Journal of Economic Perspectives—Volume 28, Number 2—Spring 2014—Pages 3–28

Big Data: New Tricks for Econometrics†

Hal R. Varian

Computers are now involved in many economic transactions and can capture data associated with these transactions, which can then be manipulated and analyzed. Conventional statistical and econometric techniques such as regression often work well, but there are issues unique to big datasets that may require different tools.

First, the sheer size of the data involved may require more powerful data manipulation tools. Second, we may have more potential predictors than appropriate for estimation, so we need to do some kind of variable selection. Third, large datasets may allow for more flexible relationships than simple linear models. Machine learning techniques such as decision trees, support vector machines, neural nets, deep learning, and so on may allow for more effective ways to model complex relationships.

In this essay, I will describe a few of these tools for manipulating and analyzing big data. I believe that these methods have a lot to offer and should be more widely known and used by economists. In fact, my standard advice to graduate students these days is go to the computer science department and take a class in machine learning. There have been very fruitful collaborations between computer scientists and statisticians in the last decade or so, and I expect collaborations between computer scientists and econometricians will also be productive in the future.

■ *Hal Varian is Chief Economist, Google Inc., Mountain View, California, and Emeritus Professor of Economics, University of California, Berkeley, California. His email address is hal@ischool.berkeley.edu.*

† To access the Appendix and disclosure statements, visit
http://dx.doi.org/10.1257/jep.28.2.3 doi=10.1257/jep.28.2.3

Tools to Manipulate Big Data

Economists have historically dealt with data that fits in a spreadsheet, but that is changing as new more-detailed data becomes available (see Einav and Levin 2013, for several examples and discussion). If you have more than a million or so rows in a spreadsheet, you probably want to store it in a relational database, such as MySQL. Relational databases offer a flexible way to store, manipulate, and retrieve data using a Structured Query Language (SQL), which is easy to learn and very useful for dealing with medium-sized datasets.

However, if you have several gigabytes of data or several million observations, standard relational databases become unwieldy. Databases to manage data of this size are generically known as "NoSQL" databases. The term is used rather loosely, but is sometimes interpreted as meaning "not only SQL." NoSQL databases are more primitive than SQL databases in terms of data manipulation capabilities but can handle larger amounts of data.

Due to the rise of computer-mediated transactions, many companies have found it necessary to develop systems to process billions of transactions per day. For example, according to Sullivan (2012), Google has seen 30 trillion URLs, crawls over 20 billion of those a day, and answers 100 billion search queries a month. Analyzing even one day's worth of data of this size is virtually impossible with conventional databases. The challenge of dealing with datasets of this size led to the development of several tools to manage and analyze big data.

A number of these tools are proprietary to Google, but have been described in academic publications in sufficient detail that open-source implementations have been developed. Table 1 contains both the Google name and the name of related open-source tools. Further details can be found in the Wikipedia entries associated with the tool names.

Though these tools can be run on a single computer for learning purposes, real applications use large clusters of computers such as those provided by Amazon, Google, Microsoft, and other cloud-computing providers. The ability to rent rather than buy data storage and processing has turned what was previously a fixed cost of computing into a variable cost and has lowered the barriers to entry for working with big data.

Tools to Analyze Data

The outcome of the big-data processing described above is often a "small" table of data that may be directly human readable or can be loaded into an SQL database, a statistics package, or a spreadsheet. If the extracted data is still inconveniently large, it is often possible to select a subsample for statistical analysis. At Google, for example, I have found that random samples on the order of 0.1 percent work fine for analysis of business data.

Once a dataset has been extracted, it is often necessary to do some exploratory data analysis along with consistency and data-cleaning tasks. This is something

Table 1
Tools for Manipulating Big Data

Google name	*Analog*	*Description*
Google File System	Hadoop File System	This system supports files so large that they must be distributed across hundreds or even thousands of computers.
Bigtable	Cassandra	This is a table of data that lives in the Google File System. It too can stretch over many computers.
MapReduce	Hadoop	This is a system for accessing and manipulating data in large data structures such as Bigtables. MapReduce allows you to access the data in parallel, using hundreds or thousands of machines to extract the data you are interested in. The query is "mapped" to the machines and is then applied in parallel to different shards of the data. The partial calculations are then combined ("reduced") to create the summary table you are interested in.
Sawzall	Pig	This is a language for creating MapReduce jobs.
Go	None	Go is flexible open-source, general-purpose computer language that makes it easier to do parallel data processing.
Dremel, BigQuery	Hive, Drill, Impala	This is a tool that allows data queries to be written in a simplified form of of Structured Query Language (SQL). With Dremel it is possible to run an SQL query on a petabtye of data (1,000 terabytes) in a few seconds.

of an art, which can be learned only by practice, but data-cleaning tools such as OpenRefine and DataWrangler can be used to assist in data cleansing.

Data analysis in statistics and econometrics can be broken down into four categories: 1) prediction, 2) summarization, 3) estimation, and 4) hypothesis testing. Machine learning is concerned primarily with prediction; the closely related field of data mining is also concerned with summarization, and particularly with finding interesting patterns in the data. Econometricians, statisticians, and data mining specialists are generally looking for insights that can be extracted from the data. Machine learning specialists are often primarily concerned with developing high-performance computer systems that can provide useful predictions in the presence of challenging computational constraints. Data science, a somewhat newer term, is concerned with both prediction and summarization, but also with data manipulation, visualization, and other similar tasks. Note that terminology is not standardized in these areas, so these descriptions reflect general usage, not hard-and-fast definitions. Other terms used to describe computer-assisted data analysis include knowledge extraction, information discovery, information harvesting, data archaeology, data pattern processing, and exploratory data analysis.

Much of applied econometrics is concerned with detecting and summarizing relationships in the data. The most common tool used for summarization is (linear) regression analysis. As we shall see, machine learning offers a set of tools that can usefully summarize various sorts of nonlinear relationships in the data. We will focus on these regression-like tools because they are the most natural for economic applications.

In the most general formulation of a statistical prediction problem, we are interested in understanding the conditional distribution of some variable y given some other variables $x = (x_1, \ldots, x_P)$. If we want a point prediction, we can use the mean or median of the conditional distribution.

In machine learning, the x-variables are usually called "predictors" or "features." The focus of machine learning is to find some function that provides a good prediction of y as a function of x. Historically, most work in machine learning has involved cross-section data where it is natural to think of the data being independent and identically distributed (IID) or at least independently distributed. The data may be "fat," which means lots of predictors relative to the number of observations, or "tall" which means lots of observations relative to the number of predictors.

We typically have some observed data on y and x, and we want to compute a "good" prediction of y given new values of x. Usually "good" means it minimizes some loss function such as the sum of squared residuals, mean of absolute value of residuals, and so on. Of course, the relevant loss is that associated with *new* out-of-sample observations of x, not the observations used to fit the model.

When confronted with a prediction problem of this sort an economist would think immediately of a linear or logistic regression. However, there may be better choices, particularly if a lot of data is available. These include nonlinear methods such as 1) classification and regression trees (CART); 2) random forests; and 3) penalized regression such as LASSO, LARS, and elastic nets. (There are also other techniques, such as neural nets, deep learning, and support vector machines, which I do not cover in this review.) Much more detail about these methods can be found in machine learning texts; an excellent treatment is available in Hastie, Tibshirani, and Friedman (2009), which can be freely downloaded. Additional suggestions for further reading are given at the end of this article.

General Considerations for Prediction

Our goal with prediction is typically to get good *out-of-sample predictions*. Most of us know from experience that it is all too easy to construct a predictor that works well in-sample but fails miserably out-of-sample. To take a trivial example, n linearly independent regressors will fit n observations perfectly but will usually have poor out-of-sample performance. Machine learning specialists refer to this phenomenon as the "overfitting problem" and have come up with several ways to deal with it.

First, since simpler models tend to work better for out-of-sample forecasts, machine learning experts have come up with various ways to penalize models for excessive complexity. In the machine learning world, this is known as "regularization," and we will describe some examples below. Economists tend to prefer simpler models for the same reason, but have not been as explicit about quantifying complexity costs.

Second, it is conventional to divide the data into separate sets for the purpose of training, testing, and validation. You use the training data to estimate a model, the validation data to choose your model, and the testing data to evaluate how well your chosen model performs. (Often validation and testing sets are combined.)

Third, if we have an explicit numeric measure of model complexity, we can view it as a parameter that can be "tuned" to produce the best out of sample predictions. The standard way to choose a good value for such a tuning parameter is to use *k-fold cross-validation.*

1. Divide the data into k roughly equal subsets (folds) and label them by $s = 1, \ldots, k$. Start with subset $s = 1$.
2. Pick a value for the tuning parameter.
3. Fit your model using the $k - 1$ subsets other than subset s.
4. Predict for subset s and measure the associated loss.
5. Stop if $s = k$, otherwise increment s by 1 and go to step 2.

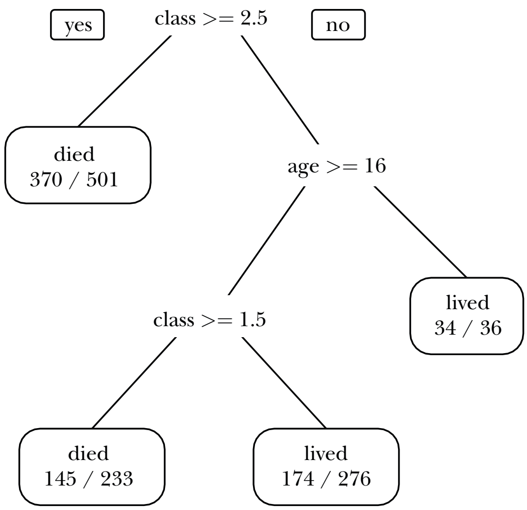
Common choices for k are 10, 5, and the sample size minus 1 ("leave one out"). After cross-validation, you end up with k values of the tuning parameter and the associated loss which you can then examine to choose an appropriate value for the tuning parameter. Even if there is no tuning parameter, it is prudent to use cross-validation to report goodness-of-fit measures since it measures out-of-sample performance, which is generally more meaningful than in-sample performance.

The test-train cycle and cross-validation are very commonly used in machine learning and, in my view, should be used much more in economics, particularly when working with large datasets. For many years, economists have reported in-sample goodness-of-fit measures using the excuse that we had small datasets. But now that larger datasets have become available, there is no reason not to use separate training and testing sets. Cross-validation also turns out to be a very useful technique, particularly when working with reasonably large data. It is also a much more realistic measure of prediction performance than measures commonly used in economics.

Classification and Regression Trees

Let us start by considering a discrete variable regression where our goal is to predict a 0–1 outcome based on some set of features (what economists would call explanatory variables or predictors). In machine learning, this is known as a

Figure 1
A Classification Tree for Survivors of the *Titanic*



Note: See text for interpretation.

classification problem. A common example would be classifying email into "spam" or "not spam" based on characteristics of the email. Economists would typically use a generalized linear model like a logit or probit for a classification problem.

A quite different way to build a classifier is to use a decision tree. Most economists are familiar with decision trees that describe a sequence of decisions that results in some outcome. A tree classifier has the same general form, but the decision at the end of the process is a choice about how to classify the observation. The goal is to construct (or "grow") a decision tree that leads to good out-of-sample predictions.

Ironically, one of the earliest papers on the automatic construction of decision trees (Morgan and Sonquist 1963) was coauthored by an economist. However, the technique did not really gain much traction until 20 years later in the work of Breiman, Friedman, Olshen, and Stone (1984). Nowadays this prediction technique is known as "classification and regression trees," or "CART."

To illustrate the use of tree models, I used the **R** package **rpart** to find a tree that predicts *Titanic* survivors using just two variables: age and class of travel.[1] The resulting tree is shown in Figure 1, and the rules depicted in the tree are shown in Table 2. The rules fit the data reasonably well, misclassifying about 30 percent of the observations in the testing set.

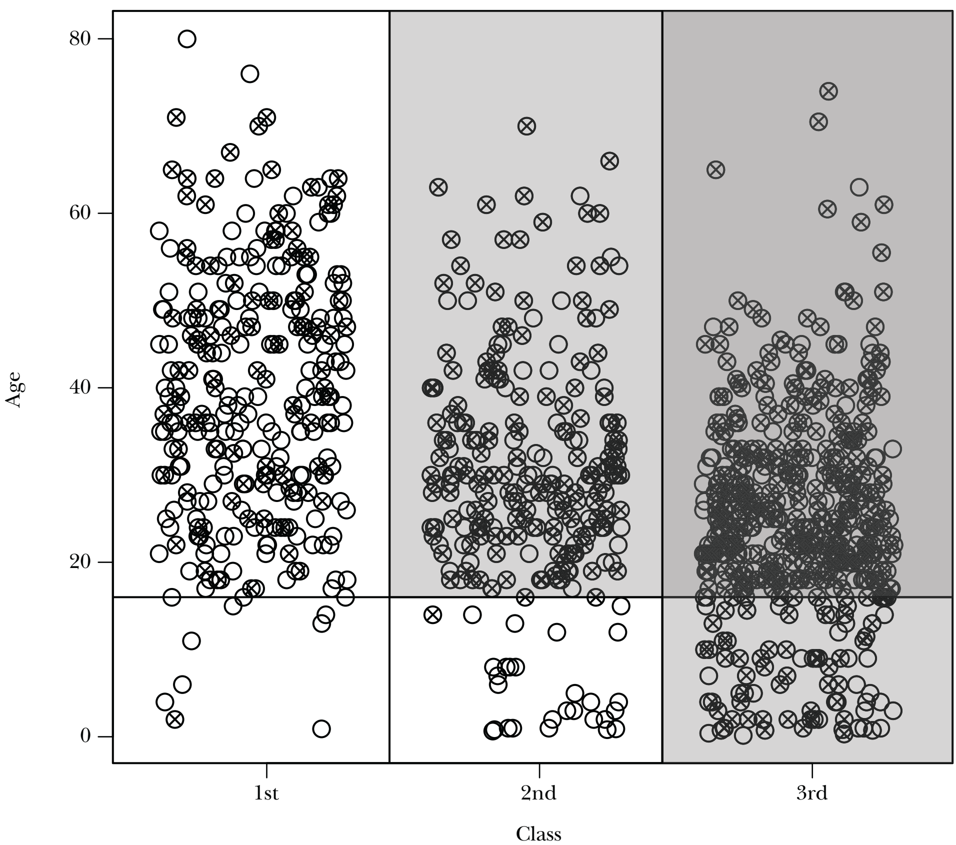
This classification can also be depicted in the "partition plot" (Figure 2), which shows how the tree divides up the space of age and class pairs into rectangular

[1] All data and code used in this paper can be found in the online Appendix available at http://e-jep.org.

Table 2
Tree Model in Rule Form

Features	*Predicted*	*Actual/Total*
Class 3	Died	370/501
Class 1–2, younger than 16	Lived	34/36
Class 2, older than 16	Died	145/233
Class 1, older than 16	Lived	174/276

Figure 2
The Simple Tree Model Predicts Death in Shaded Region
(empty circles indicate survival; circles with x*'s indicate death)*



regions. Of course, the partition plot can only be used for two variables, while a tree representation can handle an arbitrarily large number.

It turns out that there are computationally efficient ways to construct classification trees of this sort. These methods generally are restricted to binary trees (two branches

Table 3
Logistic Regression of Survival versus Age

Coefficient	*Estimate*	*Standard error*	*t value*	*p value*
Intercept	0.465	0.0350	13.291	0.000
Age	−0.002	0.001	−1.796	0.072

Note: Logistic regression relating survival (0 or 1) to age in years.

at each node). They can be used for classification with multiple outcomes ("classification trees") or with continuous dependent variables ("regression trees").

Trees tend to work well for problems where there are important nonlinearities and interactions. As an example, let us continue with the *Titanic* data and create a tree that relates survival to age. In this case, the rule generated by the tree is very simple: predict "survive" if age < 8.5 years. We can examine the same data with a logistic regression to estimate the probability of survival as a function of age, with results reported in Table 3.

The tree model suggests that age is an important predictor of survival, while the logistic model says it is barely important. This discrepancy is explained in Figure 3 where we plot survival rates by age bins. Here we see that survival rates for the youngest passengers were relatively high, and survival rates for older passengers were relatively low. For passengers between these two extremes, age didn't matter very much. So what mattered for survival is not so much age, but whether the passenger was a child or elderly. It would be difficult to discover this pattern from a logistic regression alone.[2]

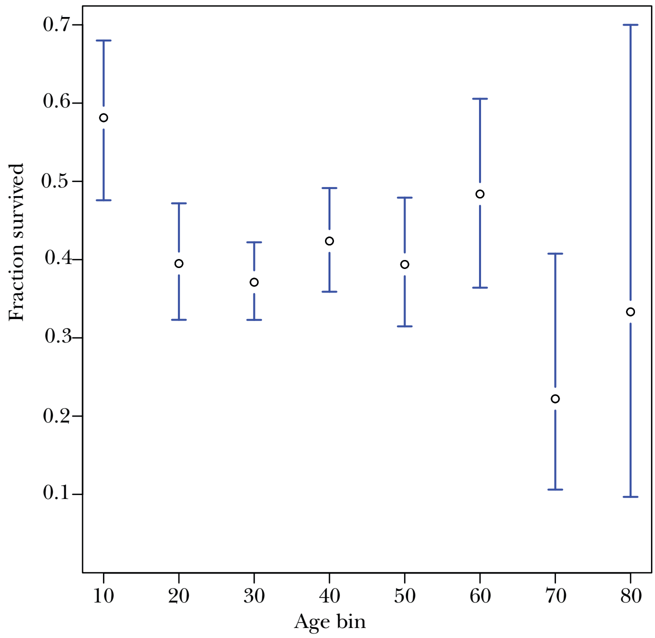
Trees also handle missing data well. Perlich, Provost, and Simonoff (2003) examined several standard datasets and found that "logistic regression is better for smaller data sets and tree induction for larger data sets." Interestingly enough, trees tend *not* to work very well if the underlying relationship really is linear, but there are hybrid models such as RuleFit (Friedman and Popescu 2005) that can incorporate both tree and linear relationships among variables. However, even if trees may not improve on predictive accuracy compared to linear models, the age example shows that they may reveal aspects of the data that are not apparent from a traditional linear modeling approach.

Pruning Trees

One problem with trees is that they tend to overfit the data. Just as a regression with n observations and n variables will give you a good fit in-sample, a tree with many branches will also fit the training data well. In either case, predictions using new data, such as the test set, could be very poor.

[2] It is true that if you *knew* that there was a nonlinearity in age, you could use age dummies in the logit model to capture this effect. However the tree formulation made this nonlinearity immediately apparent.

Figure 3
***Titanic* Survival Rates by Age Group**



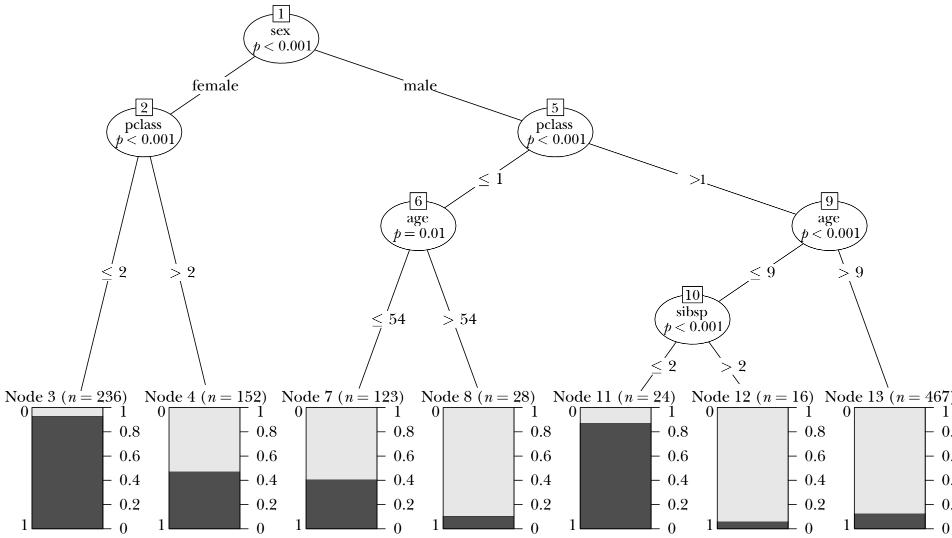
Notes: The figure shows the mean survival rates for different age groups along with confidence intervals. The age bin 10 means "10 and younger," the next age bin is "older than 10 through 20," and so on.

