates show up in the upper part of the plot pre- and post-reform, the importance of covariates other than joint market share increases post-reform. In particular, the importance of entry barriers, the post-merger HHI, the number of competitors, vertical aspects, as well as whether the merging parties are from the EU increase in the post-reform random forest compared to the pre-reform random forest. Note also that most of the industry indicator variables seem not to be very important predictors. These results are in line with Affeldt, Duso, and Szücs (2019), who, without imposing different models pre- and post-reform, find that the importance of structural indicators of market power has declined over time.

I also run simple LPM models on both the pre- and post-reform training datasets containing the same merger and market characteristics as used in training the random forests (see Appendix A.5 for the regression results). I did this in order to compare the predictive performance of a parametric linear model with the predictions resulting from the highly non-linear and non-parametric random forests.

In line with the LPM results of Affeldt, Duso, and Szücs (2019), in these simple linear models, entry barriers as well as the merging parties' joint market share positively affect the probability that DG Comp will find competitive concerns in a given market affected by the merger both pre- and post-reform.[32] Furthermore, the magnitude of the estimated coefficients on entry barriers and joint market share decreases post-reform. Lastly, the $R^2$ is equal to 0.44 for the pre-reform LPM model but decreases to 0.28 in the post-reform LPM model, suggesting that DG Comp's merger assessment became less predictable post-reform. The generalized Hausman specification test clearly rejects the null hypothesis of no significant difference in model coefficients, indicating that separate models pre- and post-reform are in order.[33]

Both in the random forest and in the LPM models, the merging parties' joint market share, as well as entry barriers, seem to be the two most important determinants of competitive concerns by DG Comp. While in the pre-reform LPM model the coefficient on the risk of foreclosure is also positive and statistically significant, this variable is less prominent in the variable importance plots of the random forests. The results are also mostly in line with the existing literature, where the presence of entry barriers and the merging parties' combined market share are often found to be the most important predictors of an intervention decision by DG Comp (Bergman, Jakobsson,

---

[32]In the pre-reform model, the risk of foreclosure as well as a merger being notified in the summer season also positively affect the probability of DG Comp finding competitive concerns, while on the other hand vertical aspects as well as an EU wide geographic market definition negatively impact this probability. Post-reform, joint ventures, conglomerate aspects as well as the indicator variable for missing competitor information negatively affect the probability of DG Comp finding competitive concerns.

[33]The generalized Hausman specification test used to compare the estimated coefficients across models is based on the joint variance/covariance matrix of the models being tested. The null hypothesis of the Hausman test is that there is no significant difference in the model coefficients. This test is clearly rejected for the two LPM models (chi-squared statistic of 62,559).

and Razo, 2005; Bergman, Coate, Jakobsson, and Ulrick, 2010; Duso, Gugler, and Szücs, 2013; Mai, 2016). Mai (2016) also finds that the joint market share is a less important predictor post-reform. Importantly, Affeldt, Duso, and Szücs (2019) find these patterns over time without imposing different models pre- and post-reform.

Note, however, as explained previously, that the goal of the random forest algorithm is to correctly predict out-of-sample. Thus, I compare the predictive performance of the random forests to the predictive performance of the LPM models in the pre-reform and post-reform prediction datasets. Tables 6 and 7 show the actual as well as predicted concern rates (i.e. the percentage of markets with competitive concerns) by the pre-reform and post-reform random forest and LPM models, respectively. The models trained on the pre-reform training data should predict well the pre-reform prediction data, while the models trained on the post-reform training data should do well in predicting the post-reform prediction data.

The actual percentage of markets in which DG Comp identified competitive concerns was 13.6% in the pre-reform prediction set and 12.8% in the post-reform prediction set. The random forest trained on the pre-reform training set predicts the concern rate in the pre-reform prediction set very well and over-predicts the concern rate in the post-reform prediction set by only 1.5 percentage points. The random forest trained on the post-reform training set over-predicts the concern rate both in the pre-reform, but particularly in the post-reform prediction set, predicting a post-reform concern rate of 24.2%. On the other hand, both LPM models, estimated based on the pre-reform and post-reform training sets, respectively, under-predict the concern rates. In particular, the LPM model estimated based on the post-reform training set most severely under-predicts the concern rates: it predicts an intervention rate of only 2.8% and 5.1% for the pre-reform and post-reform prediction sets, respectively.

Table 6: Actual and Predicted Concern Rates - RF Model

| Dataset | Actual rate | Predicted rate (Pre-reform model) | Predicted rate (Post-reform model) |
|---------|-------------|-----------------------------------|------------------------------------|
| Pre-reform prediction set | 13.6 | 13.8 | 17.4 |
| Post-reform prediction set | 12.8 | 14.3 | 24.2 |

Table 7: Actual and Predicted Concern Rates - LPM Model

| Dataset | Actual rate | Predicted rate (Pre-reform model) | Predicted rate (Post-reform model) |
|---------|-------------|-----------------------------------|------------------------------------|
| Pre-reform prediction set | 13.6 | 5.9 | 2.8 |
| Post-reform prediction set | 12.8 | 8.3 | 5.1 |

Thus, judging based on the predicted concern rates, it seems that the random forest is better than the LPM model in predicting competitive concerns pre-reform. However, it is unclear

whether the random forest or the LPM model does better at predicting competitive concerns post-reform, where the random forest heavily over-predicts while the LPM model heavily under-predicts competitive concerns. Thus, I explain in detail why the random forest clearly outperforms the LPM model in the following. In particular, judging simply based on the predicted concern rate might be misleading when drawing conclusions about the predictive performance of different models. Ultimately, we should not care about whether the average concern rate is correctly predicted but in how far the model is able to correctly classify particular observations; i.e. whether the model is able to identify both class 1 and class 2 cases correctly. Tables 8 and 9 show the percentage of correct predictions for the pre- and post-reform prediction sets based on the random forest as well as the LPM models, respectively.

Table 8: Percentage of Correct Predictions - RF Model

|  | Pre-reform model | Post-reform model |
|---|---|---|
| Pre-reform prediction set |  |  |
| Concern | 58.2 | 67.2 |
| No concern | 93.1 | 90.4 |
| Overall | 88.4 | 87.3 |
| Post-reform prediction set |  |  |
| Concern | 38.8 | 61.2 |
| No concern | 89.3 | 81.2 |
| Overall | 82.8 | 78.6 |

Table 9: Percentage of Correct Predictions - LPM Model

|  | Pre-reform model | Post-reform model |
|---|---|---|
| Pre-reform prediction set |  |  |
| Concern | 30.2 | 15.9 |
| No concern | 97.9 | 99.3 |
| Overall | 88.7 | 87.9 |
| Post-reform prediction set |  |  |
| Concern | 44.3 | 31.4 |
| No concern | 97.0 | 98.7 |
| Overall | 90.3 | 90.1 |

As discussed in Section 3, most existing literature only presents the overall rate of correct predictions but does not distinguish between correct predictions of class 1 and class 2. Mai (2016) reports an overall rate of correct predictions above 80% and up to 90%, Bergman, Coate, Mai, and Ulrick (2016) report about 82% correct predictions for the EU models and Mini (2018) even reports about 90% correctly predicted observations. Table 8 shows between 88% and 90% correct predictions in the LPM - this is therefore in line with the existing literature. However, the random forests also predict between 79% and 88% of observations correctly, which is slightly lower than for the LPM (see Table 9). However, at the extreme, if about 90% of the observations

are class 1 and the model classifies 100% of the observations as class 1, then 90% of the cases are correctly predicted, even though the model is essentially useless as it is not able to correctly predict any of the rare class 2 cases. Consequently, I distinguish between predictions for markets without competitive concerns (*No concern*) and markets for which DG Comp found competitive concerns (*Concern*).