The most common solution to this problem is to "prune" the tree by imposing a cost for complexity. There are various measures of complexity, but a common one is the number of terminal nodes (also known as "leafs"). The cost of complexity is a tuning parameter that is chosen to provide the best out-of-sample predictions, which is typically measured using the 10-fold cross-validation procedure mentioned earlier.

A typical tree estimation session might involve dividing your data into ten folds, using nine of the folds to grow a tree with a particular complexity, and then predict on the excluded fold. Repeat the estimation with different values of the complexity parameter using other folds and choose the value of the complexity parameter that minimizes the out-of-sample classification error. (Some researchers recommend being a bit more aggressive and advocate choosing the complexity parameter that is one standard deviation lower than the loss-minimizing value.)

Of course, in practice, the computer program handles most of these details for you. In the examples in this paper, I mostly use default choices to keep things simple, but in practice these defaults will often be adjusted by the analyst. As with any other statistical procedure, skill, experience, and intuition are helpful in coming up with a good answer. Diagnostics, exploration, and experimentation are just as useful with these methods as with regression techniques.

Figure 4
A ctree for Survivors of the *Titanic*
(black bars indicate fraction of the group that survived)



Note: See text for interpretation.

There are many other approaches to creating trees, including some that are explicitly statistical in nature. For example, a "conditional inference tree," or ctree for short, chooses the structure of the tree using a sequence of hypothesis tests. The resulting trees tend to need very little pruning (Hothorn, Hornik, and Zeileis 2006). An example for the *Titanic* data is shown in Figure 4.

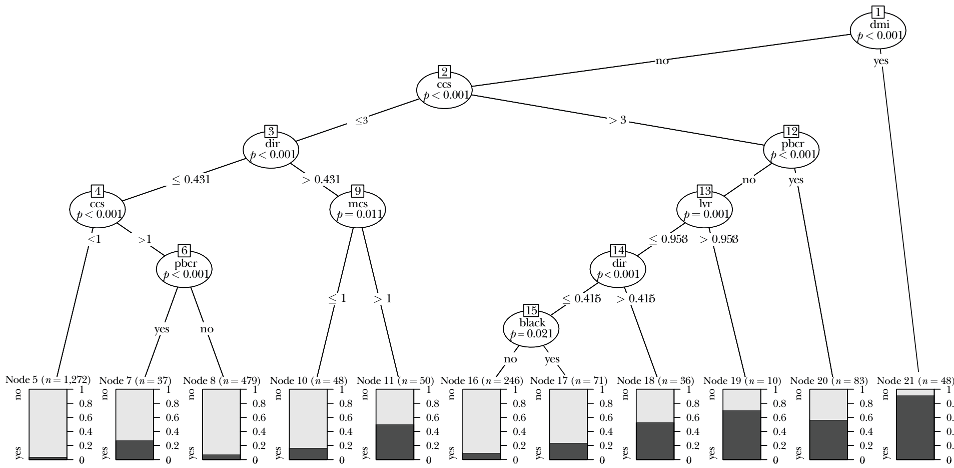
The first node divides by gender. The second node then divides by class. In the right-hand branches, the third node divides by age, and a fourth node divides by the number of siblings plus spouse aboard. The bins at the bottom of the figure show the total number of people in that leaf and a graphical depiction of their survival rate. One might summarize this tree by the following principle: "women and children first . . . particularly if they were traveling first class." This simple example again illustrates that classification trees can be helpful in summarizing relationships in data, as well as predicting outcomes.[3]

An Economic Example Using Home Mortgage Disclosure Act Data

Munnell, Tootell, Browne, and McEneaney (1996) examined mortgage lending in Boston to see if race played a significant role in determining who was approved for a mortgage. The primary econometric technique was a logistic regression where

[3] For two excellent tutorials on tree methods that use the *Titanic* data, see Stephens and Wehrley (2014).

Figure 5
Home Mortgage Disclosure Act (HMDA) Data Tree



Notes: Figure 5 shows a conditional tree estimated using the **R** package **party**. The black bars indicate the fraction of each group who were denied mortgages. The most important determinant of this is the variable "dmi," or "denied mortgage insurance." Other variables are: "dir," debt payments to total income ratio; "hir," housing expenses to income ratio; "lvr," ratio of size of loan to assessed value of property; "ccs," consumer credit score; "mcs," mortgage credit score; "pbcr," public bad credit record; "dmi," denied mortgage insurance; "self," self-employed; "single," applicant is single; "uria," 1989 Massachusetts unemployment rate applicant's industry; "condominium," unit is condominium; "black," race of applicant black; and "deny," mortgage application denied.

race was included as one of the predictors. The coefficient on race showed a statistically significant negative impact on probability of getting a mortgage for black applicants. This finding prompted considerable subsequent debate and discussion; see Ladd (1998) for an overview.

Here I examine this question using the tree-based estimators described in the previous section. The data consists of 2,380 observations of 12 predictors, one of which was race. Figure 5 shows a conditional tree estimated using the **R** package **party**.

The tree fits pretty well, misclassifying 228 of the 2,380 observations for an error rate of 9.6 percent. By comparison, a simple logistic regression does slightly better, misclassifying 225 of the 2,380 observations, leading to an error rate of 9.5 percent. As you can see in Figure 5, the most important variable is "dmi" = "denied mortgage insurance." This variable alone explains much of the variation in the data. The race variable ("black") shows up far down the tree and seems to be relatively unimportant.

One way to gauge whether a variable is important is to exclude it from the prediction and see what happens. When this is done, it turns out that the accuracy of the tree-based model doesn't change at all: exactly the same cases are misclassified. Of course, it is perfectly possible that there was racial discrimination elsewhere in the mortgage process, or that some of the variables included are highly correlated with race. But it is noteworthy that the tree model produced by standard procedures that omits race fits the observed data just as well as a model that includes race.

Boosting, Bagging, Bootstrap

There are several useful ways to improve classifier performance. Interestingly enough, some of these methods work by *adding* randomness to the data. This seems paradoxical at first, but adding randomness turns out to be a helpful way of dealing with the overfitting problem.

Bootstrap involves choosing (with replacement) a sample of size n from a dataset of size n to estimate the sampling distribution of some statistic. A variation is the "m out of n bootstrap" which draws a sample of size m from a dataset of size $n > m$.

Bagging involves averaging across models estimated with several different bootstrap samples in order to improve the performance of an estimator.

Boosting involves repeated estimation where misclassified observations are given increasing weight in each repetition. The final estimate is then a vote or an average across the repeated estimates.[4]

Econometricians are well-acquainted with the bootstrap but rarely use the other two methods. Bagging is primarily useful for nonlinear models such as trees (Friedman and Hall 2007). Boosting tends to improve predictive performance of an estimator significantly and can be used for pretty much any kind of classifier or regression model, including logits, probits, trees, and so on.

It is also possible to combine these techniques and create a "forest" of trees that can often significantly improve on single-tree methods. Here is a rough description of how such "random forests" work.

Random Forests

Random forests is a technique that uses multiple trees. A typical procedure uses the following steps.

1. Choose a bootstrap sample of the observations and start to grow a tree.
2. At each node of the tree, choose a random sample of the predictors to make the next decision. Do not prune the trees.
3. Repeat this process many times to grow a forest of trees.
4. In order to determine the classification of a new observation, have each tree make a classification and use a majority vote for the final prediction.

This method produces surprisingly good out-of-sample fits, particularly with highly nonlinear data. In fact, Howard and Bowles (2012) claim "ensembles of decision trees (often known as 'Random Forests') have been the most successful general-purpose algorithm in modern times." They go on to indicate that

[4] Boosting is often used with decision trees, where it can dramatically improve their predictive performance.

"the algorithm is very simple to understand, and is fast and easy to apply." See also Caruana and Niculescu-Mitzil (2006) who compare several different machine learning algorithms and find that ensembles of trees perform quite well. There are a number of variations and extensions of the basic "ensemble of trees" model such as Friedman's "Stochastic Gradient Boosting" (Friedman 2002).

One defect of random forests is that they are a bit of a black box—they don't offer simple summaries of relationships in the data. As we have seen earlier, a single tree can offer some insight about how predictors interact. But a forest of a thousand trees cannot be easily interpreted. However, random forests can determine which variables are "important" in predictions in the sense of contributing the biggest improvements in prediction accuracy.

Note that random forests involves quite a bit of randomization; if you want to try them out on some data, I strongly suggest choosing a particular seed for the random number generator so that your results can be reproduced. (See the online supplement for examples.)

I ran the random forest method on the HMDA data and found that it misclassified 223 of the 2,380 cases, a small improvement over the logit and the ctree. I also used the importance option in random forests to see how the predictors compared. It turned out that "dmi" was the most important predictor and race was second from the bottom, which is consistent with the ctree analysis.

Variable Selection

Let us return to the familiar world of linear regression and consider the problem of variable selection. There are many such methods available, including stepwise regression, principal component regression, partial least squares, Akaike information criterion (AIC) and Bayesian information criterion (BIC) complexity measures, and so on. Castle, Qin, and Reed (2009) describe and compare 21 different methods.

LASSO and Friends

Here we consider a class of estimators that involves penalized regression. Consider a standard multivariate regression model where we predict y_t as a linear function of a constant, b_0, and P predictor variables. We suppose that we have standardized all the (nonconstant) predictors so they have mean zero and variance one.

Consider choosing the coefficients $(b_1, \ldots, b_P)$ for these predictor variables by minimizing the sum of squared residuals plus a penalty term of the form

$$\lambda \sum_{p=1}^{P} \left[(1 - \alpha) |b_p| + \alpha |b_p|^2 \right].$$

This estimation method is called *elastic net regression*; it contains three other methods as special cases. If there is no penalty term $(\lambda = 0)$, this is *ordinary least squares*. If $\alpha = 1$, so that there is only the quadratic constraint, this is *ridge regression*.

If $\alpha = 0$, this is called the *LASSO*, an acronym for "least absolute shrinkage and selection operator."

These penalized regressions are classic examples of regularization. In this case, the complexity is the number and size of predictors in the model. All of these methods tend to shrink the least squares regression coefficients towards zero. The LASSO and elastic net typically produces regressions where some of the variables are set to be exactly zero. Hence this is a relatively straightforward way to do variable selection.

It turns out that these estimators can be computed quite efficiently, so doing variable selection on reasonably large problems is computationally feasible. They also seem to provide good predictions in practice.

Spike-and-Slab Regression

Another approach to variable selection that is novel to most economists is spike-and-slab regression, a Bayesian technique. Suppose that you have P possible predictors in some linear model. Let γ be a vector of length P composed of zeros and ones that indicate whether or not a particular variable is included in the regression.

We start with a Bernoulli prior distribution on γ; for example, initially we might think that all variables have an equally likely chance of being in the regression. Conditional on a variable being in the regression, we specify a prior distribution for the regression coefficient associated with that variable. For example, we might use a Normal prior with mean 0 and a large variance. These two priors are the source of the method's name: the "spike'" is the probability of a coefficient being nonzero; the "slab" is the (diffuse) prior describing the values that the coefficient can take on.

Now we take a draw of γ from its prior distribution, which will just be a list of variables in the regression. Conditional on this list of included variables, we take a draw from the prior distribution for the coefficients. We combine these two draws with the likelihood in the usual way, which gives us a draw from posterior distribution on both probability of inclusion and the coefficients. We repeat this process thousands of times using a Markov Chain Monte Carlo (MCMC) technique which gives us a table summarizing the posterior distribution for γ (indicating variable inclusion), β (the coefficients), and the associated prediction of y. We can summarize this table in a variety of ways. For example, we can compute the average value of γ_p which shows the posterior probability that the variable p is included in the regressions.

An Economic Example: Growth Regressions

We illustrate these different methods of variable selection using data from Sala-i-Martín (1997). This exercise involved examining a dataset of 72 counties and 42 variables in order to see which variables appeared to be important predictors of economic growth. Sala-i-Martín (1997) computed at all possible subsets of regressors of manageable size and used the results to construct an importance measure he called CDF(0). Ley and Steel (2009) investigated the same question using Bayesian

Table 4
Comparing Variable Selection Algorithms: Which Variables Appeared as Important Predictors of Economic Growth?

Predictor	*Bayesian model averaging*	*CDF(0)*	*LASSO*	*Spike-and-Slab*
GDP level 1960	1.000	1.000	-	0.9992
Fraction Confucian	0.995	1.000	2	0.9730
Life expectancy	0.946	0.942	-	0.9610
Equipment investment	0.757	0.997	1	0.9532
Sub-Saharan dummy	0.656	1.000	7	0.5834
Fraction Muslim	0.656	1.000	8	0.6590
Rule of law	0.516	1.000	-	0.4532
Open economy	0.502	1.000	6	0.5736
Degree of capitalism	0.471	0.987	9	0.4230
Fraction Protestant	0.461	0.966	5	0.3798

Source: The table is based on that in Ley and Steel (2009); the data analyzed is from Sala-i-Martín (1997).
Notes: We illustrate different methods of variable selection. This exercise involved examining a dataset of 72 counties and 42 variables in order to see which variables appeared to be important predictors of economic growth. The table shows ten predictors that were chosen by Sala-i-Martín (1997) using a CDF(0) measure defined in the 1997 paper; Ley and Steel (2009) using Bayesian model averaging, LASSO, and spike-and-slab regressions. Metrics used are not strictly comparable across the various models. The "Bayesian model averaging" and "Spike-and-Slab" columns are posterior probabilities of inclusion; the "LASSO" column just shows the ordinal importance of the variable or a dash indicating that it was not included in the chosen model; and the CDF(0) measure is defined in Sala-i-Martín (1997).

model averaging, a technique related to, but not identical with, spike-and-slab. Hendry and Krolzig (2004) examined an iterative significance test selection method.

Table 4 shows ten predictors that were chosen by Sala-i-Martín (1997) using his two million regressions, Ley and Steel (2009) using Bayesian model averaging, LASSO, and spike-and-slab. The table is based on that in Ley and Steel (2009) but metrics used are not strictly comparable across the various models. The "Bayesian model averaging" and "spike-slab" columns show posterior probabilities of inclusion; the "LASSO" column just shows the ordinal importance of the variable or a dash indicating that it was not included in the chosen model; and the CDF(0) measure is defined in Sala-i-Martín (1997).

The LASSO and the Bayesian techniques are very computationally efficient and would likely be preferred to exhaustive search. All four of these variable selection methods give similar results for the first four or five variables, after which they diverge. In this particular case, the dataset appears to be too small to resolve the question of what is "important" for economic growth.

Variable Selection in Time Series Applications

The machine learning techniques described up until now are generally applied to cross-sectional data where independently distributed data is a plausible assumption. However, there are also techniques that work with time series. Here we describe an estimation method that we call Bayesian Structural Time Series (BSTS) that seems to work well for variable selection problems in time series applications.

Our research in this area was motivated by Google Trends data, which provides an index of the volume of Google queries on specific terms. One might expect that queries on "file for unemployment" might be predictive of the actual rate of filings for initial claims, or that queries on "Orlando vacation" might be predictive of actual visits to Orlando. Indeed, in Choi and Varian (2009, 2012), Goel, Hofman, Lahaie, Pennock, and Watts (2010), Carrière-Swallow and Labbé (2011), McLaren and Shanbhoge (2011), Artola and Galan (2012), Hellerstein and Middeldorp (2012), and other papers, many researchers have shown that Google queries do have significant short-term predictive power for various economic metrics.

The challenge is that there are billions of queries so it is hard to determine exactly which queries are the most predictive for a particular purpose. Google Trends classifies the queries into categories, which helps a little, but even then we have hundreds of categories as possible predictors so that overfitting and spurious correlation are a serious concern. Bayesian Structural Time Series is designed to address these issues. We offer a very brief description here; more details are available in Scott and Varian (2013a, 2013b).

Consider a classic time series model with *constant* level, linear time trend, and regressor components:

$$yt = \mu + bt + \beta xt + et.$$

The "local linear trend" is a stochastic generalization of this model where the level and time trend can vary through time.

Observation: $y_t = \mu_t + z_t + e_{1t} = \text{level} + \text{regression}$

State variable 1: $\mu_t = \mu_{t-1} + b_{t-1} + e_{2t} = \text{random walk} + \text{trend}$

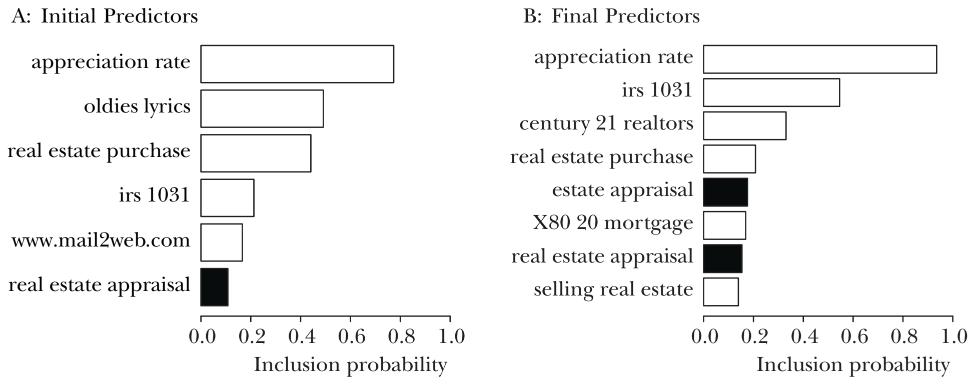
State variable 2: $z_t = \beta x_t = \text{regression}$

State variable 3: $b_t = b_{t-1} + e_{3t} = \text{random walk for trend}$

It is easy to add an additional state variable for seasonality if that is appropriate. The parameters to estimate are the regression coefficients β and the variances of (e_{it}) for $i = 1, \ldots 3$. We can then use these estimates to construct the optimal forecast based on techniques drawn from the literature on Kalman filters.

For the regression, we use the spike-and-slab variable choice mechanism described above. A draw from the posterior distribution now involves a draw of variances of (e_{1t}, e_{2t}, e_{3t}) a draw of the vector γ that indicates which variables are in the regression, and a draw of the regression coefficients β for the included variables. The draws of μ_t, b_t, and β can be used to construct estimates of y_t and forecasts for y_{t+1}. We end up with an (estimated) posterior distribution for each parameter of

Figure 6
An Example Using Bayesian Structural Time Series (BSTS)
(finding Google queries that are predictors of new home sales)



Source: Author using HSN1FNSA data from the St. Louis Federal Reserve Economic Data.
Notes: Consider the nonseasonally adjusted data for new homes sold in the United States, which is (HSN1FNSA) from the St. Louis Federal Reserve Economic Data. This time series can be submitted to Google Correlate, which then returns the 100 queries that are the most highly correlated with the series. We feed that data into the BSTS system, which identifies the predictors with the largest posterior probabilities of appearing in the housing regression; these are shown in Figure 6A. In these figures, black bars indicate a negative relationship, and white bars indicate a positive relationship. Two predictors, "oldies lyrics" and "www.mail2web" appear to be spurious so we remove them and re-estimate, yielding the results in Figure 6B.