Looking at the results presented in Table 8, it is noticeable that the *No concern* outcome is predicted well by both the pre-reform as well as the post-reform random forest both on the pre-reform as well as the post-reform prediction sets. This is also true for the prediction of the *No concern* outcome by the pre-reform and post-reform LPM models presented in Table 9. The percentage of correctly predicted *No concern* observations is above 80% in all models. However, as *No concern* is the majority class in the data, it is not surprising that all the models do well in predicting this outcome.

Hence, the predictive performance of the different models should be judged based on how well these models are able to correctly classify observations of the minority class, i.e. markets in which DG Comp identified competitive concerns. Both the pre-reform and post-reform LPM models do very poorly in this respect: the pre-reform LPM model only correctly predicts about 30% of the competitive concern markets in the pre-reform prediction set, the post-reform LPM model only predicts about 31% of the markets with concerns in the post-reform prediction set correctly. The random forest models on the other hand, do significantly better in correctly predicting the minority class. The random forest trained on pre-reform data correctly predicts 58% of the competitive concern markets in the pre-reform prediction set and the random forest trained on post-reform data correctly predicts 61% of the markets raising competitive concerns in the post-reform prediction set.

This shows that the very flexible non-parametric random forests, which are designed to maximize predictive performance, do much better in correctly classifying observations into markets with and without competitive concerns than the LPM models presented for comparison. Even though the minority class is still harder to predict than the majority class, the percentage of correct predictions of competitive concerns achieved by the random forests is doubled compared to the percentage of correct predictions by the LPM models.[34]

One reason for the difference in predictive performance between the random forests and the LPM models is that the LPM imposes a linear and additively separable relationship between the covariates and the outcome variable. However, whether a merger raises competitive concerns in a particular market might be determined by a very complex combination of market

---

[34]As a robustness check I also computed the percentage of correct predictions based on a pre- and a post-reform random forest using the default values for the node size (one) and the number of covariates considered at each split (six in this case). These random forests do similar in predicting the *No concern* outcome with about 95% correct predictions for the pre-reform random forest on the pre-reform prediction set and about 86% for the post-reform forest on the post-reform prediction set. They do however worse in predicting the minority class: the pre-reform random forest correctly predicts about 55% of the *Concern* outcomes in the pre-reform prediction set, the post-reform random forest predicts about 47% of the *Concern* outcomes in the post-reform prediction set. Nevertheless, the two random forests still do much better in correctly classifying markets with competitive concerns than the LPM models.

and merger characteristics that is very unlikely to be linear and additively separable. A random forest on the other hand, is able to find these potentially complex and non-linear combinations of characteristics that determine whether competitive concerns arise or not and "cuts" the high dimensional characteristic space into regions separating problematic from non-problematic markets. For example, within a given tree, concerns might arise if there are entry barriers and a combined market share above 50% but also if there are no entry barriers, market share above 60% and the merger takes place within a certain industry.

Of course, one could add various interaction terms of covariates in the LPM. However, this can quickly become a very large number of interaction terms (even here with 42 covariates included in the random forests) and it is unclear *ex ante* which of these interaction terms are actually relevant for predicting the outcome. The random forest, on the other hand, performs model selection - i.e. deciding which combinations of covariates are actually relevant for predicting the outcome - and estimation of the model at the same time.

A second reason for why the random forest performs better than the LPM in correctly classifying the rare class is that by using class weights as one of the tuning parameters, the random forest is able to give the correct classification of minority class observations higher weight than the correct classification of majority class cases.

One caveat of using random forests for prediction tasks must be noted though. While random forests perform very well in predicting outcomes out-of-sample for observations that are similar in their covariates to what the random forest has seen before in the training data, random forests perform poorly in predicting outcomes for out-of-sample observations that are very different in their covariates to the cases contained in the training data. Therefore, if one expects to see out-of-sample observations that look very different from the training data, random forests might not be the optimal predictor. However, in this particular application, this is not a big issue as most of the covariates included in the random forests are dummy variables. Furthermore, for the continuous covariates of joint market share and post-merger HHI, the respective training sets contain observations with market shares ranging from 0% to 100% and post-merger HHI ranging from as low as 650 to the maximum of 10,000. The training data therefore covers the entire characteristics space.

Lastly, Table 8 also shows that the random forest trained on pre-reform data does well in predicting the minority class for the pre-reform prediction set but not for the post-reform prediction set. The equivalent is however not true for the random forest trained on post-reform data. Even though the post-reform random forest is trained on post-reform data only, it actually does slightly better in classifying cases from the pre-reform prediction set into concern and no concern markets than from the post-reform prediction set. Nevertheless, the post-reform random forest is still able to filter out competitive concern markets in the post-reform prediction set much better than the pre-reform random forest. This indicates that the models - and hence also the assessment rule of DG Comp - differ pre- and post-reform. I look at this aspect next.

## 6.2. Pre-Reform versus Post-Reform Predictions

While the random forest trained on pre-reform data does well in predicting outcomes for the pre-reform prediction set, the random forest trained on post-reform data does well in predicting outcomes for the post-reform prediction set. However, the pre-reform random forests does very poorly in predicting the minority outcome in the post-reform period. Based on the actual and predicted concern rates in Table 6, the so-called case mix and policy effects can be distinguished. Following the Oaxaca decomposition of Bergman, Coate, Jakobsson, and Ulrick (2010):

1. The difference between the actual post-reform concern rate and the predicted post-reform concern rate using the pre-reform random forest is the policy effect: $12.8\% - 14.3\% = -1.5$ percentage points

2. The difference between the predicted post-reform concern rate using the pre-reform random forest and the actual pre-reform concern rate is the case-mix effect: $14.3\% - 13.6\% = 0.7$ percentage points

This decomposition shows that the merger reform led to an overall slightly less aggressive merger policy by DG Comp. Mini (2018) reports an overall policy effect of about 7 percentage points, which is higher than what I find based on the predictions of the random forests. However, the markets contained in his study are not identical to those I include in my dataset, as Mini only considers horizontal overlap markets, i.e. markets in which both parties are active and where their combined market share is 15% or more.

The Oaxaca decomposition is based only on the average concern rate rather than investigating for which observations the predictions of the pre-reform and the post-reform models differ. Mini (2018) goes a step further and investigates how the importance of market shares and concentration measured by HHI differs pre- and post-reform. He reports the concern rates as well as the policy and case-mix effects by combined market share and HHI buckets, concluding that DG Comp did not change its stance towards mergers with very low market shares nor to the ones with very high combined market share (such as mergers to monopoly). However, in the middle ranges of joint market share, DG Comp challenged fewer mergers post-reform. Table 10 reports the difference in concern rates pre- and post-reform by markets share buckets for my dataset.[35]

Unlike Mini (2018), I find that the challenge rates are statistically significantly different from each other in all market share buckets. As in the results reported by Mini (2018), DG Comp is less likely to find competitive concerns post-reform in affected markets where the merging parties' market share is above 40%. However, I find that post-reform DG Comp is more likely to find competitive concerns in markets with merging parties' combined market share below 40%, whereas Mini (2018) finds no significant difference for markets with merging parties' market

---

[35]The table is based on the entire pre- and post-reform dataset, i.e. including both the observations from the training as well as the prediction sets. Note that I use the market shares as they are coded in my dataset, which is the midpoint of the interval reported in the decision documents. Therefore, I cannot be sure that actual combined market shares fall within that range. However, the table still provides an indication of the difference in concern rates pre- and post-reform.