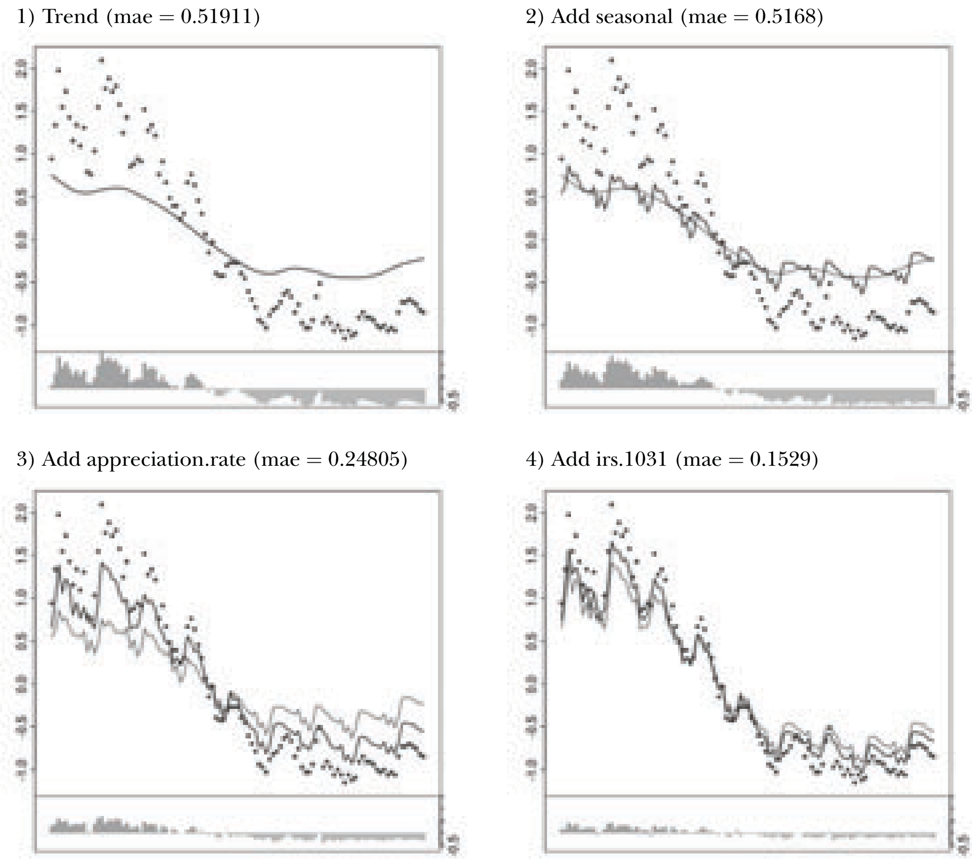
interest. If we seek a point prediction, we can average over these draws, which is essentially a form of Bayesian model averaging.

As an example, consider the nonseasonally adjusted data for new homes sold in the United States, which is (HSN1FNSA) from the St. Louis Federal Reserve Economic Data. This time series can be submitted to Google Correlate, which then returns the 100 queries that are the most highly correlated with the series. We feed that data into the BSTS system, which identifies the predictors with the largest posterior probabilities of appearing in the housing regression; these are shown in Figure 6A.In these figures, black bars indicate a negative relationship and white bars indicate a positive relationship. Two predictors, "oldies lyrics" and "www.mail2web" appear to be spurious so we remove them and re-estimate, yielding the results in Figure 6B.

The fit is shown in Figure 7, which shows the incremental contribution of the trend, seasonal, and two of the regressors. Even with only two predictors, queries on "appreciation rate" and queries on "irs 1031," we get a pretty good fit.[5]

[5] IRS section 1031 has to do with deferring capital gains on certain sorts of property exchange.

Figure 7
Fit for the Housing Regression: Incremental Contribution of Trend, Seasonal, and Two Regressors



Source: Author using (HSN1FNSA) data from the St. Louis Federal Reserve.
Notes: The plots show the impact of the trend, seasonal, and a few individual regressors. Data has been standardized to have mean zero and variance 1. The residuals are shown on the bottom. The abbreviation "mae" stands for "mean absolute error."

Econometrics and Machine Learning

There are a number of areas where there would be opportunities for fruitful collaboration between econometrics and machine learning. I mentioned above that most machine learning uses independent and identically distributed data. However, the Bayesian Structural Time Series model shows that some of these techniques can be adopted for time series models. It is also possible to use machine learning techniques to look at panel data, and there has been some work in this direction.

However, the most important area for collaboration involves causal inference. Econometricians have developed several tools for causal inference such as instrumental variables, regression discontinuity, difference-in-differences, and various forms of natural and designed experiments (Angrist and Krueger 2001). Machine learning work has, for the most part, dealt with pure prediction. In a way, this is ironic, since theoretical computer scientists, such as Pearl (2009a, b) have made significant contributions to causal modeling. However, it appears that these theoretical advances have not as yet been incorporated into machine learning practice to a significant degree.

Causality and Prediction

As economists know well, there is a big difference between correlation and causation. A classic example: there are often more police in precincts with high crime, but that does not imply that increasing the number of police in a precinct would increase crime.

The machine learning models we have described so far have been entirely about prediction. If our data were generated by policymakers who assigned police to areas with high crime, then the observed relationship between police and crime rates could be highly predictive for the *historical* data but not useful in predicting the causal impact of explicitly *assigning* additional police to a precinct.

To enlarge on this point, let us consider an experiment (natural or designed) that attempts to estimate the impact of some policy, such as adding police to precincts. There are two critical questions.

1) How will police be assigned to precincts in both the experiment and the policy implementation? Possible assignment rules could be 1) random, 2) based on perceived need, 3) based on cost of providing service, 4) based on resident requests, 5) based on a formula or set of rules, 6) based on asking for volunteers, and so on. Ideally the assignment procedure in the experiment will be similar to that used in the policy. Developing accurate predictions about which precincts will receive additional police under the proposed policy based on the experimental data can clearly be helpful in predicting the expected impact of the policy.

2) What will be the impact of these additional police in both the experiment and the policy? As Rubin (1974) and many subsequent authors have emphasized, when we want to estimate the *causal* impact of some treatment we need to compare the outcome with the intervention to what *would have happened* without the intervention. But this counterfactual cannot be observed, so it must be predicted by some model. The better predictive model you have for the counterfactual, the better you will be able to estimate the causal effect, a rule that is true for both pure experiments and natural experiments.

So even though a predictive model will not necessarily allow one to conclude anything about causality by itself, such models may help in estimating the causal impact of an intervention when it occurs.

To state this in a slightly more formal way, consider the identity from Angrist and Pischke (2009, p. 11):

$$\text{observed difference in outcome} = \text{average treatment effect on the treated} \\ + \text{ selection bias.}$$

If you want to model the average treatment effect as a function of other variables, you will usually need to model both the observed difference in outcome and the selection bias. The better your predictive model for those components, the better your estimate of the average treatment effect will be. Of course, if you have a true randomized treatment–control experiment, selection bias goes away and those treated are an unbiased random sample of the population.

To illustrate these points, let us consider the thorny problem of estimating the causal effect of advertising on sales (Lewis and Rao 2013). The difficulty is that there are many confounding variables, such as seasonality or weather, that cause both increased ad exposures and increased purchases by consumers. For example, consider the (probably apocryphal) story about an advertising manager who was asked why he thought his ads were effective. "Look at this chart," he said. "Every December I increase my ad spend and, sure enough, purchases go up." Of course, in this case, seasonality can be included in the model. However, generally there will be other confounding variables that affect both exposure to ads and the propensity of purchase, which make causal interpretations of observed relationships problematic.

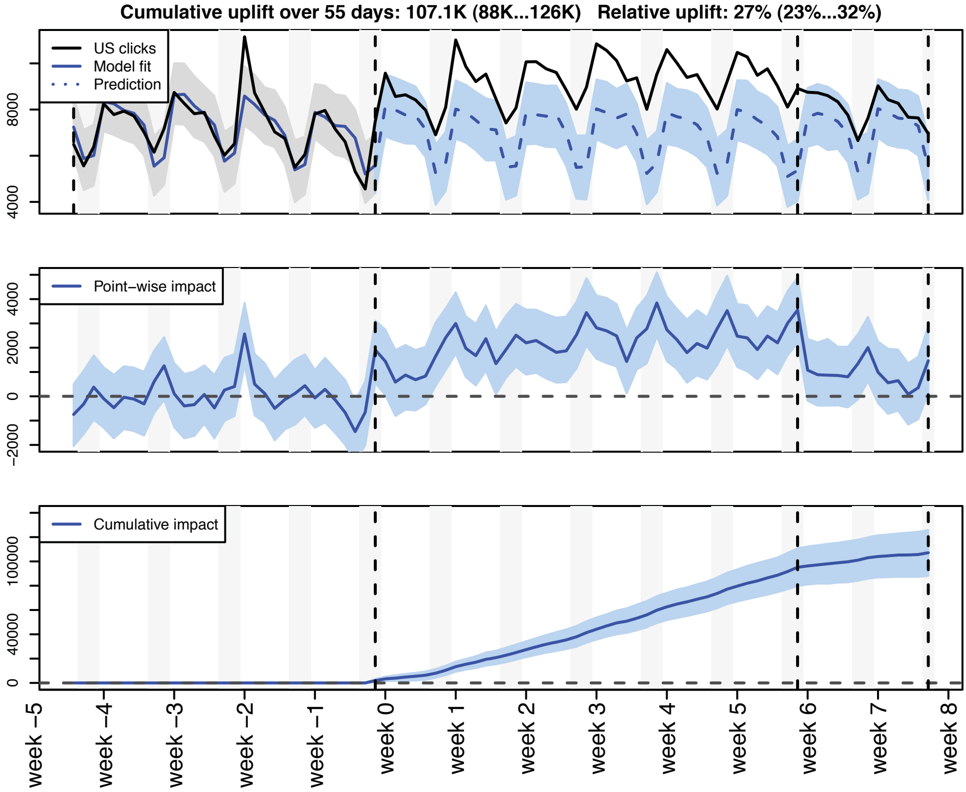
The ideal way to estimate advertising effectiveness is, of course, to run a controlled experiment. In this case the control group provides an estimate of the counterfactual: what would have happened without ad exposures. But this ideal approach can be quite expensive, so it is worth looking for alternative ways to predict the counterfactual. One way to do this is to use the Bayesian Structural Time Series (BSTS) method described earlier.

Suppose a given company wants to determine the impact of an advertising campaign on visits to its website. It first uses BSTS (or some other technique) to build a model predicting the time series of visits as a function of its past history, seasonal effects, and other possible predictors such as Google queries on its company name, its competitors' names, or products that it produces. Since there are many possible choices for predictors, it is important to use some variable selection mechanism such as those described earlier.

It next runs an ad campaign for a few weeks and records visits during this period. Finally, it makes a forecast of what visits *would have been* in the absence of the ad campaign using the model developed in the first stage. Comparing the actual visits to the counterfactual visits gives us an estimate of the causal effect of advertising.

Figure 8, shows the outcome of such a procedure. It is based on the approach proposed in Brodersen, Gallusser, Koehler, Remy, and Scott (2013), but the covariates are chosen automatically from Google Trends categories using Bayesian Structural Time Series (BSTS). Panel A shows the actual visits and the prediction

Figure 8
Actual and Predicted Website Visits



Source: This example is based on the approach proposed in Brodersen, Gallusser, Koehler, Remy, and Scott (2013), but the covariates are chosen automatically from Google Trends categories using Bayesian Structural Time Series (BSTS).
Notes: Suppose a given company wants to determine the impact of an advertising campaign on its website visits. Panel A shows the actual visits and the prediction of what the visits would have been without the campaign based on the BSTS forecasting model. Panel B shows the difference between actual and predicted visits, and Panel C shows the cumulative difference.

of what the visits would have been without the campaign based on the BSTS forecasting model. Panel B shows the difference between actual and predicted visits, and Panel C shows the cumulative difference. It is clear from this figure that there was a significant causal impact of advertising, which can then be compared to the cost of the advertising to evaluate the campaign.

This procedure does not use a control group in the conventional sense. Rather it uses a general time series model based on trend extrapolation, seasonal effects, and relevant covariates to forecast what would have happened without the ad campaign.

A good predictive model can be better than a randomly chosen control group, which is usually thought to be the gold standard. To see this, suppose that you run an ad campaign in 100 cities and retain 100 cities as a control. After the experiment is over, you discover the weather was dramatically different across the cities in the study. Should you add weather as a predictor of the counterfactual? Of course! If weather affects sales (which it does), then you will get a more accurate prediction of the counterfactual and thus a better estimate of the causal effect of advertising.

Model Uncertainty

An important insight from machine learning is that averaging over many small models tends to give better out-of-sample prediction than choosing a single model.

In 2006, Netflix offered a million dollar prize to researchers who could provide the largest improvement to their existing movie recommendation system. The winning submission involved a "complex blending of no fewer than 800 models," though they also point out that "predictions of good quality can usually be obtained by combining a small number of judiciously chosen methods" (Feuerverger, He, and Khatri 2012). It also turned out that a blend of the best- and second-best submissions outperformed either of them.

Ironically, it was recognized many years ago that averages of macroeconomic model forecasts outperformed individual models, but somehow this idea was rarely exploited in traditional econometrics. The exception is the literature on Bayesian model averaging, which has seen a steady flow of work; see Steel (2011) for a survey.

However, I think that model uncertainty has crept into applied econometrics through the back door. Many papers in applied econometrics present regression results in a table with several different specifications: which variables are included in the controls, which variables are used as instruments, and so on. The goal is usually to show that the estimate of some interesting parameter is not very sensitive to the exact specification used.

One way to think about it is that these tables illustrate a simple form of model uncertainty: how an estimated parameter varies as different models are used. In these papers, the authors tend to examine only a few representative specifications, but there is no reason why they couldn't examine many more if the data were available.

In this period of "big data," it seems strange to focus on *sampling uncertainty,* which tends to be small with large datasets, while completely ignoring *model uncertainty,* which may be quite large. One way to address this is to be explicit about examining how parameter estimates vary with respect to choices of control variables and instruments.

Summary and Further Reading

Since computers are now involved in many economic transactions, big data will only get bigger. Data manipulation tools and techniques developed for small datasets will become increasingly inadequate to deal with new problems. Researchers in machine learning have developed ways to deal with large datasets and economists interested in dealing with such data would be well advised to invest in learning these techniques.

I have already mentioned Hastie, Tibshirani, and Friedman (2009), who provide detailed descriptions of all the methods discussed here but at a relatively advanced level. James, Witten, Hastie, and Tibshirani (2013) describe many of the same topics at an undergraduate-level, along with **R** code and many examples. (There are several economic examples in the book where the tension between predictive modeling and causal inference is apparent.) Murphy (2012) examines machine learning from a Bayesian point of view.

Venables and Ripley (2002) offer good discussions of these topics with emphasis on applied examples. Leek (2013) presents a number of YouTube videos with gentle and accessible introductions to several tools of data analysis. Howe (2013) provides a somewhat more advanced introduction to data science that also includes discussions of SQL and NoSQL databases. Wu and Kumar (2009) give detailed descriptions and examples of the major algorithms in data mining, while Williams (2011) provides a unified toolkit. Domingos (2012) summarizes some important lessons including "pitfalls to avoid, important issues to focus on and answers to common questions."

■ *Thanks to Jeffrey Oldham, Tom Zhang, Rob On, Pierre Grinspan, Jerry Friedman, Art Owen, Steve Scott, Bo Cowgill, Brock Noland, Daniel Stonehill, Robert Snedegar, Gary King, Fabien Curto-Millet, and the editors of this journal for helpful comments on earlier versions of this paper. The author works for Google, and Google had the right to review this paper before publication.*

References

Angrist, Joshua D., and Alan B. Krueger. 2001. "Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments." *Journal of Economic Perspectives* 5(4): 69–85.

Angrist, Joshua D., and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Artola, Concha, and Enrique Galan. 2012. "Tracking the Future on the Web: Construction of Leading Indicators Using Internet Searches." Documentos Ocasionales 1203T, Bank of Spain. http://www.bde.es/webbde/SES/Secciones/Publicaciones/PublicacionesSeriadas/DocumentosOcasionales/12/Fich/do1203e.pdf.

Breiman, Leo, Jerome H. Friedman, R. A. Olshen, and Charles J. Stone. 1984. *Classification and Regression Trees*. Wadsworth and Brooks/Cole, Monterey.

Brodersen, Kay H., Fabian Gallusser, Jim Koehler, Nicolas Remy, and Steven L. Scott. 2013. "Inferring Causal Impact Using Bayesian Structural Time-Series Models." http://research.google.com/pubs/pub41854.html.

Carrière-Swallow, Yan, and Felipe Labbé. 2011. "Nowcasting with Google Trends in an Emerging Market." *Journal of Forecasting* 32(4): 289–98.

Caruana, Rich, and Alexandru Niculescu-Mizil. 2006. "An Empirical Comparison of Supervised Learning Algorithms." In *Proceedings of the 23rd*

International Conference on Machine Learning, Pittsburgh, PA. Available at: http://www.autonlab.org/icml2006/technical/accepted.html.

Castle, Jennifer L., Xiaochuan Qin, and W. Robert Reed. 2009. "How to Pick the Best Regression Equation: A Review and Comparison of Model Selection Algorithms." Working Paper 13/2009, Department of Economics, University of Canterbury. http://www.econ.canterbury.ac.nz/RePEc/cbt/econwp/0913.pdf.

Choi, Hyunyoung, and Hal Varian. 2009. "Predicting the Present with Google Trends." http://google.com/googleblogs/pdfs/google_predicting_the_present.pdf.

Choi, Hyunyoung, and Hal Varian. 2012. "Predicting the Present with Google Trends." *Economic Record* 88(1): 2–9.

Domingos, Pedro. 2012. "A Few Useful Things to Know about Machine Learning." *Communications of the ACM* 55(10): 78–87.

Einav, Liran, and Jonathan D. Levin. 2013. "The Data Revolution and Economic Analysis." Technical report, NBER Innovation Policy and the Economy Conference, 2013. NBER Working Paper 19035.

Feuerverger, Andrey, Yu He, and Shashi Khatri. 2012. "Statistical Significance of the Netflix Challenge." *Statistical Science* 27(2): 202–231.

Friedman, Jerome. 2002. "Stochastic Gradient Boosting." *Computational Statistics & Data Analysis* 38(4): 367–78.

Friedman, Jerome, and Peter Hall. 2007. "On Bagging and Nonlinear Estimation." *Journal of Statistical Planning and Inference* 137(3): 669–83.

Friedman, Jerome H., and Bogdan E. Popescu. 2005. "Predictive Learning via Rule Ensembles." Technical report, Stanford University. http://www-stat.stanford.edu/~jhf/ftp/RuleFit.pdf

Goel, Sharad, Jake M. Hofman, Sébastien Lahaie, David M. Pennock, and Duncan J. Watts. 2010. "Predicting Consumer Behavior with Web Search." *PNAS* 107(41).

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd edition. Springer-Verlag.

Hellerstein, Rebecca, and Menno Middeldorp. 2012. "Forecasting with Internet Search Data." *Liberty Street Economics* Blog of the Federal Reserve Bank of New York, January 4. http://libertystreeteconomics.newyorkfed.org/2012/01/forecasting-with-internet-search-data.html.

Hendry, David F., and Hans-Martin Krolzig. 2004. "We Ran One Regression." *Oxford Bulletin of Economics and Statistics* 66(5): 799–810.

Hothorn, Torsten, Kurt Hornik, and Achim Zeileis. 2006. "Unbiased Recursive Partitioning: A Conditional Inference Framework." *Journal of Computational and Graphical Statistics* 15(3): 651–74.

Howard, Jeremy, and Mike Bowles. 2012. "The Two Most Important Algorithms in Predictive Modeling Today." Strata Conference presentation, February 28. http://strataconf.com/strata2012/public/schedule/detail/22658.

Howe, Bill. 2013. Introduction to Data Science. A course from the University of Washington. https://class.coursera.org/datasci-001/lecture/index.

James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2013. *An Introduction to Statistical Learning: With Applications in R*. New York: Springer.

Ladd, Helen F. 1998. "Evidence on Discrimination in Mortgage Lending." *Journal of Economic Perspectives* 12(2): 41–62.

Leek, Jeff. 2013. Data Analysis. Videos from the course. http://blog.revolutionanalytics.com/2013/04/coursera-data-analysis-course-videos.html.

Lewis, Randall A., and Justin M. Rao. 2013. "On the Near Impossibility of Measuring the Returns to Advertising." Unpublished paper, Google, Inc. and Microsoft Research. http://justinmrao.com/lewis_rao_nearimpossibility.pdf.