Table 10: Concern Rates Pre- and Post-Reform by Combined Market Share

| Market share range (%) | Pre-reform | | | Post-reform | | | Difference | |
|---|---|---|---|---|---|---|---|---|
| | Obs. | Fraction of obs. (%) | Concern rate (%) | Obs. | Fraction of obs. (%) | Concern rate (%) | Rate diff. (post-pre) | Two-sided p-value |
| <30 | 4,734 | 55.5 | 1.80 | 7,559 | 52.9 | 3.69 | 1.90 | 0.000 |
| 30 - <40 | 1,257 | 14.7 | 5.65 | 2,182 | 15.3 | 8.02 | 2.37 | 0.009 |
| 40 - <50 | 854 | 10.0 | 23.19 | 1,418 | 9.9 | 15.09 | -8.09 | 0.000 |
| 50 - <60 | 580 | 6.8 | 34.66 | 959 | 6.7 | 27.95 | -6.71 | 0.006 |
| 60 - <70 | 395 | 4.6 | 45.06 | 584 | 4.1 | 37.16 | -7.91 | 0.013 |
| 70 - <80 | 276 | 3.2 | 57.61 | 456 | 3.2 | 42.76 | -14.85 | 0.000 |
| 80 - <90 | 201 | 2.4 | 64.68 | 349 | 2.4 | 47.85 | -16.83 | 0.000 |
| 90 - 100 | 234 | 2.7 | 58.55 | 774 | 5.4 | 40.44 | -18.11 | 0.000 |
| Total | 8,531 | 100.0 | 13.59 | 14,281 | 100.0 | 12.80 | -0.79 | 0.089 |

The table is based on the entire dataset, combining training and prediction sets pre- and post-reform. I use a two-sample test of proportions to determine whether the concern rates differ between the pre- and post-reform period.

shares below 30%. The difference to the results reported by Mini (2018) could be due to the fact that I include all markets in my dataset, whereas Mini only considers horizontal overlap markets. Furthermore, the dataset by Mini includes mergers up to and including 2013, while my dataset also contains merger cases notified in 2014. However, the EC raised the threshold for a horizontally affected market from 15% to 20% combined market share in December 2013. This could partly explain the higher post-reform concern rate for mergers with relatively low combined market share. Secondly, the higher post-reform concern rate for mergers with relatively low market shares also highlights one major goal of the merger reform: moving from the DT to the SIEC test made it easier to challenge mergers in markets where the merging parties are not dominant, while pre-reform the creation or strengthening of a dominant position was a necessary condition for the prohibition of a merger.

Mini (2018) only investigates the differences in concern rates as well as policy and case-mix effects between pre- and post-reform period by market share and concentration buckets. However, it could be the case that DG Comp's merger policy changed in more dimensions than just the importance of market share and concentration levels. In particular, the lower $R^2$ of the LPM model post-reform as well as decrease in the percentage of correct predictions post-reform also of the random forests, seems to suggest that DG Comp's assessment might be more complex and case-by-case post-reform. The fact that the post-reform random forest is still able to filter out competitive concern markets in the post-reform prediction set well and much better than the pre-reform random forest also suggests that decision criteria changed post-reform.

Table 11 shows the predicted *Concern* and *No concern* markets based on the two random forests. In particular, the models differ in their predictions of markets with competitive concerns. In order to investigate how the assessment of DG Comp differs pre- and post-reform, I look at the predictions for the post-reform prediction set only and how the predictions of the pre-reform and post-reform random forests differ for the post-reform period.

Table 11: Actual and Predicted Concerns - RF Model

|  | Pre-reform model | | Post-reform model | |
|  | Predicted Concern | Predicted No concern | Predicted Concern | Predicted No concern |
| --- | --- | --- | --- | --- |
| Pre-reform prediction set | | | | |
| Actual Concern | 135 | 97 | 156 | 76 |
| Actual No concern | 101 | 1,373 | 141 | 1,333 |
| Post-reform prediction set | | | | |
| Actual Concern | 142 | 224 | 224 | 142 |
| Actual No concern | 266 | 2,225 | 468 | 2,023 |

Table 12 shows the differences in the post-reform predictions of the two random forests distinguishing markets where DG Comp found and did not find competitive concerns. While in 106 markets, both random forests correctly predict competitive concerns (29% of actual markets with competitive concerns), they wrongly both predict no competitive concerns in 207 markets (57% of actual markets with competitive concerns). In 1,964 markets, both random forests correctly predict no concerns (79% of actual markets with no concerns) and in 106 markets they wrongly both predict competitive concerns (4% of actual markets with no concerns). This translates to an overall agreement rate in the predictions of the two random forests of about 83%. However, the interesting cases are in particular those affected markets for which the random forests trained on pre-reform and post-reform data differ in their predictions. In 118 markets where DG Comp found competitive concerns, the post-reform random forest correctly predicts those, while the pre-reform random forest predicts no concerns (32% of actual markets with competitive concerns). Secondly, in 59 markets for which DG Comp raised no concerns, the prediction of the post-reform random forest is correct while the pre-reform random forest incorrectly predicts competitive concerns in these markets (2% of actual markets with no concerns) and in 261 markets with no concerns, the prediction of the pre-reform random forest is correct while the post-reform random forest incorrectly predicts competitive concerns in these markets (10% of actual markets with no concerns).

Consequently, I study in which dimensions the affected markets differ depending on the prediction of the pre- and post-reform random forests. Table 13 compares the mean of the covariates of markets for which the pre- and the post-reform random forests predict competitive concerns, respectively. The markets for which the pre-reform random forest predicts competitive concerns are significantly more likely than the markets for which the post-reform random forest predicts competitive concerns to exhibit risk of foreclosure, to be vertically affected markets and defined

33

Table 12: Differences in Post-Reform Predictions by RF Models

| Predictions | Actual Concern | Actual No concern |
|---|---|---|
| Both pre- and post-model: Concern | 106 | 207 |
| Both pre- and post-model: No concern | 106 | 1,964 |
| Pre-model: No concern/Post-model: Concern | 118 | 261 |
| Pre-model: Concern/Post-model: No concern | 36 | 59 |
| Total | 366 | 2,491 |

as being EU wide geographic markets. On the other hand, they are significantly less likely to exhibit entry barriers and to be defined as national geographic markets. They are also markets in which the merging parties have lower joint market shares than the markets for which the post-reform model predicts competitive concerns.

Table 13: Equality of Means Test - Predicted Concern

| | Pre-reform model Concern mean | Post-reform model Concern mean | t statistic | two-sided p-value |
|---|---|---|---|---|
| Entry barriers | 0.48 | 0.59 | -3.75 | 0.000 |
| Risk of foreclosure | 0.12 | 0.04 | 4.52 | 0.000 |
| Full merger | 0.95 | 0.92 | 1.71 | 0.087 |
| Joint Venture | 0.01 | 0.03 | -3.17 | 0.002 |
| Conglomerate merger | 0.00 | 0.02 | -3.95 | 0.000 |
| Vertical merger | 0.33 | 0.21 | 4.14 | 0.000 |
| National market | 0.62 | 0.70 | -2.68 | 0.007 |
| EU wide market | 0.27 | 0.21 | 2.31 | 0.021 |
| Worldwide market | 0.11 | 0.09 | 1.06 | 0.289 |
| Number of competitors | 2.97 | 3.27 | -1.40 | 0.161 |
| No competitor information | 0.30 | 0.27 | 0.97 | 0.333 |
| EU acquirer | 0.62 | 0.40 | 7.11 | 0.000 |
| EU target | 0.78 | 0.80 | -0.94 | 0.348 |
| Indicator for July/August | 0.18 | 0.28 | -3.77 | 0.000 |
| Indicator for December | 0.17 | 0.08 | 4.13 | 0.000 |
| Joint market share | 46.93 | 51.77 | -3.30 | 0.001 |
| Post-merger HHI (high) | 5,029.36 | 4,865.25 | 1.25 | 0.211 |
| Observations | 1,100 | | | |

Table 14 compares the mean of the covariates of markets for which the pre- and the post-reform random forests predict no competitive concerns, respectively. The markets for which the pre-reform random forest predicts no competitive concerns are significantly less likely to exhibit entry barriers or risk of foreclosure than markets for which the post-reform random forests predicts no competitive concerns. Furthermore, the mergers are less likely to be full mergers

34

and the merging parties have lower combined market share in these markets than in those for which the post-reform random forest predicts no competitive concerns.