Ley, Eduardo, and Mark F. J. Steel. 2009. "On the Effect of Prior Assumptions in Bayesian Model Averaging with Applications to Growth Regression." *Journal of Applied Econometrics* 24(4): 651–74.

McLaren, Nick, and Rachana Shanbhoge. 2011. "Using Internet Search Data as Economic Indicators." *Bank of England Quarterly Bulletin* 51(2): 134–40.

Morgan, James N., and John A. Sonquist. 1963. "Problems in the Analysis of Survey Data, and a Proposal." *Journal of the American Statistical Association* 58(302): 415–34.

Munnell, Alicia H., Geoffrey M. B. Tootell, Lynne E. Browne, and James McEneaney. 1996. "Mortgage Lending in Boston: Interpreting HMDA Data." *American Economic Review* 86(1): 25–53.

Murphy, Kevin P. 2012. *Machine Learning: A Probabilistic Perspective.* MIT Press.

Pearl, Judea. 2009a. *Causality: Models, Reasoning, and Inference*, 2nd edition. Cambridge University Press.

Pearl, Judea. 2009b. "Causal Inference in Statistics: An Overview." *Statistics Surveys* 3: 96–146.

Perlich, Claudia, Foster Provost, and Jeffrey S. Simonoff. 2003. "Tree Induction vs. Logistic Regression: A Learning-Curve Analysis." *Journal of Machine Learning Research* 4: 211–55.

Rubin, Donald B. 1974. "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology* 66(5): 689–701.

Sala-i-Martín, Xavier. 1997. "I Just Ran Two Million Regressions." *American Economic Review* 87(2): 178–83.

Scott, Steve, and Hal Varian. 2013a. "Bayesian Variable Selection for Nowcasting Economic Time Series." NBER Working Paper 19567.

Scott, Steve, and Hal Varian. 2013b. "Predicting the Present with Bayesian Structural Time Series." NBER Working Paper 19567.

Steel, Mark F. J. 2011. "Bayesian Model Averaging and Forecasting." *Bulletin of E.U. and U.S. Inflation and Macroeconomic Analysis*, 200: 30–41.

Stephens, Revor, and Curt Wehrley. 2014. "Getting Started with R." *Kaggle*, September 28. https://www.kaggle.com/c/titanic-gettingStarted/details/new-getting-started-with-r.

Sullivan, Danny. 2012. "Google: 100 Billion Searches per Month, Search to Integrate Gmail, Launching Enhanced Search App for iOS." Search Engine Land. http://searchengineland.com/google-search-press-129925.

Venables, W. N., and B. D. Ripley. 2002. *Modern Applied Statistics with S,* 4th edition. New York: Springer.

Williams, Graham. 2011. *Data Mining with Rattle and R*. New York: Springer.

Wu, Xindong, and Vipin Kumar, eds. 2009. *The Top Ten Algorithms in Data Mining*. CRC Press.

This article has been cited by:

1. Shengquan Wang, Yuan Gao, Mingjin Luo. 2023. What drives income inequality? Bayesian evidence from the emerging market economies. *Journal of Applied Economics* **26**:1. . [Crossref]
2. Nhung Nghiem, June Atkinson, Binh P. Nguyen, An Tran-Duy, Nick Wilson. 2023. Predicting high health-cost users among people with cardiovascular disease using machine learning and nationwide linked social administrative datasets. *Health Economics Review* **13**:1. . [Crossref]
3. Hyoung Sun Yoo, Ye Lim Jung, Seung-Pyo Jun. 2023. Prediction of SMEs' R&D performances by machine learning for project selection. *Scientific Reports* **13**:1. . [Crossref]
4. Pablo de Pedraza, Ian Vollbracht. 2023. General theory of data, artificial intelligence and governance. *Humanities and Social Sciences Communications* **10**:1. . [Crossref]
5. Haodong Qi, Tuba Bircan. 2023. Can Google Trends predict asylum-seekers' destination choices?. *EPJ Data Science* **12**:1. . [Crossref]
6. Mikhail Stolbov, Maria Shchepeleva. 2023. Does one size fit all? Comparing the determinants of the FinTech market segments expansion. *The Journal of Finance and Data Science* **9**, 100095. [Crossref]
7. Jinsong Li, Kenji Takeuchi. 2023. Do municipal mergers reduce the cost of waste management? Evidence from Japan. *Regional Science and Urban Economics* **103**, 103939. [Crossref]
8. Roberto Nai, Rosa Meo, Gabriele Morina, Paolo Pasteris. 2023. Public tenders, complaints, machine learning and recommender systems: a case study in public administration. *Computer Law & Security Review* **51**, 105887. [Crossref]
9. Kentaro Kawasaki. 2023. Impact of Climate Change on Crop Pests and Diseases: Ensemble Modeling of Time-Varying Weather Effects. *Journal of the Association of Environmental and Resource Economists* **10**:6, 1515-1543. [Crossref]
10. Leonardo Madio, Francesco Principe. 2023. Who supports liberal policies? A tale of two referendums in Italy. *Economics Letters* **232**, 111338. [Crossref]
11. Benyan Tan, Ziqi Gan, Yan Wu. 2023. The measurement and early warning of daily financial stability index based on XGBoost and SHAP: Evidence from China. *Expert Systems with Applications* **227**, 120375. [Crossref]
12. Chengyan Gu. 2023. Market segmentation and dynamic price discrimination in the U.S. airline industry. *Journal of Revenue and Pricing Management* **22**:5, 338-361. [Crossref]
13. Christoph Basten, Mike Mariathasan. 2023. Interest rate pass-through and bank risk-taking under negative-rate policies with tiered remuneration of central bank reserves. *Journal of Financial Stability* **68**, 101160. [Crossref]
14. Paolo Brunori, Paul Hufe, Daniel Mahler. 2023. The roots of inequality: estimating inequality of opportunity from regression trees and forests *. *The Scandinavian Journal of Economics* **125**:4, 900-932. [Crossref]
15. Dien Giau Bui, De-Rong Kong, Chih-Yung Lin, Tse-Chun Lin. 2023. Momentum in machine learning: Evidence from the Taiwan stock market. *Pacific-Basin Finance Journal* **73**, 102178. [Crossref]
16. Hiwot Mesfin, Francesco Cecchi, Eleonora Nillesen, Nyasha Tirivayi. 2023. Overconfidence, Trust, and Information-Seeking among Smallholder Farmers: Experimental Evidence from Ethiopia. *Economic Development and Cultural Change* **72**:1, 79-122. [Crossref]
17. Matteo Farnè, Angelos Vouldis. 2023. ROBOUT: a conditional outlier detection methodology for high-dimensional data. *Statistical Papers* **26**. . [Crossref]
18. Abeer M. Abdelhalim. 2023. How management accounting practices integrate with big data analytics and its impact on corporate sustainability. *Journal of Financial Reporting and Accounting* **15**. . [Crossref]

19. Ihsaan Bassier, Joshua Budlender, Rocco Zizzamia, Ronak Jain. 2023. The labour market and poverty impacts of COVID-19 in South Africa. *South African Journal of Economics* **14**. . [Crossref]
20. Mehmet Güney Celbiş. 2023. Unemployment in Rural Europe: A Machine Learning Perspective. *Applied Spatial Analysis and Policy* **16**:3, 1071-1095. [Crossref]
21. Martin Magris, Mostafa Shabani, Alexandros Iosifidis. 2023. Bayesian bilinear neural network for predicting the mid-price dynamics in limit-order book markets. *Journal of Forecasting* **42**:6, 1407-1428. [Crossref]
22. Diep Hoang Phan. 2023. Lights and GDP relationship: What does the computer tell us?. *Empirical Economics* **65**:3, 1215-1252. [Crossref]
23. Ming Gao, Xingyu Chen, Yiyin Xu, Tianyu Xia, Ping Wang, Boyang Chen. 2023. A multi-dimensional analysis on potential drivers of China's city-level low-carbon economy from the perspective of spatial spillover effects. *Journal of Cleaner Production* **419**, 138300. [Crossref]
24. Stuart T. Jones, Edward H. Allison, Kailin Kroetz, Yoshitaka Ota, Sunny L. Jardine. 2023. Enrollment, retention, and inclusivity of Marine Stewardship Council (MSC) eco-labelling certifications. *Marine Policy* **155**, 105734. [Crossref]
25. Brian C. Prest, Casey J. Wichman, Karen Palmer. 2023. RCTs against the Machine: Can Machine Learning Prediction Methods Recover Experimental Treatment Effects?. *Journal of the Association of Environmental and Resource Economists* **10**:5, 1231-1264. [Crossref]
26. Yuqing Zheng, Azucena Gracia, Lijiao Hu. 2023. Predicting Foodborne Disease Outbreaks with Food Safety Certifications: Econometric and Machine Learning Analyses. *Journal of Food Protection* **86**:9, 100136. [Crossref]
27. Yanyan Gao, Lin Zhang, Yongqing Nan. 2023. Travel to breathe the fresh air? Big data evidence on the short-term migration effect of air pollution from China. *China Economic Review* **14**, 102070. [Crossref]
28. Rama K. Malladi. 2023. Benchmark Analysis of Machine Learning Methods to Forecast the U.S. Annual Inflation Rate During a High-Decile Inflation Period. *Computational Economics* **4**. . [Crossref]
29. Emanuel Kohlscheen. 2023. Forecasting oil prices with random forests. *Empirical Economics* **30**. . [Crossref]
30. . Data-Driven Research 129-150. [Crossref]
31. Rafael Boix-Domenech, Vittorio Galletto, Fabio Sforzi, Francesco Capone. 2023. Living innovation machines: modelling innovation in time and space variable-geometry territorial units using machine learning. *European Planning Studies* **31**:7, 1422-1442. [Crossref]
32. Yanjie Fan. 2023. Towards Enterprise Business Management Strategy Based on Big Data Technique. *Journal of Education, Humanities and Social Sciences* **16**, 200-207. [Crossref]
33. Byron Botha, Rulof Burger, Kevin Kotzé, Neil Rankin, Daan Steenkamp. 2023. Big data forecasting of South African inflation. *Empirical Economics* **65**:1, 149-188. [Crossref]
34. Thomas Zellweger, Todd Zenger. 2023. Entrepreneurs as Scientists: A Pragmatist Approach to Producing Value Out of Uncertainty. *Academy of Management Review* **48**:3, 379-408. [Crossref]
35. Wen Lin. 2023. The effect of product quantity on willingness to pay: A meta-regression analysis of beef valuation studies. *Agribusiness* **39**:3, 646-663. [Crossref]
36. Michele Rabasco, Pietro Battiston. 2023. Predicting the deterrence effect of tax audits. A machine learning approach. *Metroeconomica* **74**:3, 531-556. [Crossref]
37. Anoop Valiya Veettil, Ashok K. Mishra. 2023. Quantifying Thresholds for Advancing Impact-based Drought Assessment Using Classification and Regression Tree (CART) Models. *Journal of Hydrology* **26**, 129966. [Crossref]

38. David De Cremer, Devesh Narayanan. 2023. How AI tools can—and cannot—help organizations become more ethical. *Frontiers in Artificial Intelligence* **6**. . [Crossref]

39. Masanori Kuroki. 2023. Integrating data science into an econometrics course with a Kaggle competition. *The Journal of Economic Education* 1-15. [Crossref]

40. Kyle Lauriston Smith. 2023. Thomas Aquinas, Ronald Dworkin, and the Fourth Revolution: The Foundations of Law in the Age of Surveillance Capitalism. *Laws* **12**:3, 40. [Crossref]

41. Muhammad Asif Khan, Juan E.Trinidad Segovia, M.Ishaq Bhatti, Asif Kabir. 2023. Corporate vulnerability in the US and China during COVID-19: A machine learning approach. *The Journal of Economic Asymmetries* **27**, e00302. [Crossref]

42. Hamed Taherdoost. 2023. Enhancing Social Media Platforms with Machine Learning Algorithms and Neural Networks. *Algorithms* **16**:6, 271. [Crossref]

43. Juan Laborda, Sonia Ruano, Ignacio Zamanillo. 2023. Multi-Country and Multi-Horizon GDP Forecasting Using Temporal Fusion Transformers. *Mathematics* **11**:12, 2625. [Crossref]

44. Gustavo Silva Araujo, Wagner Piazza Gaglianone. 2023. Machine learning methods for inflation forecasting in Brazil: New contenders versus classical models. *Latin American Journal of Central Banking* **4**:2, 100087. [Crossref]

45. Georgios Sermpinis, Serafeim Tsoukas, Yiqun Zhang. 2023. Modelling failure rates with machine-learning models: Evidence from a panel of UK firms. *European Financial Management* **29**:3, 734-763. [Crossref]

46. Julie C. Lauffenburger, Zhigang Lu, Mufaddal Mahesri, Erin Kim, Angela Tong, Seoyoung C. Kim. 2023. Using Data-Driven Approaches to Classify and Predict Health Care Spending in Patients With Gout Using Urate-Lowering Therapy. *Arthritis Care & Research* **75**:6, 1300-1310. [Crossref]

47. Shu Wang, Hao Luo, Shanshan Huang, Qingsong Li, Li Liu, Guoxin Su, Ming Liu. 2023. Counterfactual-based minority oversampling for imbalanced classification. *Engineering Applications of Artificial Intelligence* **122**, 106024. [Crossref]

48. Mikko Ranta, Mika Ylinen, Marko Järvenpää. 2023. Machine Learning in Management Accounting Research: Literature Review and Pathways for the Future. *European Accounting Review* **32**:3, 607-636. [Crossref]

49. Muhammad Fahim Uddin. An Enhanced Machine Learning Approach to Identify Noise and Detect Relevant Structures for Predictive Modeling 55-60. [Crossref]

50. Md. Mohsan Khudri, Kang Keun Rhee, Mohammad Shabbir Hasan, Karar Zunaid Ahsan. 2023. Predicting nutritional status for women of childbearing age from their economic, health, and demographic features: A supervised machine learning approach. *PLOS ONE* **18**:5, e0277738. [Crossref]

51. Raman Verma. The Human Side of Big Data: A Detailed Technical Analysis 2240-2244. [Crossref]

52. Xingchi Shen, Yueming Lucy Qiu, Xing Bo, Anand Patwardhan, Nathan Hultman, Bing Dong. 2023. The impact of co-adopting electric vehicles, solar photovoltaics, and battery storage on electricity consumption patterns: Empirical evidence from Arizona. *Resources, Conservation and Recycling* **192**, 106914. [Crossref]

53. Hongyan Dai, Qin Xiao, Songlin Chen, Weihua Zhou. 2023. Data-driven demand forecast for O2O operations: An adaptive hierarchical incremental approach. *International Journal of Production Economics* **259**, 108833. [Crossref]

54. Jochen Wirtz, Werner H. Kunz, Nicole Hartley, James Tarbit. 2023. Corporate Digital Responsibility in Service Firms and Their Ecosystems. *Journal of Service Research* **26**:2, 173-190. [Crossref]

55. Paul Corral, Heath Henderson, Sandra Segovia. Poverty Mapping in the Age of Machine Learning **25**, . [Crossref]

56. Khrystyna Bochkay, Stephen V. Brown, Andrew J. Leone, Jennifer Wu Tucker. 2023. Textual Analysis in Accounting: What's Next?*. *Contemporary Accounting Research* **40**:2, 765-805. [Crossref]
57. Shuiyang Pan, Suwan(Cheng) Long, Yiming Wang, Ying Xie. 2023. Nonlinear asset pricing in Chinese stock market: A deep learning approach. *International Review of Financial Analysis* **87**, 102627. [Crossref]
58. Nuttanan Wichitaksorn, Yingyue Kang, Faqiang Zhang. 2023. Random feature selection using random subspace logistic regression. *Expert Systems with Applications* **217**, 119535. [Crossref]
59. Wen-Jie Xie, Na Wei, Wei-Xing Zhou. 2023. An interpretable machine-learned model for international oil trade network. *Resources Policy* **82**, 103513. [Crossref]
60. Xu Shao. 2023. An Empirical Study of the Role of Big Data Analytics in Corporate Decision Making. *Journal of Global Information Management* **31**:6, 1-19. [Crossref]
61. Lennon H. T. Choy, Winky K. O. Ho. 2023. The Use of Machine Learning in Real Estate Research. *Land* **12**:4, 740. [Crossref]
62. Guillaume Belly, Lukas Boeckelmann, Carlos Mateo Caicedo Graciano, Alberto Di Iorio, Klodiana Istrefi, Vasileios Siakoulis, Arthur Stalla-Bourdillon. 2023. Forecasting sovereign risk in the Euro area via machine learning. *Journal of Forecasting* **42**:3, 657-684. [Crossref]
63. Mehmet Güney Celbiş, Pui-hang Wong, Karima Kourtit, Peter Nijkamp. 2023. Impacts of the COVID-19 outbreak on older-age cohorts in European Labor Markets: A machine learning exploration of vulnerable groups. *Regional Science Policy & Practice* **15**:3, 559-584. [Crossref]
64. Wenjie Han, Yong Li, Yunpeng Li, Tao Huang. 2023. A deep learning model based on multi-source data for daily tourist volume forecasting. *Current Issues in Tourism* **261**, 1-19. [Crossref]
65. Binrong Wu, Lin Wang, Rui Tao, Yu-Rong Zeng. 2023. Interpretable tourism volume forecasting with multivariate time series under the impact of COVID-19. *Neural Computing and Applications* **35**:7, 5437-5463. [Crossref]
66. David D.C. Tarn, Juefan Wang. 2023. Can data analytics raise marketing agility?-A sense-and-respond perspective. *Information & Management* **60**:2, 103743. [Crossref]
67. Kenneth David Strang. An Empirical Illustration of How Socioeconomic Stakeholders Can Leverage AI and Big Data 22-43. [Crossref]
68. Yuexi Liu. 2023. A Machine Learning Approach for Selecting Directors of Chinese Listed Company. *Highlights in Business, Economics and Management* **5**, 380-389. [Crossref]
69. Jung Ryeol Park, Yituo Feng. 2023. Trajectory tracking of changes digital divide prediction factors in the elderly through machine learning. *PLOS ONE* **18**:2, e0281291. [Crossref]
70. Daniel Hoang, Kevin Wiegratz. 2023. Machine learning methods in finance: Recent applications and prospects. *European Financial Management* **24**. . [Crossref]
71. Mario D. Schultz, Peter Seele. 2023. Towards AI ethics' institutionalization: knowledge bridges from business ethics to advance organizational AI ethics. *AI and Ethics* **3**:1, 99-111. [Crossref]
72. Amaia Palencia-Esteban, Pedro Salas-Rojo. Intergenerational Mobility and Life Satisfaction in Spain 109-137. [Crossref]
73. Hamza Heni, S. Arona Diop, Jacques Renaud, Leandro C. Coelho. 2023. Measuring fuel consumption in vehicle routing: new estimation models using supervised learning. *International Journal of Production Research* **61**:1, 114-130. [Crossref]
74. T. M. Schromm, C. U. Grosse. 2023. From 2D projections to the 3D rotation matrix: an attempt for finding a machine learning approach for the efficient evaluation of mechanical joining elements in X-ray computed tomography volume data. *SN Applied Sciences* **5**:1. . [Crossref]