Table 14: Equality of Means Test - Predicted No Concern

| | Pre-reform model No concern mean | Post-reform model No concern mean | t statistic | two-sided p-value |
|---|---|---|---|---|
| Entry barriers | 0.03 | 0.06 | -5.20 | 0.000 |
| Risk of foreclosure | 0.01 | 0.04 | -5.81 | 0.000 |
| Full merger | 0.71 | 0.74 | -2.42 | 0.015 |
| Joint Venture | 0.12 | 0.11 | 1.87 | 0.061 |
| Conglomerate merger | 0.01 | 0.01 | 1.90 | 0.058 |
| Vertical merger | 0.36 | 0.37 | -1.11 | 0.269 |
| National market | 0.58 | 0.57 | 0.57 | 0.571 |
| EU wide market | 0.28 | 0.29 | -0.67 | 0.501 |
| Worldwide market | 0.13 | 0.13 | -0.07 | 0.945 |
| Number of competitors | 2.31 | 2.34 | -0.34 | 0.737 |
| No competitor information | 0.43 | 0.42 | 0.70 | 0.486 |
| EU acquirer | 0.59 | 0.63 | -2.74 | 0.006 |
| EU target | 0.66 | 0.67 | -0.70 | 0.485 |
| Indicator for July/August | 0.17 | 0.16 | 1.39 | 0.166 |
| Indicator for December | 0.12 | 0.14 | -2.06 | 0.039 |
| Joint market share | 24.19 | 26.02 | -3.38 | 0.001 |
| Post-merger HHI (high) | 5,469.79 | 5,446.05 | 0.33 | 0.738 |
| Observations | 4,614 | | | |

These differences between the predictions based on the pre-reform and post-reform random forests confirm that pre-reform DG Comp seems to have relied more on market shares in its merger assessment. Thus, the pre-reform model already predicts competitive concerns at lower market shares than the post-reform model and, conversely, the markets for which the pre-reform model predicts no competitive concerns have lower average combined market share than those markets for which the post-reform model still predicts no competitive concerns. Furthermore, the results show that post-reform DG Comp seems to assess mergers more on a case-by-case basis, where cases are not automatically considered to raise competitive concerns just because markets are concentrated, exhibit entry barriers, or are geographically narrowly defined, while mergers in broad geographic markets with relatively low market shares or vertical aspects might still raise competitive concerns. Therefore, although the overall intervention rate decreased slightly post-reform, DG Comp seems to base its assessment of competitive concerns more on a case-by-case economic analysis where a complex interaction between all merger characteristics determines the intervention decision. Of course these potentially complex interactions cannot be easily detected by a simple comparison of means as in Tables 13 and 14. However, as the prediction results show, the random-forest trained on the post-reform data is able to find these

relevant interactions of characteristics and filter out problematic markets much better than the pre-reform random forest or any simple linear model.

## 7. Conclusion

One goal of the 2004 EU merger reform was to bring merger control closer to economic principles. Another was to increase legal certainty and transparency of the merger review process as evidenced by the publication of merger guidelines and the institutional changes made. However, the effect of the reform on the predictability of DG Comp's decisions is ambiguous, as the use of a "more economic approach" in the merger review implies a shift from simple general rules, such as concentration thresholds, toward a more in depth case-by-case economic analysis. Thus, the question is whether the merger reform increased the *ex ante* predictability of decisions based on market and merger characteristics and also how the merger reform changed the decision criteria on which DG Comp bases its merger assessment.

In this paper, I study the predictability of DG Comp's merger policy and assess how it changed following the 2004 merger reform based on a comprehensive dataset covering almost all mergers notified to the EC between 1990 and 2014. Unlike most of the existing literature, rather than assessing mergers at the aggregate, the data is collected at a more fine-grained level, defining an observation as a particular product and geographic market combination concerned by a merger. This allows studying the factors that cause competitive concerns in specific sub-markets, as mergers typically concern several product and geographic markets that can be affected differently by the merger.

In addition, and unlike the existing literature studying the determinants of DG Comp's merger intervention decisions and their predictability, I use non-parametric random forests to predict DG Comp's assessment of competitive concerns arising in affected markets due to the merger. This machine learning algorithm is designed to maximize predictive performance rather than estimating causal effects and allows for highly flexible, non-linear interactions between covariates.

First, I find that the predictive performance of the random forests is much better than the performance of the LPM models I estimate for comparison reasons. While all models are able to predict the majority outcome of *no competitive concerns* very well (between 80% and 90% correct predictions), the LPM models do very poorly in predicting the minority outcome of *competitive concerns* with only between 16% and 44% correct predictions. Furthermore, based on these predictions as well as the $R^2$, the LPM models would wrongly suggest that the predictability of DG Comp's assessment decreased after the 2004 reform. The random forests however, are able to correctly classify the minority class cases in about 60% of the cases both pre- and post-reform. Thus, it is not true that the predictability of DG Comp's merger policy decreases post-reform.

Secondly, I study how the predictions of the pre-reform and post-reform random forests differ for the post-reform period. Based on the two random forests and correcting for the case mix effect, the policy effect shows a decrease in the concern rate of about 1.5 percentage points

36

post-reform. However, this decomposition only considers the average concern rate rather than investigating for which type of mergers and affected markets the predictions of the pre-reform and the post-reform random forests differ. Studying post-reform cases for which the predictions of the two random forests differ, I find that pre-reform DG Comp seems to have relied more on structural indicators, such as market shares and HHI, in its merger assessment, while post-reform DG Comp seems to base its assessment of competitive concerns more on a case-by-case analysis and less on simple structural indicators such as market shares or concentration measures. The highly flexible random forest algorithm is able to detect these potentially complex interactions between merger and market characteristics on which DG Comp's decision is based, therefore still allowing for high prediction precision compared to the overly simplistic LPM models.

# References

AFFELDT, P., T. DUSO, AND F. SZÜCS (2018): "EU Merger Control Database: 1990-2014," *DIW Berlin Data Documentation 95*.

——— (2019): "25 Years of European Merger Control," *DIW Discussion Paper No. 1797*.

ATHEY, S. (2018): "The Impact of Machine Learning on Economics," Mimeographed.

BERGMAN, M., M. B. COATE, A. MAI, AND S. W. ULRICK (2016): "Does Merger Policy Converge after the 2004 European Union Reforms?," Mimeographed.

BERGMAN, M. A., M. B. COATE, M. JAKOBSSON, AND S. W. ULRICK (2010): "Comparing Merger Policies in the European Union and the United States," *Review of Industrial Organization*, 36(4), 305–331.

BERGMAN, M. A., M. JAKOBSSON, AND C. RAZO (2005): "An econometric analysis of the European Commission's merger decisions," *International Journal of Industrial Organization*, 23(9-10), 717–737.

BJÖRKEGREN, D., AND D. GRISSEN (2018): "Behavior Revealed in Mobile Phone Usage Predicts Credit Repayment," Mimeographed.

BRADFORD, A., R. J. JACKSON, AND J. ZYTNICK (2018): "Is E.U. Merger Control Used for Protectionism? An Empirical Analysis," *Journal of Empirical Legal Studies*, 15(1), 165–191.

BREIMAN, L. (2001): "Random Forests," *Machine Learning*, 45(1), 5–32.

CHALFIN, A., O. DANIELI, A. HILLIS, Z. JELVEH, M. LUCA, J. LUDWIG, AND S. MULLAINATHAN (2016): "Productivity and Selection of Human Capital with Machine Learning," *American Economic Review: Papers and Proceedings*, 106(5), 124–127.