75. Marianne Wanamaker, Carola Frydman, Christian M. Dahl. 2023. Preface. *Explorations in Economic History* **87**, 101505. [Crossref]
76. Helen Margetts, Cosmina Dorobantu. Computational Social Science for Public Policy 3-18. [Crossref]
77. Sebastiano Manzan. Big Data and Computational Social Science for Economic Analysis and Policy 231-242. [Crossref]
78. Matheus Pereira Libório, Petr Iakovlevitch Ekel, Carlos Augusto Paiva da Silva Martins. 2023. Economic analysis through alternative data and big data techniques: what do they tell about Brazil?. *SN Business & Economics* **3**:1. . [Crossref]
79. Andrés Alonso, José Manuel Carbó, J. Manuel Marqués. 2023. Machine Learning methods in climate finance: a systematic review. *SSRN Electronic Journal* **225**. . [Crossref]
80. Christine Biju Jacob, K. Asha. Real-Estate Housing Market Analytics and Prediction Using Big Data for Post Pandemic Era 80-89. [Crossref]
81. Andrew J. Patton, Yasin Simsek. 2023. Generalized Autoregressive Score Trees and Forests. *SSRN Electronic Journal* **5**. . [Crossref]
82. Sunčica Rogić, Ljiljana Kašćelan. Decoding Customer Behaviour: Relevance of Web and Purchasing Behaviour in Predictive Response Modeling 369-380. [Crossref]
83. Georgios A. Papanastasopoulos, John Sorros, Antonios M. Vasilatos. 2023. Do Raw Accounting Data Convey Information About R&D Costs Accounting Treatment? An Investigation With Machine Learning. *SSRN Electronic Journal* **36**. . [Crossref]
84. Felipe Tumenas Marques, Francisco Louzada Neto. 2023. Formação estatística nos programas de pós-graduação brasileiros: análise das disciplinas oferecidas. *Educação e Pesquisa* **49**. . [Crossref]
85. Roberto Casarin, Stefano Grassi, Francesco Ravazzolo, Herman K. van Dijk. 2023. A flexible predictive density combination for large financial data sets in regular and crisis periods. *Journal of Econometrics* **44**, 105370. [Crossref]
86. Michele Lenza, Inès Moutachaker, Joan Paredes. 2023. Density Forecasts of Inflation: A Quantile Regression Forest Approach. *SSRN Electronic Journal* **109**. . [Crossref]
87. Leonardo Madio, Francesco Principe. 2023. Who Supports Liberal Policies? A Tale of Two Referendums in Italy. *SSRN Electronic Journal* **56**. . [Crossref]
88. Douglas Araujo. 2023. Gingado: A Machine Learning Library Focused on Economics and Finance. *SSRN Electronic Journal* **59**. . [Crossref]
89. Xinyuan Lin, Wenjun Wang, Fa-Hsiang Chang. The Demand for Big Data Skills in China 711-727. [Crossref]
90. Augusto Cerqua, Marco Letta, Fiammetta Menchetti. 2023. Losing Control (Group)? The Machine Learning Control Method for Counterfactual Forecasting. *SSRN Electronic Journal* **59**. . [Crossref]
91. Sanjay Goswami, Jyoti Mishra, Mahendra Tiwari. Data Analytics Incorporated with Machine Learning Approaches in Finance 73-93. [Crossref]
92. Bowei Guo, Ao Sun, Feng Song. 2023. Time is Money: The Social Benefits of Time-of-Use Tariffs. *SSRN Electronic Journal* **105**. . [Crossref]
93. Hicham Sadok, Fadi Sakka, Mohammed El Hadi El Maknouzi. 2022. Artificial intelligence and bank credit analysis: A review. *Cogent Economics & Finance* **10**:1. . [Crossref]
94. Nimisha Malik, Bhavik Agarwal. Time Series Nowcasting of India's GDP with Machine Learning 1-6. [Crossref]
95. Ilse Botha. 2022. Time series forecasting in the artificial intelligence milieu. *Journal of Economic and Financial Sciences* **15**:1. . [Crossref]

96. Kenneth David Strang. How Could Machine Learning Help Healthcare Informatics Predict Coronavirus? 23-44. [Crossref]
97. Sunčica Rogić, Ljiljana Kašćelan, Vladimir Kašćelan, Vladimir Đurišić. 2022. Automatic customer targeting: a data mining solution to the problem of asymmetric profitability distribution. *Information Technology and Management* **23**:4, 315-333. [Crossref]
98. Francesco Decarolis, Cristina Giorgiantonio. 2022. Corruption red flags in public procurement: new evidence from Italian calls for tenders. *EPJ Data Science* **11**:1. . [Crossref]
99. Micha Kaiser, Steffen Otterbach, Alfonso Sousa-Poza. 2022. Using machine learning to uncover the relation between age and life satisfaction. *Scientific Reports* **12**:1. . [Crossref]
100. Ebru Çağlayan Akay, Naciye Tuba Yılmaz Soydan, Burcu Kocarık Gacar. 2022. Bibliometric analysis of the published literature on machine learning in economics and econometrics. *Social Network Analysis and Mining* **12**:1. . [Crossref]
101. Giovanni Cerulli. 2022. Machine learning using Stata/Python. *The Stata Journal: Promoting communications on statistics and Stata* **22**:4, 772-810. [Crossref]
102. Devesh Singh, Maciej Turała. 2022. Machine Learning and Regularization Technique to Determine Foreign Direct Investment in Hungarian Counties. *DANUBE* **13**:4, 269-291. [Crossref]
103. Chien-Ming Chi, Patrick Vossler, Yingying Fan, Jinchi Lv. 2022. Asymptotic properties of high-dimensional random forests. *The Annals of Statistics* **50**:6. . [Crossref]
104. Tong Feng, Huibin Du, Zhongguo Lin, Xudong Chen, Zhenni Chen, Qiang Tu. 2022. Green recovery or pollution rebound? Evidence from air pollution of China in the post-COVID-19 era. *Journal of Environmental Management* **324**, 116360. [Crossref]
105. Lidia Ceriani, Sergio Olivieri, Marco Ranzani. 2022. Housing, imputed rent, and household welfare. *The Journal of Economic Inequality* **38**. . [Crossref]
106. Katharine A. Anderson, Seth Richards-Shubik. 2022. Collaborative Production in Science: An Empirical Analysis of Coauthorships in Economics. *The Review of Economics and Statistics* **104**:6, 1241-1255. [Crossref]
107. Massimiliano Caporin, Mikhail Stolbov, Maria Shchepeleva. 2022. What drives the expansion of research on banking crises? Cross-country evidence. *Applied Economics* **54**:52, 6054-6064. [Crossref]
108. Zhe Jing, Yan Luo, Xiaotong Li, Xin Xu. 2022. A multi-dimensional city data embedding model for improving predictive analytics and urban operations. *Industrial Management & Data Systems* **122**:10, 2199-2216. [Crossref]
109. Binru Zhang, Nao Li, Rob Law, Heng Liu. 2022. A hybrid MIDAS approach for forecasting hotel demand using large panels of search data. *Tourism Economics* **28**:7, 1823-1847. [Crossref]
110. Patrick Krennmair, Timo Schmid. 2022. Flexible domain prediction using mixed effects random forests. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **71**:5, 1865-1894. [Crossref]
111. Guido de Blasio, Alessio D'Ignazio, Marco Letta. 2022. Gotham city. Predicting 'corrupted' municipalities with machine learning. *Technological Forecasting and Social Change* **184**, 122016. [Crossref]
112. Xue-Zhong He, Shen Lin. 2022. Reinforcement Learning Equilibrium in Limit Order Markets. *Journal of Economic Dynamics and Control* **144**, 104497. [Crossref]
113. David Mitre-Becerril, Sarah Tahamont, Jason Lerner, Aaron Chalfin. 2022. Can deterrence persist? Long-term evidence from a randomized experiment in street lighting. *Criminology & Public Policy* **21**:4, 865-891. [Crossref]
114. Rama K. Malladi. 2022. Application of Supervised Machine Learning Techniques to Forecast the COVID-19 U.S. Recession and Stock Market Crash. *Computational Economics* **8**. . [Crossref]

115. Christina H. Maaß. 2022. Shedding light on dark figures: Steps towards a methodology for estimating actual numbers of COVID-19 infections in Germany based on Google Trends. *PLOS ONE* **17**:10, e0276485. [Crossref]

116. Eder J. A. L. Pereira, Paulo Ferreira, Ivan C. da Cunha Lima, Thiago B. Murari, Marcelo A. Moret, Hernane B. de B. Pereira. 2022. Conservation in the Amazon rainforest and Google searches: A DCCA approach. *PLOS ONE* **17**:10, e0276675. [Crossref]

117. Anja K. Leist, Matthias Klee, Jung Hyun Kim, David H. Rehkopf, Stéphane P. A. Bordas, Graciela Muniz-Terrera, Sara Wade. 2022. Mapping of machine learning approaches for description, prediction, and causal inference in the social and health sciences. *Science Advances* **8**:42. . [Crossref]

118. Sathya Uma Lakshmi Kandasamy, Piyush Kumar Singh, Dillip Kumar Swain. 2022. Determination of Factors Affecting the Adoption of Integrated Farming System in Dryland Areas of Southern India by Using Supervised Learning Techniques. *Journal of Asian and African Studies* **66**, 002190962211303. [Crossref]

119. Steen Nielsen. 2022. Management accounting and the concepts of exploratory data analysis and unsupervised machine learning: a literature study and future directions. *Journal of Accounting & Organizational Change* **18**:5, 811-853. [Crossref]

120. Jorge Iván Pérez Rave, Gloria Patricia Jaramillo Álvarez, Favián González Echavarría. 2022. A psychometric data science approach to study latent variables: a case of class quality and student satisfaction. *Total Quality Management & Business Excellence* **33**:13-14, 1547-1572. [Crossref]

121. Alessandra Garbero, Marco Letta. 2022. Predicting household resilience with machine learning: preliminary cross-country tests. *Empirical Economics* **63**:4, 2057-2070. [Crossref]

122. Simon Blöthner, Mario Larch. 2022. Economic determinants of regional trade agreements revisited using machine learning. *Empirical Economics* **63**:4, 1771-1807. [Crossref]

123. Werner Kristjanpoller, Nicole Astudillo, Josephine E. Olson. 2022. An empirical application of a hybrid ANFIS model to predict household over-indebtedness. *Neural Computing and Applications* **34**:20, 17343-17353. [Crossref]

124. Neal Hughes, Wei Ying Soh, Kenton Lawson, Michael Lu. 2022. Improving the performance of micro-simulation models with machine learning: The case of Australian farms. *Economic Modelling* **115**, 105957. [Crossref]

125. Ankita Raj, Sunil Kumar Singh. Forecasting GDP of India and its neighbouring countries using Time Series Analysis 1-6. [Crossref]

126. Shuoli Zhao, Yufeng Lai, Chenglong Ye, Keehyun Lee. 2022. Machine learning applications in household-level demand prediction. *Applied Economics Letters* **8**, 1-7. [Crossref]

127. Francisco Corona, Graciela González-Farías, Jesús López-Pérez. 2022. Timely Estimates of the Monthly Mexican Economic Activity. *Journal of Official Statistics* **38**:3, 733-765. [Crossref]

128. Samuel Shamiri, Leanne Ngai, Peter Lake, Yin Shan, Amee McMillan, Therese Smith, Kishor Sharma. 2022. Nowcasting the Australian Labour Market at Disaggregated Levels. *Australian Economic Review* **55**:3, 389-404. [Crossref]

129. Zongying Wang. 2022. Evaluation and Analysis of the Informatization Degree of College English Education Based on Big Data Technology. *Security and Communication Networks* **2022**, 1-11. [Crossref]

130. Liang Lu, Guang Tian, Patrick Hatzenbuehler. 2022. How agricultural economists are using big data: a review. *China Agricultural Economic Review* **14**:3, 494-508. [Crossref]

131. David P Daniels, Daniella Kupor. 2022. The Magnitude Heuristic: Larger Differences Increase Perceived Causality. *Journal of Consumer Research* **10**. . [Crossref]

132. Takaaki Ikeda, Upul Cooray, Masanori Hariyama, Jun Aida, Katsunori Kondo, Masayasu Murakami, Ken Osaka. 2022. An Interpretable Machine Learning Approach to Predict Fall Risk Among Community-Dwelling Older Adults: a Three-Year Longitudinal Study. *Journal of General Internal Medicine* **37**:11, 2727-2735. [Crossref]

133. Chengbin Wang, Jianguo Chen, Yongpeng Ouyang. 2022. Determination of Predictive Variables in Mineral Prospectivity Mapping Using Supervised and Unsupervised Methods. *Natural Resources Research* **31**:4, 2081-2102. [Crossref]

134. Matthias X. Hanauer, Marina Kononova, Marc Steffen Rapp. 2022. Boosting agnostic fundamental analysis: Using machine learning to identify mispricing in European stock markets. *Finance Research Letters* **48**, 102856. [Crossref]

135. Mona Aghdaee, Bonny Parkinson, Kompal Sinha, Yuanyuan Gu, Rajan Sharma, Emma Olin, Henry Cutler. 2022. An examination of machine learning to map non-preference based patient reported outcome measures to health state utility values. *Health Economics* **31**:8, 1525-1557. [Crossref]

136. Alexandre Guedes, Samuel Faria, Sofia Gouveia, João Rebelo. 2022. The effect of virtual proximity and digital adoption on international tourism flows to Southern Europe. *Tourism Economics* **69**, 135481662211167. [Crossref]

137. Muhammed Sehid Gorus, Erdal Tanas Karagol. 2022. Factors affecting per capita ecological footprint in OECD countries: Evidence from machine learning techniques. *Energy & Environment* **69**, 0958305X2211129. [Crossref]

138. Donatus U Ekwueme, Michael T Halpern, Harrell W Chesson, Mahima Ashok, Jeffrey Drope, Young-Rock Hong, Michael Maciosek, Michael F Pesko, Donald S Kenkel. 2022. Health Economics Research in Primary Prevention of Cancer: Assessment, Current Challenges, and Future Directions. *JNCI Monographs* **2022**:59, 28-41. [Crossref]

139. Fotios Petropoulos, Daniele Apiletti, Vassilios Assimakopoulos, Mohamed Zied Babai, Devon K. Barrow, Souhaib Ben Taieb, Christoph Bergmeir, Ricardo J. Bessa, Jakub Bijak, John E. Boylan, Jethro Browell, Claudio Carnevale, Jennifer L. Castle, Pasquale Cirillo, Michael P. Clements, Clara Cordeiro, Fernando Luiz Cyrino Oliveira, Shari De Baets, Alexander Dokumentov, Joanne Ellison, Piotr Fiszeder, Philip Hans Franses, David T. Frazier, Michael Gilliland, M. Sinan Gönül, Paul Goodwin, Luigi Grossi, Yael Grushka-Cockayne, Mariangela Guidolin, Massimo Guidolin, Ulrich Gunter, Xiaojia Guo, Renato Guseo, Nigel Harvey, David F. Hendry, Ross Hollyman, Tim Januschowski, Jooyoung Jeon, Victor Richmond R. Jose, Yanfei Kang, Anne B. Koehler, Stephan Kolassa, Nikolaos Kourentzes, Sonia Leva, Feng Li, Konstantia Litsiou, Spyros Makridakis, Gael M. Martin, Andrew B. Martinez, Sheik Meeran, Theodore Modis, Konstantinos Nikolopoulos, Dilek Önkal, Alessia Paccagnini, Anastasios Panagiotelis, Ioannis Panapakidis, Jose M. Pavía, Manuela Pedio, Diego J. Pedregal, Pierre Pinson, Patrícia Ramos, David E. Rapach, J. James Reade, Bahman Rostami-Tabar, Michał Rubaszek, Georgios Sermpinis, Han Lin Shang, Evangelos Spiliotis, Aris A. Syntetos, Priyanga Dilini Talagala, Thiyanga S. Talagala, Len Tashman, Dimitrios Thomakos, Thordis Thorarinsdottir, Ezio Todini, Juan Ramón Trapero Arenas, Xiaoqian Wang, Robert L. Winkler, Alisa Yusupova, Florian Ziel. 2022. Forecasting: theory and practice. *International Journal of Forecasting* **38**:3, 705-871. [Crossref]

140. Catia Nicodemo, Albert Satorra. 2022. Exploratory data analysis on large data sets: The example of salary variation in Spanish Social Security Data. *BRQ Business Research Quarterly* **25**:3, 283-294. [Crossref]

141. Kim Christensen, Mathias Siggaard, Bezirgen Veliyev. 2022. A Machine Learning Approach to Volatility Forecasting. *Journal of Financial Econometrics* **39**. . [Crossref]

142. Zhixuan Xiao, Chengyi Li, Shihua Pan, Gaoda Wei, Mengmeng Tian, Runjiu Hu. 2022. Exploring the Spatial Impact of Multisource Data on Urban Vitality: A Causal Machine Learning Method. *Wireless Communications and Mobile Computing* **2022**, 1-24. [Crossref]

143. Stephen Jarvis, Olivier Deschenes, Akshaya Jha. 2022. The Private and External Costs of Germany's Nuclear Phase-Out. *Journal of the European Economic Association* **20**:3, 1311-1346. [Crossref]

144. Florian M. Artinger, Gerd Gigerenzer, Perke Jacobs. 2022. Satisficing: Integrating Two Traditions. *Journal of Economic Literature* **60**:2, 598-635. [Abstract] [View PDF article] [PDF with links]

145. Matteo Marsili, Yasser Roudi. 2022. Quantifying relevance in learning and inference. *Physics Reports* **963**, 1-43. [Crossref]

146. Monica Billio, Roberto Casarin, Fausto Corradin. Understanding Economic Instability during the Pandemic: A Factor Model Approach 1-55. [Crossref]

147. Alexandre Rubesam. 2022. Machine learning portfolios with equal risk contributions: Evidence from the Brazilian market. *Emerging Markets Review* **51**, 100891. [Crossref]

148. Samira Bounid, Mohammed Oughanem, Salman Bourkadi. Advanced Financial Data Processing and Labeling Methods for Machine Learning 1-6. [Crossref]

149. Ryan Engstrom, Jonathan Hersh, David Newhouse. 2022. Poverty from Space: Using High Resolution Satellite Imagery for Estimating Economic Well-being. *The World Bank Economic Review* **36**:2, 382-412. [Crossref]

150. Karsten Müller. 2022. German forecasters' narratives: How informative are German business cycle forecast reports?. *Empirical Economics* **62**:5, 2373-2415. [Crossref]

151. Manuel Monge, Carlos Poza, Sofía Borgia. 2022. A proposal of a suspicion of tax fraud indicator based on Google trends to foresee Spanish tax revenues. *International Economics* **169**, 1-12. [Crossref]

152. Matteo Alpino, Karen Evelyn Hauge, Andreas Kotsadam, Simen Markussen. 2022. Effects of dialogue meetings on sickness absence—Evidence from a large field experiment. *Journal of Health Economics* **83**, 102615. [Crossref]

153. Yoshiyuki Suimon, Hiroto Tanabe. Construction of real-time manufacturing industry production activity estimation models using high-frequency electricity demand data 1-7. [Crossref]

154. Yang Yi, Le Wen, Shan He. 2022. Partitioning for "Common but Differentiated" Precise Air Pollution Governance: A Combined Machine Learning and Spatial Econometric Approach. *Energies* **15**:9, 3346. [Crossref]

155. Valentas Gružauskas, Aurelija Burinskienė. 2022. Managing Supply Chain Complexity and Sustainability: The Case of the Food Industry. *Processes* **10**:5, 852. [Crossref]

156. Laion Lima Boaventura, Paulo Henrique Ferreira, Rosemeire Leovigildo Fiaccone. 2022. On flexible Statistical Process Control with Artificial Intelligence: Classification control charts. *Expert Systems with Applications* **194**, 116492. [Crossref]

157. Siyavash Filom, Amir M. Amiri, Saiedeh Razavi. 2022. Applications of machine learning methods in port operations – A systematic literature review. *Transportation Research Part E: Logistics and Transportation Review* **161**, 102722. [Crossref]

158. XI CHEN, YANG HA (TONY) CHO, YIWEI DOU, BARUCH LEV. 2022. Predicting Future Earnings Changes Using Machine Learning and Detailed Financial Data. *Journal of Accounting Research* **60**:2, 467-515. [Crossref]

159. MIAO LIU. 2022. Assessing Human Information Processing in Lending Decisions: A Machine Learning Approach. *Journal of Accounting Research* **60**:2, 607-651. [Crossref]

160. Joey Blumberg, Gary Thompson. 2022. Nonparametric segmentation methods: Applications of unsupervised machine learning and revealed preference. *American Journal of Agricultural Economics* **104**:3, 976-998. [Crossref]

161. Felipe D. Calainho, Alex M. van de Minne, Marc K. Francke. 2022. A Machine Learning Approach to Price Indices: Applications in Commercial Real Estate. *The Journal of Real Estate Finance and Economics* **46**. . [Crossref]

162. Steve J. Bickley, Ho Fai Chan, Benno Torgler. 2022. Artificial intelligence in the field of economics. *Scientometrics* **127**:4, 2055-2084. [Crossref]

163. Kim Anh Thi Nguyen, Tram Anh Thi Nguyen, Brice M. Nguelifack, Curtis M. Jolly. 2022. Machine Learning Approaches for Predicting Willingness to Pay for Shrimp Insurance in Vietnam. *Marine Resource Economics* **37**:2, 155-182. [Crossref]

164. Xiaohang Ren, Kun Duan, Lizhu Tao, Yukun Shi, Cheng Yan. 2022. Carbon prices forecasting in quantiles. *Energy Economics* **108**, 105862. [Crossref]

165. Ines Levin. 2022. Learning about Spatial and Temporal Proximity using Tree-Based Methods. *Statistics, Politics and Policy* **13**:1, 73-95. [Crossref]

166. Samira Barzin, Paolo Avner, Jun Rentschler, Neave O'Clery. Where are All the Jobs? A Machine Learning Approach for High Resolution Urban Employment Prediction in Developing Countries **86**, . [Crossref]

167. Li Zhang. Big Data, Factor Agglomeration of Urban Economics and Spatial Chaos 45-49. [Crossref]

168. Florian Wozny. 2022. The Impact of COVID-19 on Airfares—A Machine Learning Counterfactual Analysis. *Econometrics* **10**:1, 8. [Crossref]

169. Márton Gosztonyi, Csákné Filep Judit. 2022. Profiling (Non-)Nascent Entrepreneurs in Hungary Based on Machine Learning Approaches. *Sustainability* **14**:6, 3571. [Crossref]

170. Jan Abrell, Mirjam Kosch, Sebastian Rausch. 2022. How effective is carbon pricing?—A machine learning approach to policy evaluation. *Journal of Environmental Economics and Management* **112**, 102589. [Crossref]

171. Aaron Chalfin, Benjamin Hansen, Jason Lerner, Lucie Parker. 2022. Reducing Crime Through Environmental Design: Evidence from a Randomized Experiment of Street Lighting in New York City. *Journal of Quantitative Criminology* **38**:1, 127-157. [Crossref]

172. Ti-Ching Peng, Chun-Chieh Wang. 2022. The Application of Machine Learning Approaches on Real-Time Apartment Prices in the Tokyo Metropolitan Area. *Social Science Japan Journal* **25**:1, 3-28. [Crossref]

173. Miaoying Shi, Jintao Xu, Shilei Liu, Zhenci Xu. 2022. Productivity-Based Land Suitability and Management Sensitivity Analysis: The Eucalyptus E. urophylla × E. grandis Case. *Forests* **13**:2, 340. [Crossref]

174. Zhewei Liu, Jianxiao Liu, Xiao Huang, Erchen Zhang, Biyu Chen. 2022. Measuring Chinese cities' economic development with mobile application usage. *Journal of Geographical Sciences* **32**:12, 2415-2429. [Crossref]

175. ANDREAS FUSTER, PAUL GOLDSMITH-PINKHAM, TARUN RAMADORAI, ANSGAR WALTHER. 2022. Predictably Unequal? The Effects of Machine Learning on Credit Markets. *The Journal of Finance* **77**:1, 5-47. [Crossref]

176. Faridoon Khan, Amena Urooj, Saud Ahmed Khan, Saima K. Khosa, Sara Muhammadullah, Zahra Almaspoor. 2022. Evaluating the Performance of Feature Selection Methods Using Huge Big Data: A Monte Carlo Simulation Approach. *Mathematical Problems in Engineering* **2022**, 1-10. [Crossref]

177. Dweepobotee Brahma, Debasri Mukherjee. 2022. Early warning signs: targeting neonatal and infant mortality using machine learning. *Applied Economics* **54**:1, 57-74. [Crossref]

178. Yang Bao, Gilles Hilary, Bin Ke. Artificial Intelligence and Fraud Detection 223-247. [Crossref]

179. Gareth Macartney. Econometrics in Litigation: Challenges at Class Certification 311-346. [Crossref]

180. Jiachuan Wang, Tanmay Bagla, Sneha Srivastava, Adam Teehan, Khurshid Ahmad. Market Movements at High Frequencies and Latency in Response Times 943-961. [Crossref]

181. Daniel Hain, Roman Jurowetzki. Introduction to Rare-Event Predictive Modeling for Inferential Statisticians—A Hands-On Application in the Prediction of Breakthrough Patents 49-83. [Crossref]
182. Ben Vinod. Artificial Intelligence and Emerging Technologies in Hospitality 279-313. [Crossref]
183. Karsten Lübke, Bianca Krol. Empirisch-quantitative Abschlussarbeiten – Ein Blick nach vorne 499-509. [Crossref]
184. Zhigang Qiu, Xiaolin Huo, Yue Dai. Development of FinTech in Academia 71-84. [Crossref]
185. Rishabh Kalai, Rajeev Ramesh, Karthik Sundararajan. Machine Learning Models for Predictive Analytics in Personal Finance 241-254. [Crossref]
186. Marçal Farré, Federico Todeschini, Didier Grimaldi, Carlos Carrasco-Farré. Data-driven policy evaluation 197-225. [Crossref]
187. Nicholas J. Hallman, Antonis Kartapanis, Jaime J. Schmidt. 2022. How do auditors respond to competition? Evidence from the bidding process. *Journal of Accounting and Economics* 101475. [Crossref]
188. Montserrat González Garibay, Andrej Srakar, Tjaša Bartolj, Jože Sambt. 2022. Does Machine Learning Offer Added Value Vis-à-Vis Traditional Statistics? An Exploratory Study on Retirement Decisions Using Data from the Survey of Health, Ageing, and Retirement in Europe (SHARE). *Mathematics* **10**:1, 152. [Crossref]
189. Fausto Corradin, Monica Billio, Roberto Casarin. 2022. Forecasting Economic Indicators with Robust Factor Models. *National Accounting Review* **4**:2, 167-190. [Crossref]
190. Kenneth David Strang, Zhaohao Sun. Managerial Controversies in Artificial Intelligence and Big Data Analytics 1745-1764. [Crossref]
191. Yigit Aydogan. Generating Big Data in (Micro)Economics 126-148. [Crossref]
192. Benjamin Bluhm, Jannic Alexander Cutura. 2022. Econometrics at Scale: Spark up Big Data in Economics. *Journal of Data Science* **49**, 413-436. [Crossref]
193. Emilio Lehoucq. 2022. Do Americans Think the Digital Economy is Fair? Using Supervised Learning to Explore Evaluations of Predictive Automation. *Journal of Data Science* **35**, 381-399. [Crossref]
194. Mariano Maisonnave, Fernando Delbianco, Fernando Tohme, Evangelos Milios, Ana G. Maguitman. 2022. Causal graph extraction from news: a comparative study of time-series causality learning techniques. *PeerJ Computer Science* **8**, e1066. [Crossref]
195. Baojun Yu, Changming Li, Nawazish Mirza, Muhammad Umar. 2022. Forecasting credit ratings of decarbonized firms: Comparative assessment of machine learning models. *Technological Forecasting and Social Change* **174**, 121255. [Crossref]
196. Augusto Cerqua, Marco Letta. 2022. Local inequalities of the COVID-19 crisis. *Regional Science and Urban Economics* **92**, 103752. [Crossref]
197. Duo Qin. 2022. Redirect the Probability Approach in Econometrics Towards PAC Learning. *SSRN Electronic Journal* **105**. . [Crossref]
198. Nuttanan Wichitaksorn, Yingyue Kang, Faqiang Zhang. 2022. Random Feature Selection Using Random Subspace Logistic Regression. *SSRN Electronic Journal* **39**. . [Crossref]
199. Jens Ludwig, Sendhil Mullainathan. 2022. Algorithmic Behavioral Science: Machine Learning as a Tool for Scientific Discovery. *SSRN Electronic Journal* **18**. . [Crossref]
200. Johanna Barop. 2022. Correlation or Causation? Identification! - Directed acyclic graphs as an identification framework in econometrics. *SSRN Electronic Journal* **128**. . [Crossref]
201. Kajal Lahiri, Cheng Yang. Time Series Models 1-36. [Crossref]
202. Veli Andirin, Yusuf Neggers, Mehdi Shadmehr, Jesse Shapiro. 2022. Measuring the Tolerance of the State: Theory and Application to Protest. *SSRN Electronic Journal* **96**. . [Crossref]

203. Jon Ellingsen, Vegard H. Larsen, Leif Anders Thorsrud. 2022. News media versus FRED-MD for macroeconomic forecasting. *Journal of Applied Econometrics* **37**:1, 63-81. [Crossref]
204. Simone Plak, Ilja Cornelisz, Martijn Meeter, Chris Klaveren. 2022. Early warning systems for more effective student counselling in higher education: Evidence from a Dutch field experiment. *Higher Education Quarterly* **76**:1, 131-152. [Crossref]
205. Scott Guernsey, Feng Guo, Tingting Liu, Matthew Serfling. 2022. Thirty Years of Change: The Evolution of Classified Boards. *SSRN Electronic Journal* **95**. . [Crossref]
206. Faridoon Khan, Amena Urooj, Saud Ahmed Khan, Abdelaziz Alsubie, Zahra Almaspoor, Sara Muhammadullah. 2021. Comparing the Forecast Performance of Advanced Statistical and Machine Learning Techniques Using Huge Big Data: Evidence from Monte Carlo Experiments. *Complexity* **2021**, 1-11. [Crossref]
207. Faridoon Khan, Amena Urooj, Kalim Ullah, Badr Alnssyan, Zahra Almaspoor. 2021. A Comparison of Autometrics and Penalization Techniques under Various Error Distributions: Evidence from Monte Carlo Simulation. *Complexity* **2021**, 1-8. [Crossref]
208. Youren Zhang, Wenxi Xie, Zhengxun He, Yifan Ren, Ziyan Jiang. A Study of Counterfactual Inference Based on Instrumental Variables and Machine Learning 30-34. [Crossref]
209. Francesco Bloise, Paolo Brunori, Patrizio Piraino. 2021. Estimating intergenerational income mobility on sub-optimal data: a machine learning approach. *The Journal of Economic Inequality* **19**:4, 643-665. [Crossref]
210. Joep Steegmans. 2021. The Pearls and Perils of Google Trends: A Housing Market Application. *Big Data* **9**:6, 443-453. [Crossref]
211. Onder Ozgur, Erdal Tanas Karagol, Fatih Cemil Ozbugday. 2021. Machine learning approach to drivers of bank lending: evidence from an emerging economy. *Financial Innovation* **7**:1. . [Crossref]
212. Mehmet Güney Celbiş, Pui-Hang Wong, Karima Kourtit, Peter Nijkamp. 2021. Innovativeness, Work Flexibility, and Place Characteristics: A Spatial Econometric and Machine Learning Approach. *Sustainability* **13**:23, 13426. [Crossref]
213. Sumit Agarwal, Long Wang, Yang Yang. 2021. Impact of transboundary air pollution on service quality and consumer satisfaction. *Journal of Economic Behavior & Organization* **192**, 357-380. [Crossref]
214. Yunsong Chen, Xiaogang Wu, Anning Hu, Guangye He, Guodong Ju. 2021. Social prediction: a new research paradigm based on machine learning. *The Journal of Chinese Sociology* **8**:1. . [Crossref]
215. Jiaming Zhang, Zhanfeng Li, Xinyuan Song, Hanwen Ning. 2021. Deep Tobit networks: A novel machine learning approach to microeconometrics. *Neural Networks* **144**, 279-296. [Crossref]
216. Antonio Rodríguez Andrés, Abraham Otero, Voxi Heinrich Amavilah. 2021. Using deep learning neural networks to predict the knowledge economy index for developing and emerging economies. *Expert Systems with Applications* **184**, 115514. [Crossref]
217. Maja Micevska. 2021. Revisiting forced migration: A machine learning perspective. *European Journal of Political Economy* **70**, 102044. [Crossref]
218. Christian Handke, Lucie Guibault, Joan-Josep Vallbé. 2021. Copyright's impact on data mining in academic research. *Managerial and Decision Economics* **42**:8, 1999-2016. [Crossref]
219. Pratumtip Thongcharoen. Lessons Learned on Knowledge Management to Produce Online Teaching Materials for PSU-MOOC: A Case Study of the Subject 'Social Etiquette in the Digital Age' 76-81. [Crossref]
220. Deon Filmer, Vatsal Nahata, Shwetlena Sabarwal. Preparation, Practice, and Beliefs: A Machine Learning Approach to Understanding Teacher Effectiveness **1**, . [Crossref]

221. Chengge Wu. Empowering Financial Technical Analysis using Computer Vision Techniques 179-184. [Crossref]
222. Weiqing Zhuang. 2021. The Influence of Big Data Analytics on E-Commerce: Case Study of the U.S. and China. *Wireless Communications and Mobile Computing* **2021**, 1-20. [Crossref]
223. Mehmet Güney CELBİŞ. 2021. Social Networks, Female Unemployment, and the Urban-Rural Divide in Turkey: Evidence from Tree-Based Machine Learning Algorithms. *Sosyoekonomi* **29**:50, 73-93. [Crossref]
224. Jia Li, Yuhong Jiang. 2021. The Research Trend of Big Data in Education and the Impact of Teacher Psychology on Educational Development During COVID-19: A Systematic Review and Future Perspective. *Frontiers in Psychology* **12**. . [Crossref]
225. Prodosh E. Simlai. 2021. Predicting owner-occupied housing values using machine learning: an empirical investigation of California census tracts data. *Journal of Property Research* **38**:4, 305-336. [Crossref]
226. Lauri Paavola, Richard Cuthbertson. 2021. Algorithms Creating Paradoxes of Power: Explore, Exploit, Embed, Embalm. *Information Systems Management* **38**:4, 358-371. [Crossref]
227. Augusto Cerqua, Roberta Di Stefano, Marco Letta, Sara Miccoli. 2021. Local mortality estimates during the COVID-19 pandemic in Italy. *Journal of Population Economics* **34**:4, 1189-1217. [Crossref]
228. Jesús Gonzalo, Jean-Yves Pitarakis. 2021. Spurious relationships in high-dimensional systems with strong or mild persistence. *International Journal of Forecasting* **37**:4, 1480-1497. [Crossref]
229. Alvaro Escribano, Daniel Peña, Esther Ruiz. 2021. 30 years of cointegration and dynamic factor models forecasting and its future with big data: Editorial. *International Journal of Forecasting* **37**:4, 1333-1337. [Crossref]
230. Hal Varian. 2021. Economics at Google. *Business Economics* **56**:4, 195-199. [Crossref]
231. Ranjith Vijayakumar, Mike W.-L. Cheung. 2021. Assessing Replicability of Machine Learning Results: An Introduction to Methods on Predictive Accuracy in Social Sciences. *Social Science Computer Review* **39**:5, 768-801. [Crossref]
232. Jennifer L. Castle, Jurgen A. Doornik, David F. Hendry. 2021. Modelling non-stationary 'Big Data'. *International Journal of Forecasting* **37**:4, 1556-1575. [Crossref]
233. Amit Kumar Kushwaha, Prashant Kumar, Arpan Kumar Kar. 2021. What impacts customer experience for B2B enterprises on using AI-enabled chatbots? Insights from Big data analytics. *Industrial Marketing Management* **98**, 207-221. [Crossref]
234. Alexandre Bonnet R. Costa, Pedro Cavalcanti G. Ferreira, Wagner P. Gaglianone, Osmani Teixeira C. Guillén, João Victor Issler, Yihao Lin. 2021. Machine learning and oil price point and density forecasting. *Energy Economics* **102**, 105494. [Crossref]
235. Matteo Iacopini, Carlo R.M.A. Santagiustina. 2021. Filtering the Intensity of Public Concern from Social Media Count Data with Jumps. *Journal of the Royal Statistical Society Series A: Statistics in Society* **184**:4, 1283-1302. [Crossref]
236. Nobuyuki Nakamura, Aya Suzuki. 2021. COVID-19 and the intentions to migrate from developing countries: Evidence from online search activities in Southeast Asia. *Journal of Asian Economics* **76**, 101348. [Crossref]
237. Giovanni Cerulli. 2021. Improving econometric prediction by machine learning. *Applied Economics Letters* **28**:16, 1419-1425. [Crossref]
238. Raúl Rodríguez-Luna, Margareth Mercado-Pérez, Mariana Escobar-Borja. 2021. Big data y cadenas de suministros un binomio complejo para américa latina. *Aibi revista de investigación, administración e ingeniería* **8**:S1, 16-23. [Crossref]

239. Marina Johnson, Rashmi Jain, Peggy Brennan-Tonetta, Ethne Swartz, Deborah Silver, Jessica Paolini, Stanislav Mamonov, Chelsey Hill. 2021. Impact of Big Data and Artificial Intelligence on Industry: Developing a Workforce Roadmap for a Data Driven Economy. *Global Journal of Flexible Systems Management* **22**:3, 197-217. [Crossref]

240. Wen-Jie Xie, Mu-Yao Li, Wei-Xing Zhou. 2021. Learning representation of stock traders and immediate price impacts. *Emerging Markets Review* **48**, 100791. [Crossref]

241. Galina Alova, Ben Caldecott. 2021. A machine learning model to investigate factors contributing to the energy transition of utility and independent power producer sectors internationally. *iScience* **24**:9, 102929. [Crossref]

242. Feras A. Batarseh, Munisamy Gopinath, Anderson Monken, Zhengrong Gu. 2021. Public policymaking for international agricultural trade using association rules and ensemble machine learning. *Machine Learning with Applications* **5**, 100046. [Crossref]

243. Rafael Quintana. 2021. Who Belongs in School? Using Statistical Learning Techniques to Identify Linear, Nonlinear and Interactive Effects. *The Quantitative Methods for Psychology* **17**:3, 312-328. [Crossref]

244. Hossein Hassani, Xu Huang, Steve MacFeely, Mohammad Reza Entezarian. 2021. Big Data and the United Nations Sustainable Development Goals (UN SDGs) at a Glance. *Big Data and Cognitive Computing* **5**:3, 28. [Crossref]

245. James T. Bang, Atin Basuchoudhary, Aniruddha Mitra. 2021. Validating Game-Theoretic Models of Terrorism: Insights from Machine Learning. *Games* **12**:3, 54. [Crossref]

246. John M. Griffin, Samuel Kruger, Gonzalo Maturana. 2021. What drove the 2003–2006 house price boom and subsequent collapse? Disentangling competing explanations. *Journal of Financial Economics* **141**:3, 1007-1035. [Crossref]

247. Yucheng Zhang, Shan Xu, Long Zhang, Mengxi Yang. 2021. Big data and human resource management research: An integrative review and new directions for future research. *Journal of Business Research* **133**, 34-50. [Crossref]

248. Akash Malhotra. 2021. A hybrid econometric–machine learning approach for relative importance analysis: prioritizing food policy. *Eurasian Economic Review* **11**:3, 549-581. [Crossref]

249. Virgilio Galdo, Yue Li, Martin Rama. 2021. Identifying urban areas by combining human judgment and machine learning: An application to India. *Journal of Urban Economics* **125**, 103229. [Crossref]

250. Kristian Bondo Hansen, Christian Borch. 2021. The absorption and multiplication of uncertainty in machine-learning-driven finance. *The British Journal of Sociology* **72**:4, 1015-1029. [Crossref]

251. Zhang Wu, Terence Tai-Leung Chong, Yuchen Liu. 2021. Market Reaction to iPhone Rumors. *Algorithmic Finance* **9**:1-2, 1-23. [Crossref]

252. Fengjun Tian, Yang Yang, Zhenxing Mao, Wenyue Tang. 2021. Forecasting daily attraction demand using big data from search engines and social media. *International Journal of Contemporary Hospitality Management* **33**:6, 1950-1976. [Crossref]

253. Guilherme Lindenmeyer, Pedro Pablo Skorin, Hudson da Silva Torrent. 2021. Using boosting for forecasting electric energy consumption during a recession: a case study for the Brazilian State Rio Grande do Sul. *Letters in Spatial and Resource Sciences* **14**:2, 111-128. [Crossref]