CHEN, C., A. LIAW, AND L. BREIMAN (2004): "Using Random Forests to Learn Imbalanced Data," Mimeographed.

DUSO, T. (2012): "A Decade of Ex-post Merger Policy Evaluations: A Progress Report," in *The Pros and Cons of Merger Control*, ed. by D. Sjöblom. Swedish Competition Authority.

DUSO, T., K. GUGLER, AND F. SZÜCS (2012): "An Empirical Assessment of the 2004 EU Merger Poliy Reform," *DICE Discussion Paper No.58*.

——— (2013): "An Empirical Assessment of the 2004 EU Merger Poliy Reform," *Economic Journal*, 123(572), F596–F619.

DUSO, T., K. GUGLER, AND B. B. YURTOGLU (2011): "How Effective is European Merger Control?," *European Economic Review*, 55(7), 980–1006.

DUSO, T., D. J. NEVEN, AND L.-H. RÖLLER (2007): "The Political Economy of European Merger Control: Evidence using Stock Market Data," *Journal of Law & Economics*, 50(3), 455–489.

ELMAN, P. (1965): "The Need for Certainty and Predictability in the Application of the Merger Law," *New York University Law Review*, 40, 613–627.

GERBER, D. J. (2014): "Searching for a Modernized Voice: Economics, Institutions, and Predictability in European Competition Law," *Fordham International Law Journal*, 37(5), 1421–1450.

Hastie, T., R. Tibshirani, and J. Friedman (2008): *The Elements of Statistical Learning - Data Mining, Inference, and Prediction*, Springer Series in Statistics. Springer New York Inc.

Kleinberg, J., H. Lakkaraju, J. Leskovec, J. Ludwig, and S. Mullainathan (2018): "Human Decisions and Machine Prediction," *Quarterly Journal of Economics*, 133(1), 237–293.

Kleinberg, J., J. Ludwig, S. Mullainathan, and Z. Obermeyer (2015): "Prediction Policy Problems," *American Economic Review*, 105(5), 491–495.

Kuhn, M. (2008): "Building Predictive Models in R Using the caret Package," *Journal of Statistical Software, Articles*, 28(5), 1–26.

Kwoka, J. E. (2013): "Does Merger Control Work? A Retrospective on U.S. Enforcement Actions and Merger Outcomes," *Antitrust Law Journal*, 78(3), 619–650.

Liaw, A., and M. Wiener (2002): "Classification and Regression by randomForest," *R News*, 2(3), 18–22.

Lyons, B. R. (2004): "Reform of European merger policy," *Review of International Economics*, 12(2), 246–261.

Mai, A. T. V. (2016): "Is EU Merger Policy Less Stringent After Its 2004 Reform?," *PESO Working Papers 2016:1*.

McAfee, P. R. (2010): "Transparency and Antitrust Policy," Mimeographed.

Mini, F. (2018): "Fifty is the New Forty: EU Merger Policy Permits Higher Market Shares After the 2004 Reform," *Review of Industrial Organization*, 53(3), 535–561.

Patterson, D., and C. Shapiro (2001): "Transatlantic Divergence in GE/Honeywell: Causes and Lessons," *Antitrust Magazine*, 16, 18–26.

Ribers, M., and H. Ullrich (2018): "Battling Resistance: Using Machine Prediction to Improve Antibiotic Prescribing," Mimeographed.

Smith, B. (1958): "Precedent, Public Policy and Predictability," *Georgetown Law Journal*, 46, 632–645.

Szücs, F. (2012): "Investigating transatlantic merger policy convergence," *International Journal of Industrial Organization*, 30(6), 654–662.

Tuszynski, J. (2014): *caTools: Tools: moving window statistics, GIF, Base64, ROC AUC, etc.* R package version 1.17.1.

# A. Appendix

## A.1. Summary Statistics Entire Dataset

Table 15: Summary Statistics Variables at Market Level (Entire Dataset)

|  | mean | sd | min | max | observations |
|---|---|---|---|---|---|
| Concern | 0.11 | 0.31 | 0 | 1 | 30,995 |
| Vertical merger | 0.26 | 0.44 | 0 | 1 | 30,995 |
| Conglomerate merger | 0.02 | 0.13 | 0 | 1 | 30,995 |
| National market | 0.58 | 0.49 | 0 | 1 | 30,995 |
| EU wide market | 0.20 | 0.40 | 0 | 1 | 30,995 |
| Worldwide market | 0.10 | 0.29 | 0 | 1 | 30,995 |
| Left open market | 0.12 | 0.33 | 0 | 1 | 30,995 |
| Entry barriers | 0.12 | 0.32 | 0 | 1 | 30,995 |
| Risk of foreclosure | 0.03 | 0.16 | 0 | 1 | 30,995 |
| Number of competitors | 1.58 | 2.33 | 0 | 34 | 30,995 |
| No competitor information | 0.56 | 0.50 | 0 | 1 | 30,995 |
| Joint market share | 32.46 | 23.60 | 0 | 100 | 22,812 |
| Post-merger HHI (low) | 2,147.73 | 2,368.30 | 0 | 10,000 | 22,812 |
| Post-merger HHI (high) | 5,638.99 | 2,251.08 | 650 | 10,000 | 22,812 |

Table 16: Summary Statistics Variables at Merger Level (Entire Dataset)

|  | mean | sd | min | max | observations |
|---|---|---|---|---|---|
| Intervention | 0.07 | 0.26 | 0 | 1 | 5,109 |
| Full merger | 0.55 | 0.50 | 0 | 1 | 5,109 |
| Joint Venture | 0.37 | 0.48 | 0 | 1 | 5,109 |
| Number of concerned markets | 6.07 | 13.43 | 1 | 245 | 5,109 |
| EU acquirer | 0.66 | 0.47 | 0 | 1 | 5,109 |
| EU target | 0.72 | 0.45 | 0 | 1 | 5,109 |
| Indicator for July/August | 0.18 | 0.39 | 0 | 1 | 5,109 |
| Indicator for December | 0.06 | 0.23 | 0 | 1 | 5,109 |

## A.2. Random Forest Tuning Stage

The two plots show the results of the tuning of the node size and the number of covariates considered at each split using 5-fold cross-validation. For the pre-reform random forest, the highest kappa is achieved with a node size of 15 and considering 12 randomly selected covariates at each split. For the post-reform random forest, the highest kappa is achieved for a node size of 20 and considering only 7 randomly selected predictors at each split.