254. Marja-Liisa Halko, Olli Lappalainen, Lauri Sääksvuori. 2021. Do non-choice data reveal economic preferences? Evidence from biometric data and compensation-scheme choice. *Journal of Economic Behavior & Organization* **188**, 87-104. [Crossref]

255. Max Besbris, Ariela Schachter, John Kuk. 2021. The Unequal Availability of Rental Housing Information Across Neighborhoods. *Demography* **58**:4, 1197-1221. [Crossref]

256. Mehmet Güney Celbiş. 2021. A machine learning approach to rural entrepreneurship. *Papers in Regional Science* **100**:4, 1079-1104. [Crossref]
257. Semen Yu. BOGATYREV. 2021. Heuristics as a new way of adjusting the end market value. *Finance and Credit* **27**:7, 1581-1599. [Crossref]
258. Dante Adalberto Avaro. 2021. Algoritmos y pandemia. Tres claves emergentes para futuros análisis sobre opinión pública. *Revista Mexicana de Opinión Pública* :31, 41-53. [Crossref]
259. Mirta Galesic, Wändi Bruine de Bruin, Jonas Dalege, Scott L. Feld, Frauke Kreuter, Henrik Olsson, Drazen Prelec, Daniel L. Stein, Tamara van der Does. 2021. Human social sensing is an untapped resource for computational social science. *Nature* **595**:7866, 214-222. [Crossref]
260. Marco Guerzoni, Consuelo R. Nava, Massimiliano Nuccio. 2021. Start-ups survival through a crisis. Combining machine learning with econometrics to measure innovation. *Economics of Innovation and New Technology* **30**:5, 468-493. [Crossref]
261. Giorgio Gnecco, Federico Nutarelli, Daniela Selvi. 2021. Optimal data collection design in machine learning: the case of the fixed effects generalized least squares panel data model. *Machine Learning* **110**:7, 1549-1584. [Crossref]
262. Giorgio Gnecco, Federico Nutarelli. 2021. On the trade-off between number of examples and precision of supervision in machine learning problems. *Optimization Letters* **15**:5, 1711-1733. [Crossref]
263. Richard G. Newell, Brian C. Prest, Steven E. Sexton. 2021. The GDP-Temperature relationship: Implications for climate change damages. *Journal of Environmental Economics and Management* **108**, 102445. [Crossref]
264. Tom Coupe. 2021. How global is the affordable housing crisis?. *International Journal of Housing Markets and Analysis* **14**:3, 429-445. [Crossref]
265. David Easley, Marcos López de Prado, Maureen O'Hara, Zhibai Zhang. 2021. Microstructure in the Machine Age. *The Review of Financial Studies* **34**:7, 3316-3363. [Crossref]
266. René Böheim, Philipp Stöllinger. 2021. Decomposition of the gender wage gap using the LASSO estimator. *Applied Economics Letters* **28**:10, 817-828. [Crossref]
267. Giovanni Di Franco, Michele Santurro. 2021. Machine learning, artificial neural networks and social research. *Quality & Quantity* **55**:3, 1007-1025. [Crossref]
268. Gang Xie, Xin Li, Yatong Qian, Shouyang Wang. 2021. Forecasting tourism demand with KPCA-based web search indexes. *Tourism Economics* **27**:4, 721-743. [Crossref]
269. Chiara Binelli. 2021. Estimating Causal Effects When the Treatment Affects All Subjects Simultaneously: An Application. *Big Data and Cognitive Computing* **5**:2, 22. [Crossref]
270. S.Y BOGATYREV. 2021. BEHAVIORAL DEVIATIONS OF THE MARKET VALUE OF MACHINERY, EQUIPMENT, VEHICLES AND CURRENCY ASSETS. *AZIMUTH OF SCIENTIFIC RESEARCH: ECONOMICS AND ADMINISTRATION* **10**:35. . [Crossref]
271. Semen Yu. BOGATYREV. 2021. Simulation of emotional differences in the structured query language for databases of financial markets. *Financial Analytics: Science and Experience* **14**:2, 156-173. [Crossref]
272. Semen Yu. BOGATYREV. 2021. New finance: Psychological measurement of value. *Finance and Credit* **27**:5, 1156-1177. [Crossref]
273. Chonbadee Juthamanee, Krerk Piromsopa, Prabhas Chongstitvatana. Token Allocation for Course Bidding With Machine Learning Method 1168-1171. [Crossref]
274. Danilo Bertoni, Giacomo Aletti, Daniele Cavicchioli, Alessandra Micheletti, Roberto Pretolani. 2021. Estimating the CAP greening effect by machine learning techniques: A big data ex post analysis. *Environmental Science & Policy* **119**, 44-53. [Crossref]

275. Eirik Sjåholm Knudsen, Lasse B. Lien, Bram Timmermans, Ivan Belik, Sujit Pandey. 2021. Stability in turbulent times? The effect of digitalization on the sustainability of competitive advantage. *Journal of Business Research* **128**, 360-369. [Crossref]
276. Wolfram Höpken, Tobias Eberle, Matthias Fuchs, Maria Lexhagen. 2021. Improving Tourist Arrival Prediction: A Big Data and Artificial Neural Network Approach. *Journal of Travel Research* **60**:5, 998-1017. [Crossref]
277. Yash Raj Shrestha, Vivianna Fang He, Phanish Puranam, Georg von Krogh. 2021. Algorithm Supported Induction for Building Theory: How Can We Use Prediction Models to Theorize?. *Organization Science* **32**:3, 856-880. [Crossref]
278. Sergei Yu. BOGATYREV. 2021. Looking into bubbles in financial markets and the emotional side of corporate forecast completion through modeling in the structured query language of financial databases. *Finance and Credit* **27**:4, 833-850. [Crossref]
279. Miriam Steurer, Robert J. Hill, Norbert Pfeifer. 2021. Metrics for evaluating the performance of machine learning based automated valuation models. *Journal of Property Research* **38**:2, 99-129. [Crossref]
280. Małgorzata Grządzielewska. 2021. Using Machine Learning in Burnout Prediction: A Survey. *Child and Adolescent Social Work Journal* **38**:2, 175-180. [Crossref]
281. Shunqin Chen, Zhengfeng Guo, Xinlei Zhao. 2021. Predicting mortgage early delinquency with machine learning methods. *European Journal of Operational Research* **290**:1, 358-372. [Crossref]
282. Sergei Yu. BOGATYREV. 2021. The sentiment analysis method in finance: The psychological-financial index. *Finance and Credit* **27**:3, 561-584. [Crossref]
283. Jochen Hartmann. Classification Using Decision Tree Ensembles 103-117. [Crossref]
284. Jalayer Khalilzadeh. 2021. Predictive policing in hospitality and tourism venues — The case of Orlando. *Journal of Destination Marketing & Management* **19**, 100535. [Crossref]
285. Yuying Lin, Yanhai Zhou, Mingshui Lin, Shidai Wu, Baoyin Li. 2021. Exploring the disparities in park accessibility through mobile phone data: Evidence from Fuzhou of China. *Journal of Environmental Management* **281**, 111849. [Crossref]
286. Mark D. Flood. Financial Crises and the Macroeconomy 1-7. [Crossref]
287. Lulin Xu, Zhongwu Li. 2021. A New Appraisal Model of Second-Hand Housing Prices in China's First-Tier Cities Based on Machine Learning Algorithms. *Computational Economics* **57**:2, 617-637. [Crossref]
288. Modhurima Dey Amin, Syed Badruddoza, Jill J. McCluskey. 2021. Predicting access to healthful food retailers with machine learning. *Food Policy* **99**, 101985. [Crossref]
289. Toshiaki Aizawa. 2021. Decomposition of Improvements in Infant Mortality in Asian Developing Countries Over Three Decades. *Demography* **58**:1, 137-163. [Crossref]
290. Galina Alova, Philipp A. Trotter, Alex Money. 2021. A machine-learning approach to predicting Africa's electricity mix based on planned power plants and their chances of success. *Nature Energy* **6**:2, 158-166. [Crossref]
291. Ti-Ching Peng. 2021. The effect of hazard shock and disclosure information on property and land prices: a machine-learning assessment in the case of Japan. *Review of Regional Research* **39**. . [Crossref]
292. Andreas Fagereng, Martin Blomhoff Holm, Kjersti Næss Torstensen. 2021. Housing wealth in Norway, 1993–20151. *Journal of Economic and Social Measurement* **45**:1, 65-81. [Crossref]
293. Achim Ahrens, Christopher Aitken, Jan Ditzen, Erkal Ersoy, David Kohns, Mark E. Schaffer. A Theory-Based Lasso for Time-Series Data 3-36. [Crossref]
294. Vadim I. Marshev. Actual Problems and Concepts of Management 619-699. [Crossref]

295. Peter Romero, Stephen Fitz. The Use of Psychometrics and Artificial Intelligence in Alternative Finance 511-587. [Crossref]
296. Peng Cheng, Laurent Ferrara, Alice Froidevaux, Thanh-Long Huynh. Massive Data Analytics for Macroeconomic Nowcasting 145-167. [Crossref]
297. Hirofumi Fukuyama, William L. Weber. Network DEA and Big Data with an Application to the Coronavirus Pandemic 175-197. [Crossref]
298. R. B. Barman. Sustaining High Economic Growth Requires a Different Strategy: An Integrated Approach for Broad-Based Knowledge Economy 1-20. [Crossref]
299. Jaehyun Yoon. 2021. Forecasting of Real GDP Growth Using Machine Learning Models: Gradient Boosting and Random Forest Approach. *Computational Economics* **57**:1, 247-265. [Crossref]
300. Jing Liang, Yueming Qiu, Bo Xing. 2021. Social Versus Private Benefits of Energy Efficiency Under Time-of-Use and Increasing Block Pricing. *Environmental and Resource Economics* **78**:1, 43-75. [Crossref]
301. Solomon Y. Deku, Alper Kara, Artur Semeyutin. 2021. The predictive strength of MBS yield spreads during asset bubbles. *Review of Quantitative Finance and Accounting* **56**:1, 111-142. [Crossref]
302. Christophre Georges, Javier Pereira. 2021. Market stability with machine learning agents. *Journal of Economic Dynamics and Control* **122**, 104032. [Crossref]
303. Yin Wenjing. 2021. The Development of Interior Decoration Design Under the Background of Big Data. *E3S Web of Conferences* **253**, 02084. [Crossref]
304. Kristof Lommers, Ouns El Harzli, Jack Kim. 2021. Confronting Machine Learning with Financial Research. *SSRN Electronic Journal* **1**. . [Crossref]
305. Edward Oughton, William Lehr. 2021. Next-G Wireless: Learning from 5G Techno-Economics to Inform Next Generation Wireless Technologies. *SSRN Electronic Journal* **36**. . [Crossref]
306. William Lehr. 2021. Smart Contracts, Real-Virtual World Convergence and Economic Implications. *SSRN Electronic Journal* **3**. . [Crossref]
307. M. Dziamulych, T. Shmatkovska, O. Borysiuk. 2021. Big data and their role in the digital economy formation. *Galic'kij ekonomičnij visnik* **70**:3, 16-21. [Crossref]
308. Domenico Giannone, Michele Lenza, Giorgio E. Primiceri. 2021. Economic Predictions With Big Data: The Illusion of Sparsity. *Econometrica* **89**:5, 2409-2437. [Crossref]
309. Alessandro Sontuoso, Sudeep Bhatia. 2021. A notion of prominence for games with natural-language labels. *Quantitative Economics* **12**:1, 283-312. [Crossref]
310. Namratha Birudaraju, Adiraju Prasanth Rao, Sathiyamoorthi V.. Architecture for Analyzing Agriculture Data Using Data Analytics 111-122. [Crossref]
311. Jan Kalina. Managerial Decision Support in the Post-COVID-19 Era 225-241. [Crossref]
312. Diego Vicentin. Da cibernética às sociedades de controle 181-210. [Crossref]
313. Ben Vinod. Artificial Intelligence and Emerging Technologies in Travel 313-337. [Crossref]
314. Tony Liu, Lyle Ungar, Konrad Kording. 2021. Quantifying causality in data science with quasi-experiments. *Nature Computational Science* **1**:1, 24-32. [Crossref]
315. Mehmet Güney Celbiş. Applications of Machine Learning Models in Regional and Demographic Economic Analysis: A Literature Survey 219-229. [Crossref]
316. Scott Mongeau, Andrzej Hajdasinski. Phase III: CSDS Gap-Prescriptions—Design Science Problem-Solving 201-316. [Crossref]
317. Mikko Ranta, Mika Ylinen, Marko Järvenpää. 2021. Machine learning in management accounting research: Literature review and pathways for the future. *SSRN Electronic Journal* **53**. . [Crossref]

318. Paul Hünermund, Jermain Kaminski, Carla Schmitt. 2021. Causal Machine Learning and Business Decision Making. *SSRN Electronic Journal* **33**. . [Crossref]
319. William Lehr. 2021. Smart Contracts, Real-Virtual World Convergence and Economic Implications. *SSRN Electronic Journal* **3**. . [Crossref]
320. Simon Blöthner, Mario Larch. 2021. Economic Determinants of Regional Trade Agreements Revisited Using Machine Learning. *SSRN Electronic Journal* **7**. . [Crossref]
321. James T. E. Chapman, Ajit Desai. 2021. Macroeconomic Predictions using Payments Data and Machine Learning. *SSRN Electronic Journal* **29**. . [Crossref]
322. Abdollah Farhoodi, Nazanin Khazra, Peter Christensen. 2021. Does Airbnb Reduce Matching Frictions in the Housing Market?. *SSRN Electronic Journal* **355**. . [Crossref]
323. Grazia Cecere, Nicoletta Corrocher, Clara Jean. 2021. Fair or Unbiased Algorithmic Decision-Making? A Review of the Literature on Digital Economics. *SSRN Electronic Journal* **13**. . [Crossref]
324. Emanuel Kohlscheen. 2021. What does machine learning say about the drivers of inflation?. *SSRN Electronic Journal* **45**. . [Crossref]
325. Karthik Babu Nattamai Kannan, Govinda Dhungana, Glenn B. Voss. 2021. Managing Supply and Demand for the Performing Arts in the Time of COVID. *SSRN Electronic Journal* **72**. . [Crossref]
326. Guillaume Belly, Boeckelmann Lukas, Carlos Mateo Caicedo Graciano, Alberto Di Iorio, Klodiana Istrefi, Vasileios Siakoulis, Arthur Stalla-Bourdillon. 2021. Forecasting Sovereign Risk in the Euro Area via Machine Learning. *SSRN Electronic Journal* **34**. . [Crossref]
327. Felix Chopra. 2021. Media Persuasion and Consumption: Evidence from the Dave Ramsey Show. *SSRN Electronic Journal* **80**. . [Crossref]
328. Kathy Baylis, Thomas Heckelei, Hugo Storm. Machine learning in agricultural economics 4551-4612. [Crossref]
329. Paolo Giordani. 2021. Smartboost Learning for Tabular Data. *SSRN Electronic Journal* **68**. . [Crossref]
330. Andrew Harvey, Stephen Thiele. 2021. Cointegration and control: Assessing the impact of events using time series data. *Journal of Applied Econometrics* **36**:1, 71-85. [Crossref]
331. Alex Singleton, Daniel Arribas-Bel. 2021. Geographic Data Science. *Geographical Analysis* **53**:1, 61-75. [Crossref]
332. Tuğba KARABOĞA, Cemal ZEHİR. 2020. Büyük Verinin Etkin Yönetiminde Stratejik Uyum ve Veri Odaklı Kültür. *IBAD Sosyal Bilimler Dergisi* :8, 63-76. [Crossref]
333. Grazia Cecere, Thierry Pénard. 2020. Introduction to the Special Issue: "From The digital economy to the digitalization of the economy". *Revue d'économie industrielle* :172, 11-17. [Crossref]
334. Jun Yang, Xiaoming Li, Shoujun Huang. 2020. Impacts on environmental quality and required environmental regulation adjustments: A perspective of directed technical change driven by big data. *Journal of Cleaner Production* **275**, 124126. [Crossref]
335. Songul Cinaroglu. 2020. The impact of oversampling with "ubSMOTE" on the performance of machine learning classifiers in prediction of catastrophic health expenditures. *Operations Research for Health Care* **27**, 100275. [Crossref]
336. Sandra Achten, Christian Lessmann. 2020. Spatial inequality, geography and economic activity. *World Development* **136**, 105114. [Crossref]
337. Federico Belotti, Franco Peracchi. 2020. Fast leave-one-out methods for inference, model selection, and diagnostic checking. *The Stata Journal: Promoting communications on statistics and Stata* **20**:4, 785-804. [Crossref]
338. Ran Li, Bingcheng Yang, Jerrod Penn, Bailey Houghtaling, Juan Chen, Witoon Prinyawiwatkul, Brian E. Roe, Danyi Qi. 2020. Perceived vulnerability to COVID-19 infection from event attendance:

results from Louisiana, USA, two weeks preceding the national emergency declaration. *BMC Public Health* **20**:1. . [Crossref]

339. Julie C. Lauffenburger, Mufaddal Mahesri, Niteesh K. Choudhry. 2020. Not there yet: using data-driven methods to predict who becomes costly among low-cost patients with type 2 diabetes. *BMC Endocrine Disorders* **20**:1. . [Crossref]
340. Wanderson Rocha Bittencourt, Pedro H. M. Albuquerque. 2020. Evaluating company bankruptcies using causal forests. *Revista Contabilidade & Finanças* **31**:84, 542-559. [Crossref]
341. Manuel J. García Rodríguez, Vicente Rodríguez Montequín, Francisco Ortega Fernández, Joaquín M. Villanueva Balsera. 2020. Bidders Recommender for Public Procurement Auctions Using Machine Learning: Data Analysis, Algorithm, and Case Study with Tenders from Spain. *Complexity* **2020**, 1-20. [Crossref]
342. Michael T. Kiley. 2020. Financial Conditions and Economic Activity: Insights from Machine Learning. *Finance and Economics Discussion Series* **2020**:095, 1-40. [Crossref]
343. Nicholas Charron, Paola Annoni. 2020. What is the Influence of News Media on People's Perception of Corruption? Parametric and Non-Parametric Approaches. *Social Indicators Research* **70**. . [Crossref]
344. Giorgio Gnecco, Federico Nutarelli, Daniela Selvi. 2020. Optimal trade-off between sample size, precision of supervision, and selection probabilities for the unbalanced fixed effects panel data model. *Soft Computing* **24**:21, 15937-15949. [Crossref]
345. Ross Brown, Augusto Rocha. 2020. Entrepreneurial uncertainty during the Covid-19 crisis: Mapping the temporal dynamics of entrepreneurial finance. *Journal of Business Venturing Insights* **14**, e00174. [Crossref]
346. Jenny W. Sun, Jessica M. Franklin, Kathryn Rough, Rishi J. Desai, Sonia Hernández-Díaz, Krista F. Huybrechts, Brian T. Bateman. 2020. Predicting overdose among individuals prescribed opioids using routinely collected healthcare utilization data. *PLOS ONE* **15**:10, e0241083. [Crossref]
347. Winky K.O. Ho, Bo-Sin Tang, Siu Wai Wong. 2020. Predicting property prices with machine learning algorithms. *Journal of Property Research* **147**, 1-23. [Crossref]
348. Seth Richards-Shubik. Application and Computation of a Flexible Class of Network Formation Models 111-142. [Crossref]
349. Barteld Braaksma, Kees Zeelenberg, Sofie De Broe. Big Data in Official Statistics 303-338. [Crossref]
350. Julie C. Lauffenburger, Mufaddal Mahesri, Niteesh K. Choudhry. 2020. Use of Data-Driven Methods to Predict Long-term Patterns of Health Care Spending for Medicare Patients. *JAMA Network Open* **3**:10, e2020291. [Crossref]
351. Jens Prüfer, Patricia Prüfer. 2020. Data science for entrepreneurship research: studying demand dynamics for entrepreneurial skills in the Netherlands. *Small Business Economics* **55**:3, 651-672. [Crossref]
352. Jermain C. Kaminski, Christian Hopp. 2020. Predicting outcomes in crowdfunding campaigns with textual, visual, and linguistic signals. *Small Business Economics* **55**:3, 627-649. [Crossref]
353. Vessela Daskalova, Nicolaas J. Vriend. 2020. Categorization and coordination. *European Economic Review* **129**, 103519. [Crossref]
354. Arpan Kumar Kar, Yogesh K. Dwivedi. 2020. Theory building with big data-driven research – Moving away from the "What" towards the "Why". *International Journal of Information Management* **54**, 102205. [Crossref]
355. Carlos Poza, Manuel Monge. 2020. A real time leading economic indicator based on text mining for the Spanish economy. Fractional cointegration VAR and Continuous Wavelet Transform analysis. *International Economics* **163**, 163-175. [Crossref]