Figure 2: Parameter Tuning Pre-Reform Random Forest
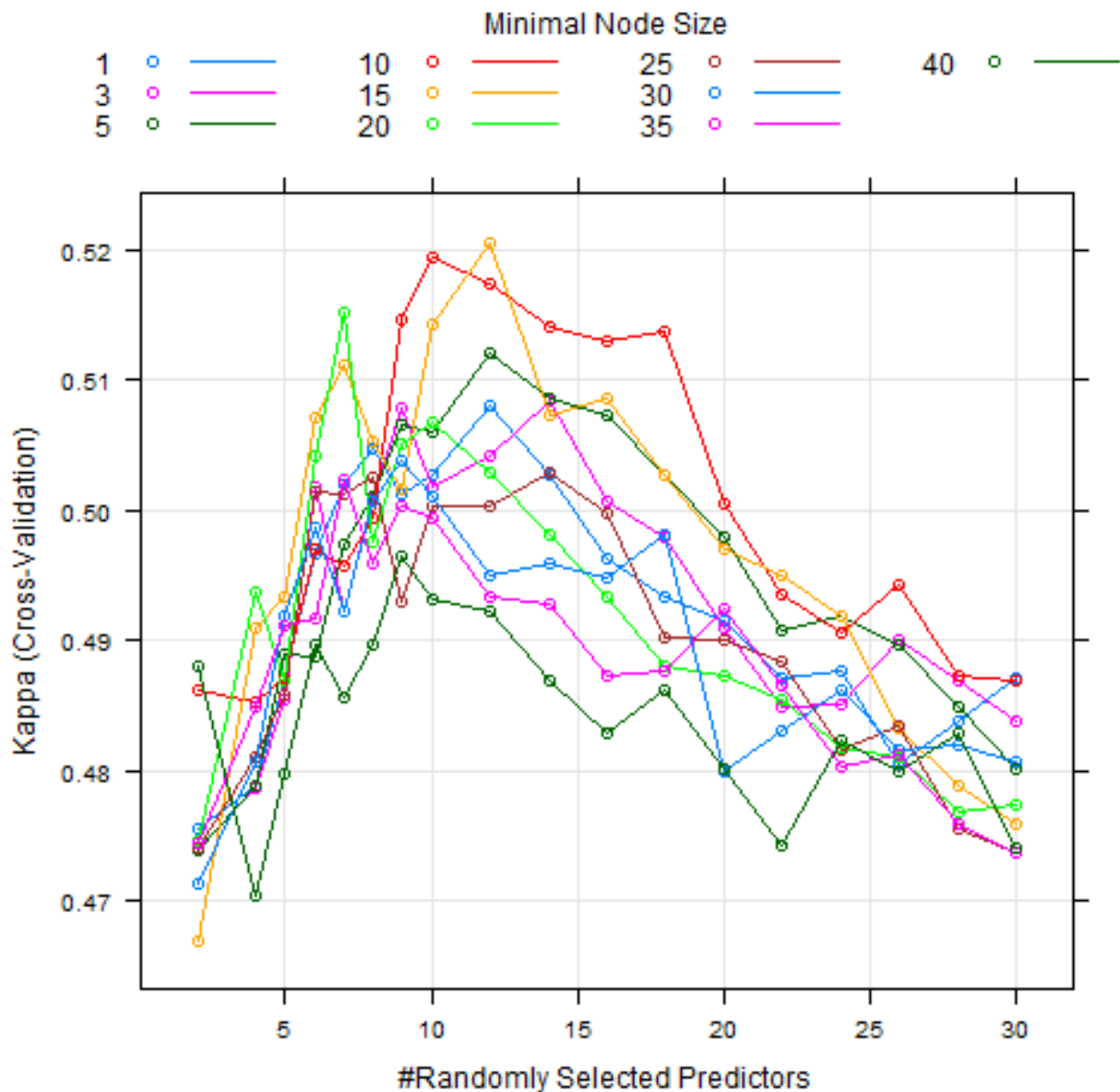
Figure 3: Parameter Tuning Post-Reform Random Forest
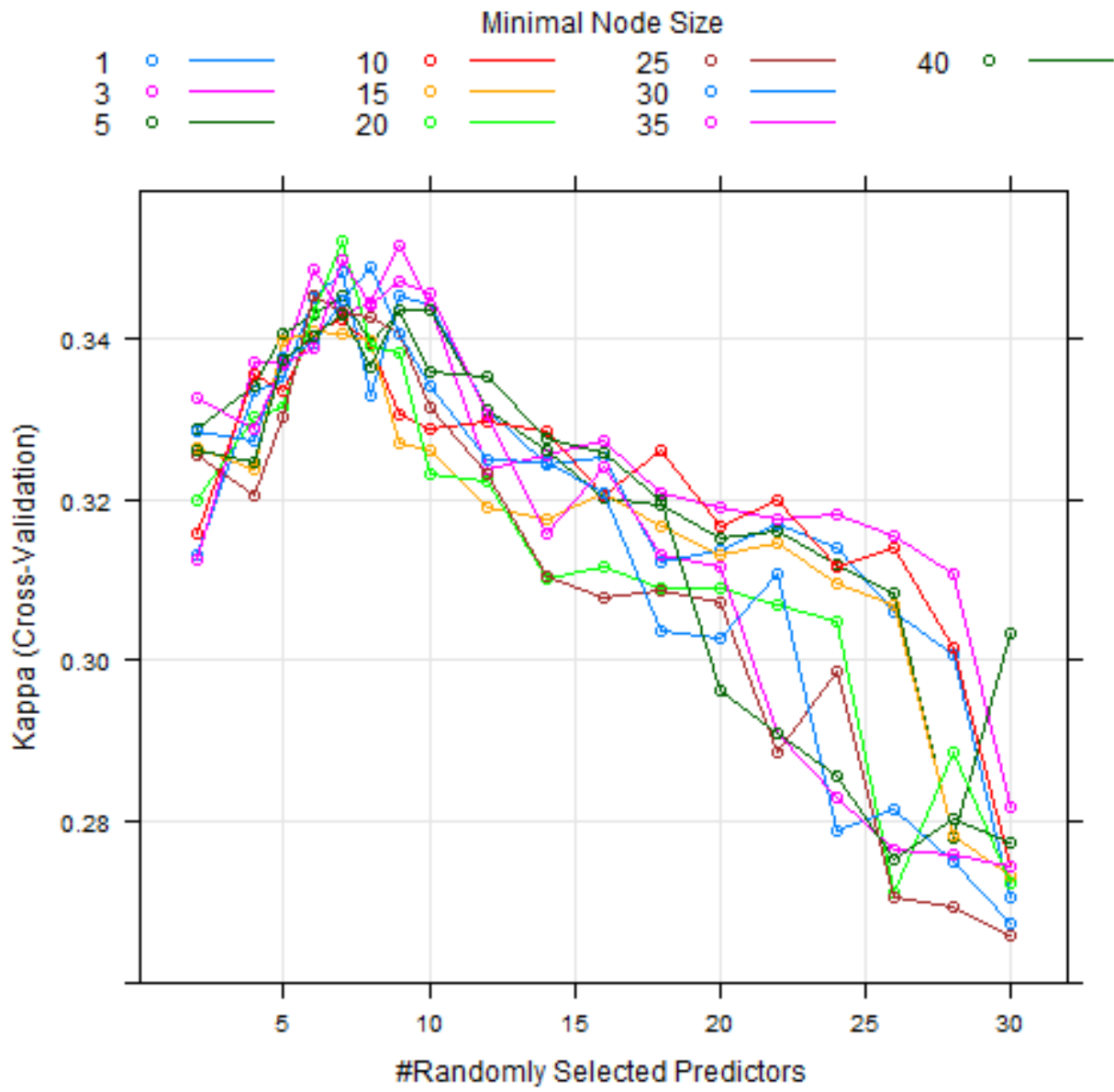
## A.3. OOB Error Rate

The two plots show the development of the out-of-bag error rate as the number of trees included in each of the random forests increases. As random forests bootstrap the training data when constructing the individual trees, one can evaluate the prediction error for an observation by computing the mean prediction error using only the trees in the forest which do not include this particular observation - this is the out-of-bag error (OOB error).
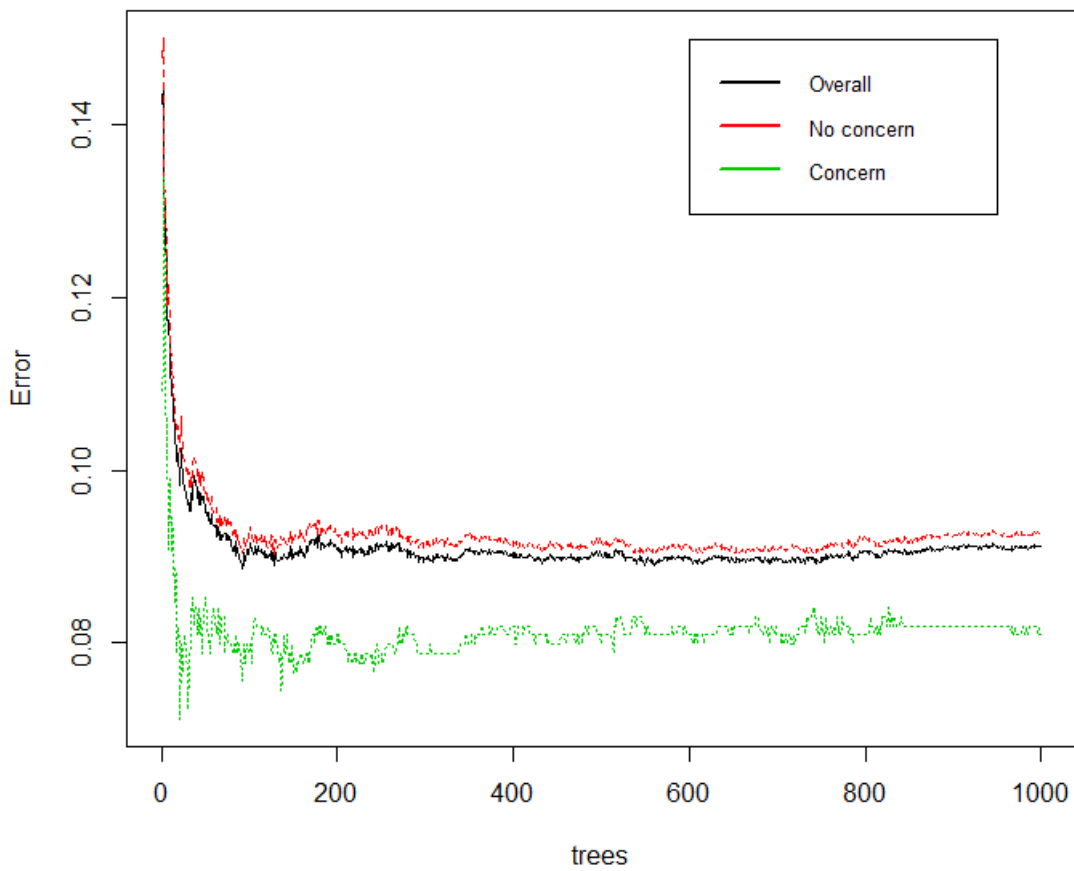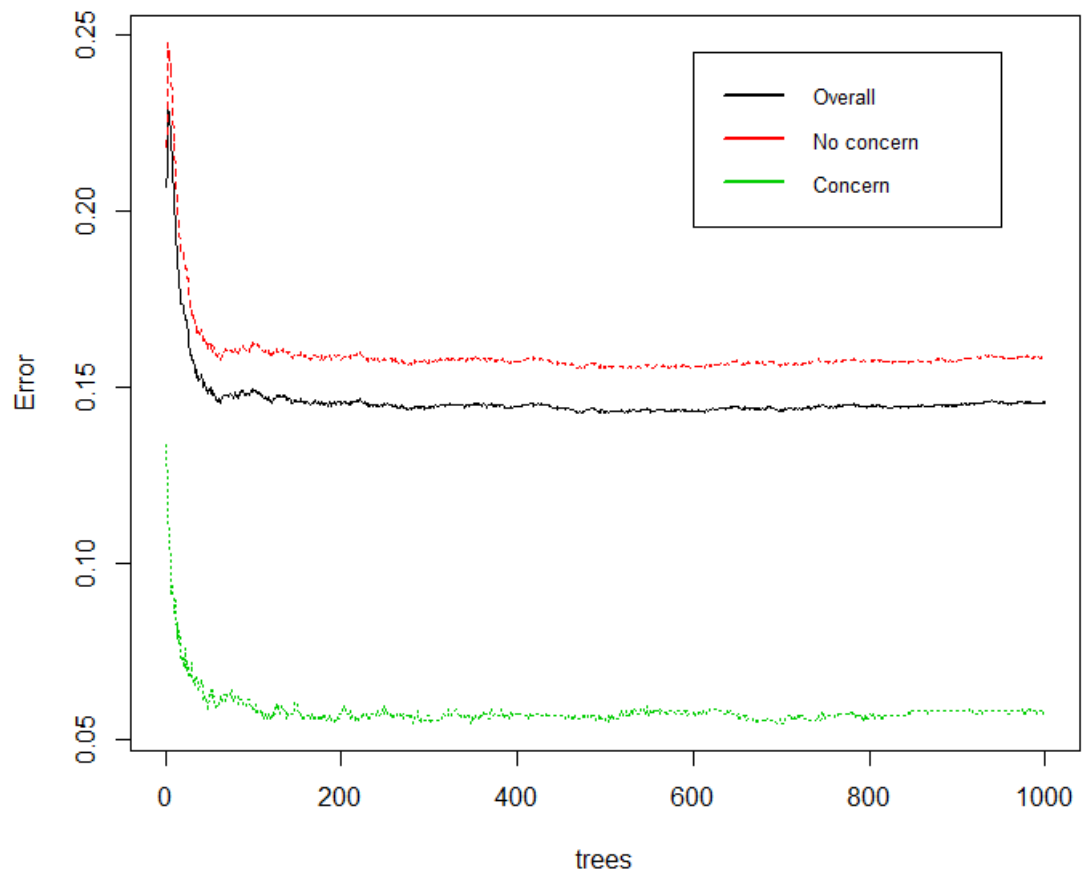
Figure 4: OOB Error for Pre-Reform Random Forest



43

Figure 5: OOB Error for Post-Reform Random Forest

## A.4. ROC Curve

The two plots show the receiver operating characteristic (ROC) curves for each of the two trained random forests. The ROC curve plots the model's true positive rate (sensitivity) versus the false positive rate (1-specificity) as the classification threshold varies over $[0, 1]$. While the 45 degree line represents random guessing, any point above the 45 degree line is better than random guessing. Points in the top left corner would represent perfect prediction with no false positives and no false negatives. The area under the curve (AUC) is a commonly used measure of precision. For the pre-reform random forest, the AUC is 0.9711, for the post-reform random forests, the AUC is 0.9512, which is very high compared to for example an AUC of 0.707 reported by Kleinberg, Lakkaraju, Leskovec, Ludwig, and Mullainathan (2018) or an AUC between 0.61-0.77 reported by Björkegren and Grissen (2018). Again, an AUC of 0.5 represents random prediction, while an AUC of 1 represents perfect prediction.
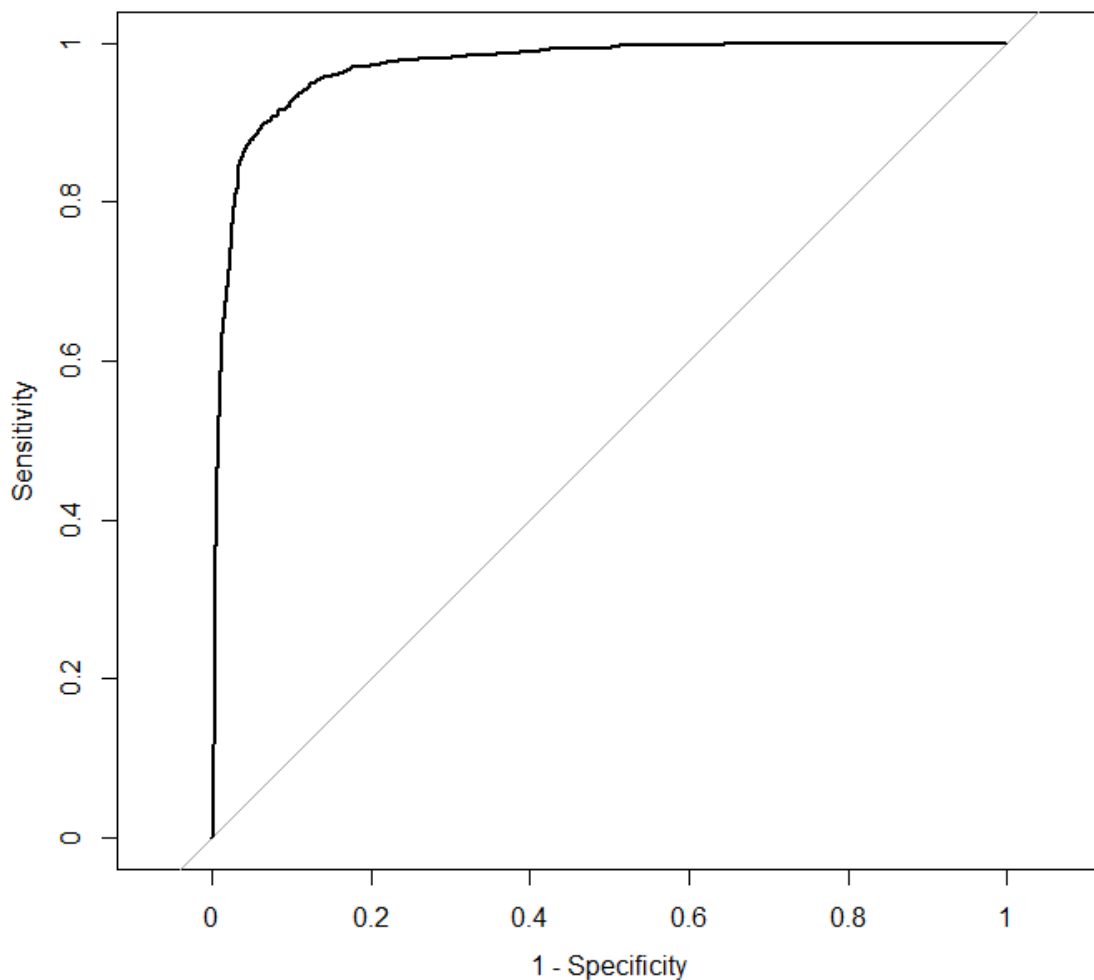
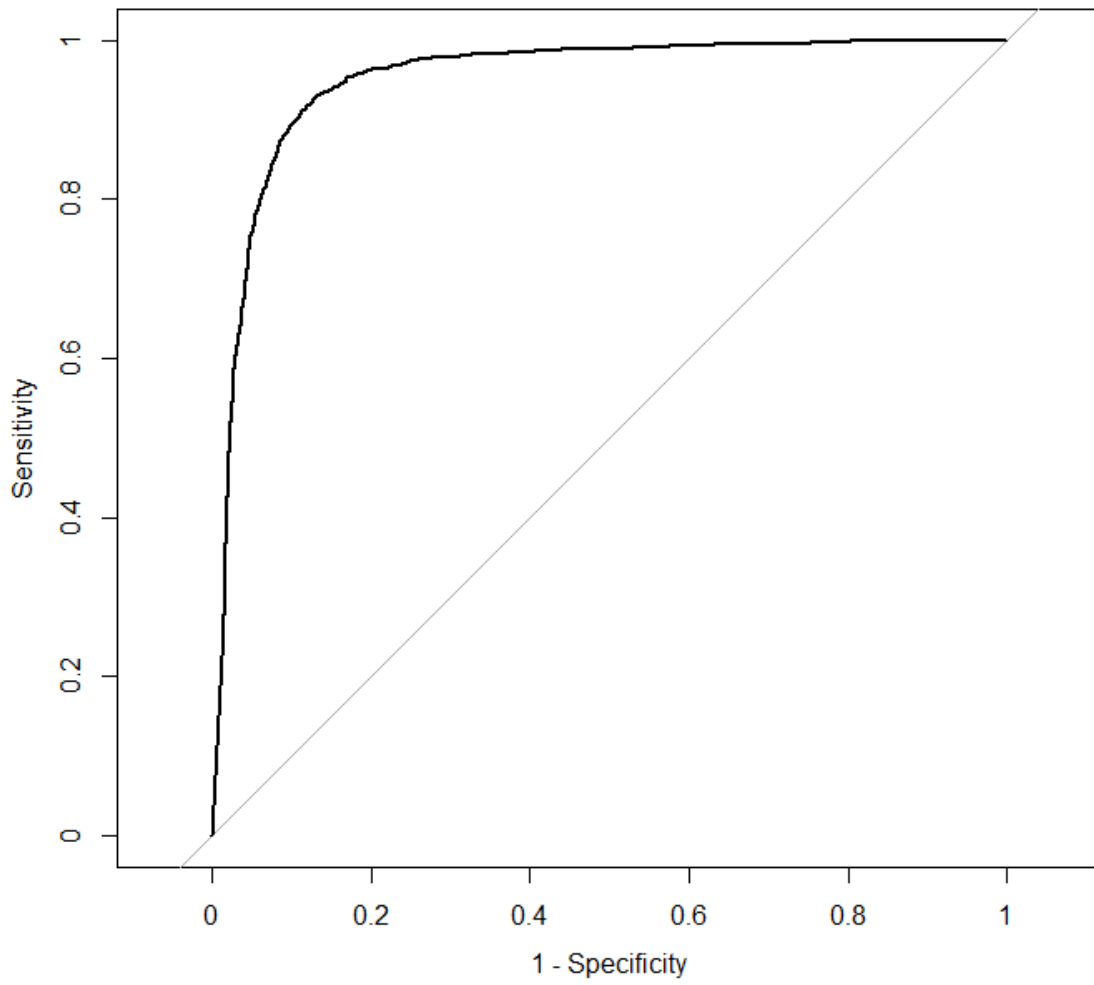Figure 6: ROC Curve for Pre-Reform Random Forest



45

Figure 7: ROC Curve for Post-Reform Random Forest

## A.5. Linear Probability Model

Table 17: Linear Probability Model for Concern (Market Level)

| | (1) Pre-reform | (2) Post-reform |
|---|---|---|
| Joint market share | 0.0050*** | 0.0032*** |
| | (0.0006) | (0.0005) |
| Post-merger HHI (high) | 0.0000** | 0.0000* |
| | (0.0000) | (0.0000) |
| Entry barriers | 0.3927*** | 0.3148*** |
| | (0.0511) | (0.0865) |
| Risk of foreclosure | 0.1435*** | 0.0478 |
| | (0.0345) | (0.1715) |
| Full merger | -0.0019 | -0.0319 |
| | (0.0363) | (0.0260) |
| Joint Venture | -0.0466 | -0.0846** |
| | (0.0320) | (0.0308) |
| Conglomerate merger | -0.0024 | -0.1008** |
| | (0.0308) | (0.0469) |
| Vertical merger | -0.0359** | 0.0136 |
| | (0.0148) | (0.0326) |
| National market | -0.0164 | -0.1224 |
| | (0.0153) | (0.0914) |
| EU wide market | -0.0309*** | -0.0861 |
| | (0.0099) | (0.0917) |
| Worldwide market | 0.0404 | -0.1373 |
| | (0.0337) | (0.1093) |
| Number of competitors | -0.0042 | -0.0008 |
| | (0.0036) | (0.0050) |
| No competitor information | -0.0408 | -0.0637*** |
| | (0.0328) | (0.0213) |
| EU acquirer | 0.0319* | 0.0518 |
| | (0.0155) | (0.0347) |
| EU target | 0.0111 | -0.0045 |
| | (0.0186) | (0.0288) |
| Indicator for July/August | 0.0533** | -0.0010 |
| | (0.0227) | (0.0396) |
| Indicator for December | -0.0102 | 0.1070** |
| | (0.0412) | (0.0496) |
| Constant | -0.0445 | 0.1160 |
| | (0.0590) | (0.1201) |
| Industry Group FE | Yes | Yes |
| R2 | 0.439 | 0.280 |
| Observations | 6,825 | 11,424 |

We report heteroskedasticity robust standard errors clustered at the industry group level.

Significance at the 1%, 5%, and 10% levels is represented by ***,** and * respectively.