356. Evgeny A. Antipov, Elena B. Pokryshevskaya. 2020. Interpretable machine learning for demand modeling with high-dimensional data using Gradient Boosting Machines and Shapley values. *Journal of Revenue and Pricing Management* **19**:5, 355-364. [Crossref]

357. Zhongqi Deng, Yu Zhang, Ao Yu. 2020. The New Economy in China: An Intercity Comparison. *SAGE Open* **10**:4, 215824402097787. [Crossref]

358. SeyedSoroosh Azizi, Kiana Yektansani. 2020. Artificial Intelligence and Predicting Illegal Immigration to the USA. *International Migration* **58**:5, 183-193. [Crossref]

359. Sangchul Park, Haksoo Ko. 2020. Machine Learning and Law and Economics: A Preliminary Overview. *Asian Journal of Law and Economics* **11**:2. . [Crossref]

360. Sangchul Park, Haksoo Ko. 2020. Machine Learning and Law and Economics: A Preliminary Overview. *Asian Journal of Law and Economics* **11**:2. . [Crossref]

361. Mark F. J. Steel. 2020. Model Averaging and Its Use in Economics. *Journal of Economic Literature* **58**:3, 644-719. [Abstract] [View PDF article] [PDF with links]

362. Mikhail Stolbov, Maria Shchepeleva. 2020. What predicts the legal status of cryptocurrencies?. *Economic Analysis and Policy* **67**, 273-291. [Crossref]

363. Huamao Wang, Yumei Yao, Said Salhi. 2020. Tension in big data using machine learning: Analysis and applications. *Technological Forecasting and Social Change* **158**, 120175. [Crossref]

364. Mitsuru Igami. 2020. Artificial intelligence as structural estimation: Deep Blue, Bonanza, and AlphaGo. *The Econometrics Journal* **23**:3, S1-S24. [Crossref]

365. Jeppe Druedahl, Anders Munk-Nielsen. 2020. Higher-order income dynamics with linked regression trees. *The Econometrics Journal* **23**:3, S25-S58. [Crossref]

366. Gary Smith. 2020. Data mining fool's gold. *Journal of Information Technology* **35**:3, 182-194. [Crossref]

367. Carolina Rojas-Córdova, Boris Heredia-Rojas, Patricio Ramírez-Correa. 2020. Predicting Business Innovation Intention Based on Perceived Barriers: A Machine Learning Approach. *Symmetry* **12**:9, 1381. [Crossref]

368. Velibor V. Mišić. 2020. Optimization of Tree Ensembles. *Operations Research* **68**:5, 1605-1624. [Crossref]

369. Robert J. Hill, Miriam Steurer. 2020. Commercial Property Price Indices and Indicators: Review and Discussion of Issues Raised in the CPPI Statistical Report of Eurostat (2017). *Review of Income and Wealth* **66**:3, 736-751. [Crossref]

370. Ryan Engstrom, David Newhouse, Vidhya Soundararajan. 2020. Estimating small-area population density in Sri Lanka using surveys and Geo-spatial data. *PLOS ONE* **15**:8, e0237063. [Crossref]

371. Matthew A. Cole, Robert J R Elliott, Bowen Liu. 2020. The Impact of the Wuhan Covid-19 Lockdown on Air Pollution and Health: A Machine Learning and Augmented Synthetic Control Approach. *Environmental and Resource Economics* **76**:4, 553-580. [Crossref]

372. 2020. GlobalSearchRegression.jl: \ Building bridges between Machine Learning and Econometrics in Fat-Data scenarios. *JuliaCon Proceedings* **2**:13, 53. [Crossref]

373. Werickson Fortunato de Carvalho Rocha, Charles Bezerra do Prado, Niksa Blonder. 2020. Comparison of Chemometric Problems in Food Analysis Using Non-Linear Methods. *Molecules* **25**:13, 3025. [Crossref]

374. Andres Algaba, David Ardia, Keven Bluteau, Samuel Borms, Kris Boudt. 2020. ECONOMETRICS MEETS SENTIMENT: AN OVERVIEW OF METHODOLOGY AND APPLICATIONS. *Journal of Economic Surveys* **34**:3, 512-547. [Crossref]

375. Silvia Emili, Attilio Gardini, Enrico Foscolo. 2020. High spatial and temporal detail in timely prediction of tourism demand. *International Journal of Tourism Research* **22**:4, 451-463. [Crossref]
376. Ron Tidhar, Kathleen M. Eisenhardt. 2020. Get rich or die trying… finding revenue model fit using machine learning and multiple cases. *Strategic Management Journal* **41**:7, 1245-1273. [Crossref]
377. Hugo Storm, Kathy Baylis, Thomas Heckelei. 2020. Machine learning in agricultural and applied economics. *European Review of Agricultural Economics* **47**:3, 849-892. [Crossref]
378. Jian-qiang Guo, Shu-hen Chiang, Min Liu, Chi-Chun Yang, Kai-yi Guo. 2020. CAN MACHINE LEARNING ALGORITHMS ASSOCIATED WITH TEXT MINING FROM INTERNET DATA IMPROVE HOUSING PRICE PREDICTION PERFORMANCE?. *International Journal of Strategic Property Management* **24**:5, 300-312. [Crossref]
379. D. D. Li, D. X. Yu, Z. J. Qu, S. H. Yu. 2020. Feature Selection and Model Fusion Approach for Predicting Urban Macro Travel Time. *Mathematical Problems in Engineering* **2020**, 1-13. [Crossref]
380. Gary Smith. 2020. The paradox of big data. *SN Applied Sciences* **2**:6. . [Crossref]
381. Juan D. Montoro-Pons, Manuel Cuadrado-García. 2020. Music festivals as mediators and their influence on consumer awareness. *Poetics* **80**, 101424. [Crossref]
382. Abigail Devereaux, Linan Peng. 2020. Give us a little social credit: to design or to discover personal ratings in the era of Big Data. *Journal of Institutional Economics* **16**:3, 369-387. [Crossref]
383. Kenneth David Strang, Zhaohao Sun. 2020. Hidden big data analytics issues in the healthcare industry. *Health Informatics Journal* **26**:2, 981-998. [Crossref]
384. Hideaki IWATA. 2020. Non-Steady Trading Day Detection Based on Stock Index Time-Series Information. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences* **E103.A**:6, 821-828. [Crossref]
385. Marios Poulos, Nikolaos Korfiatis, Sozon Papavlassopoulos. 2020. Assessing stationarity in web analytics: A study of bounce rates. *Expert Systems* **37**:3. . [Crossref]
386. Misheck Mutize, McBride Peter Nkhalamba. 2020. A comparative study of economic growth as a key determinant of sovereign credit ratings in Africa. *International Journal of Emerging Markets* **ahead-of-print**:ahead-of-print. . [Crossref]
387. Heiko Kirchhain, Jan Mutl, Joachim Zietz. 2020. The Impact of Exogenous Shocks on House Prices: the Case of the Volkswagen Emissions Scandal. *The Journal of Real Estate Finance and Economics* **60**:4, 587-610. [Crossref]
388. Vivek Anand Asokan, Masaru Yarime, Motoharu Onuki. 2020. A review of data-intensive approaches for sustainability: methodology, epistemology, normativity, and ontology. *Sustainability Science* **15**:3, 955-974. [Crossref]
389. Christophe Croux, Julapa Jagtiani, Tarunsai Korivi, Milos Vulanovic. 2020. Important factors determining Fintech loan default: Evidence from a lendingclub consumer platform. *Journal of Economic Behavior & Organization* **173**, 270-296. [Crossref]
390. Feiyu Hu, Jim Warren, Daniel J. Exeter. 2020. Predicting Lipid-Lowering Medication Persistence after the First Cardiovascular Disease Hospitalization. *Methods of Information in Medicine* **59**:02/03, 061-074. [Crossref]
391. Marcos M. López de Prado. Machine Learning for Asset Managers **44**, . [Crossref]
392. Georges Bresson. Comments on "An Econometrician's Perspective on Big Data" by Cheng Hsiao 431-443. [Crossref]
393. Xia Liu. 2020. Analyzing the impact of user-generated content on B2B Firms' stock performance: Big data analysis with machine learning methods. *Industrial Marketing Management* **86**, 30-39. [Crossref]

394. Kenneth David Strang. 2020. Problems with research methods in medical device big data analytics. *International Journal of Data Science and Analytics* **9**:2, 229-240. [Crossref]
395. Chinmay Kakatkar, Volker Bilgram, Johann Füller. 2020. Innovation analytics: Leveraging artificial intelligence in the innovation process. *Business Horizons* **63**:2, 171-181. [Crossref]
396. Hailong Cui, Sampath Rajagopalan, Amy R. Ward. 2020. Predicting product return volume using machine learning methods. *European Journal of Operational Research* **281**:3, 612-627. [Crossref]
397. Achim Ahrens, Christian B. Hansen, Mark E. Schaffer. 2020. lassopack: Model selection and prediction with regularized regression in Stata. *The Stata Journal: Promoting communications on statistics and Stata* **20**:1, 176-235. [Crossref]
398. Mathias Bärtl, Simone Krummaker. 2020. Prediction of Claims in Export Credit Finance: A Comparison of Four Machine Learning Techniques. *Risks* **8**:1, 22. [Crossref]
399. S.Yu. Bogatyrev. 2020. The behavioral valuation apparatus. *Finance and Credit* **26**:2, 257-269. [Crossref]
400. David Easley, Eleonora Patacchini, Christopher Rojas. 2020. Multidimensional diffusion processes in dynamic online networks. *PLOS ONE* **15**:2, e0228421. [Crossref]
401. David Lenz, Peter Winker. 2020. Measuring the diffusion of innovations with paragraph vector topic models. *PLOS ONE* **15**:1, e0226685. [Crossref]
402. Kong Lu. Computer Performance Determination System Based on Big Data Distributed File 877-884. [Crossref]
403. Giorgio Gnecco, Federico Nutarelli. On the Trade-Off Between Number of Examples and Precision of Supervision in Regression 1-6. [Crossref]
404. W. O. K. I. S. Wijesinghe, C. U. Kumarasinghe, J. Mannapperuma, K. L. D. U. Liyanage. Socioeconomic Status Classification of Geographic Regions in Sri Lanka Through Anonymized Call Detail Records 299-311. [Crossref]
405. Federico Bassetti, Roberto Casarin, Francesco Ravazzolo. Density Forecasting 465-494. [Crossref]
406. Clement Bellet, Paul Frijters. Big Data and Wellbeing: An Economic Perspective 175-206. [Crossref]
407. Mitja Kovač. Introduction to the Autonomous Artificial Intelligence Systems 47-63. [Crossref]
408. Jan Kalina. On Sensitivity of Metalearning: An Illustrative Study for Robust Regression 261-270. [Crossref]
409. Tiemo Thiess, Oliver Müller. Designing Causal Inference Systems for Value-Based Spare Parts Pricing 191-204. [Crossref]
410. Giorgio Gnecco, Stefano Amato, Alessia Patuelli, Nicola Lattanzi. Machine Learning Application to Family Business Status Classification 25-36. [Crossref]
411. Uwe Rudolf Fingerlos, Guido Golla, Alexander Pastwa, Peter Gluchowski, Roland Gabriel. Technologische und konzeptionelle Ansätze zur Analyse und zum Reporting von Risikodaten in Kreditinstituten 19-96. [Crossref]
412. Uwe Rudolf Fingerlos, Guido Golla, Alexander Pastwa, Peter Gluchowski, Roland Gabriel. Detailliertes Fallbeispiel zur Kreditdatenanalyse auf Basis von RStudio 155-383. [Crossref]
413. Cinzia Daraio. Nonparametric Methods and Higher Education 2109-2114. [Crossref]
414. Vikram Dayal. Introduction 3-8. [Crossref]
415. Vikram Dayal. From Trees to Random Forests 315-326. [Crossref]
416. Marcus H. Böhme, André Gröger, Tobias Stöhr. 2020. Searching for a better life: Predicting international migration with online search keywords. *Journal of Development Economics* **142**, 102347. [Crossref]

417. Carlos León, Paolo Barucca, Oscar Acero, Gerardo Gage, Fabio Ortega. 2020. Pattern recognition of financial institutions' payment behavior. *Latin American Journal of Central Banking* **1**:1-4, 100011. [Crossref]
418. Kristian Bondo Hansen. 2020. The virtue of simplicity: On machine learning models in algorithmic trading. *Big Data & Society* **7**:1, 205395172092655. [Crossref]
419. Jorge Iván Pérez Rave, Favián González Echavarría, Juan Carlos Correa Morales. 2020. Modeling of apartment prices in a Colombian context from a machine learning approach with stable-important attributes. *DYNA* **87**:212, 63-72. [Crossref]
420. Emerson Escolar, Yasuaki Hiraoka, Mitsuru Igami, Yasin Ozcan. 2020. Mapping Firms' Locations in Technological Space: A Topological Analysis of Patent Statistics. *SSRN Electronic Journal* **31**. . [Crossref]
421. Elena-Ivona Dumitrescu, Sullivan Hué, Christophe Hurlin, sessi tokpavi. 2020. Machine Learning or Econometrics for Credit Scoring: Let's Get the Best of Both Worlds. *SSRN Electronic Journal* . [Crossref]
422. Yiannis Dendramis, Elias Tzavalis, Aikaterini Cheimarioti. 2020. Measuring the Default Risk of Small Business Loans: Improved Credit Risk Prediction using Deep Learning. *SSRN Electronic Journal* **13**. . [Crossref]
423. Yang Bao, Gilles Hilary, Bin Ke. 2020. Artificial Intelligence and Fraud Detection. *SSRN Electronic Journal* . [Crossref]
424. Muhammad Mahboob Ali. 2020. Digitization of the emerging economy: An exploratory and explanatory case study. *Journal of Governance and Regulation* **9**:4, 25-36. [Crossref]
425. Marian Socoliuc, Cristina-Gabriela Cosmulese, Marius-Sorin Ciubotariu, Svetlana Mihaila, Iulia-Diana Arion, Veronica Grosu. 2020. Sustainability Reporting as a Mixture of CSR and Sustainable Development. A Model for Micro-Enterprises within the Romanian Forestry Sector. *Sustainability* **12**:2, 603. [Crossref]
426. Valerio Lemma. Fintech, Chain Transactions and Open Banking 245-297. [Crossref]
427. Laurie A. Schintler. Regional Policy Analysis in the Era of Spatial Big Data 93-109. [Crossref]
428. Tianyi Li, Munther Dahleh. 2020. Automation of Data Acquisition Strategies in Model Calibration for System Models: Sensor Placement. *SSRN Electronic Journal* **32**. . [Crossref]
429. Yulin Liu, Luyao Zhang. 2020. Cryptocurrency Valuation and Machine Learning. *SSRN Electronic Journal* **6**. . [Crossref]
430. Xi Chen, Yang Ha Cho, Yiwei Dou, Baruch Itamar Lev. 2020. Fundamental Analysis of XBRL Data: A Machine Learning Approach. *SSRN Electronic Journal* **58**. . [Crossref]
431. Catia Nicodemo, albert satorra. 2020. Exploratory Data Analysis on Large Data Sets: The Example of Salary Variation in Spanish Social Security Data. *SSRN Electronic Journal* **2**. . [Crossref]
432. Thiago Christiano Silva, Benjamin Miranda Tabak, Idamar Magalhães Ferreira. 2019. Modeling Investor Behavior Using Machine Learning: Mean-Reversion and Momentum Trading Strategies. *Complexity* **2019**, 1-14. [Crossref]
433. Jiafu An, Raghavendra Rau. 2019. Finance, technology and disruption. *The European Journal of Finance* **12**, 1-12. [Crossref]
434. Huy Duc Dang, Au Hai Thi Dam, Thuyen Thi Pham, Tra My Thi Nguyen. 2019. Determinants of credit demand of farmers in Lam Dong, Vietnam. *Agricultural Finance Review* **80**:2, 255-274. [Crossref]
435. Celso Martínez Musiño. 2019. Big Data-Análisis informétrico de documentos indexados en Scopus y Web of Science. *Investigación Bibliotecológica: archivonomía, bibliotecología e información* **34**:82, 87. [Crossref]

436. Arthur Lewbel. 2019. The Identification Zoo: Meanings of Identification in Econometrics. *Journal of Economic Literature* **57**:4, 835-903. [Abstract] [View PDF article] [PDF with links]
437. Stefan P. Penczynski. 2019. Using machine learning for communication classification. *Experimental Economics* **22**:4, 1002-1029. [Crossref]
438. Kurt Stockinger, Nils Bundi, Jonas Heitz, Wolfgang Breymann. 2019. Scalable architecture for Big Data financial analytics: user-defined functions vs. SQL. *Journal of Big Data* **6**:1. . [Crossref]
439. Wenbo Wu, Jiaqi Chen, Liang Xu, Qingyun He, Michael L. Tindall. 2019. A statistical learning approach for stock selection in the Chinese stock market. *Financial Innovation* **5**:1. . [Crossref]
440. Shengying Zhai, Qihui Chen, Wenxin Wang. 2019. What Drives Green Fodder Supply in China?—A Nerlovian Analysis with LASSO Variable Selection. *Sustainability* **11**:23, 6692. [Crossref]
441. Horacio E. Rousseau, Pascual Berrone, Liliana Gelabert. 2019. Localizing Sustainable Development Goals: Nonprofit Density and City Sustainability. *Academy of Management Discoveries* **5**:4, 487-513. [Crossref]
442. Ron Adner, Phanish Puranam, Feng Zhu. 2019. What Is Different About Digital Strategy? From Quantitative to Qualitative Change. *Strategy Science* **4**:4, 253-261. [Crossref]
443. Jorge Mejia, Shawn Mankad, Anandasivam Gopal. 2019. A for Effort? Using the Crowd to Identify Moral Hazard in New York City Restaurant Hygiene Inspections. *Information Systems Research* **30**:4, 1363-1386. [Crossref]
444. Alessandro Roncaglia. The Age of Fragmentation **2**, . [Crossref]
445. Manuel J. García Rodríguez, Vicente Rodríguez Montequín, Francisco Ortega Fernández, Joaquín M. Villanueva Balsera. 2019. Public Procurement Announcements in Spain: Regulations, Data Analysis, and Award Price Estimator Using Machine Learning. *Complexity* **2019**, 1-20. [Crossref]
446. L. Maria Michael Visuwasam, D. Paul Raj. 2019. NMA: integrating big data into a novel mobile application using knowledge extraction for big data analytics. *Cluster Computing* **22**:S6, 14287-14298. [Crossref]
447. David McKenzie, Dario Sansone. 2019. Predicting entrepreneurial success is hard: Evidence from a business plan competition in Nigeria. *Journal of Development Economics* **141**, 102369. [Crossref]
448. Cathy W.S. Chen, Manh Cuong Dong, Nathan Liu, Songsak Sriboonchitta. 2019. Inferences of default risk and borrower characteristics on P2P lending. *The North American Journal of Economics and Finance* **50**, 101013. [Crossref]
449. Ashwin Madhou, Tayushma Sewak, Imad Moosa, Vikash Ramiah. 2019. Forecasting the GDP of a small open developing economy: an application of FAVAR models. *Applied Economics* **1**, 1-12. [Crossref]
450. Nicolas Huck. 2019. Large data sets and machine learning: Applications to statistical arbitrage. *European Journal of Operational Research* **278**:1, 330-342. [Crossref]
451. Katsuyuki Tanaka, Takuo Higashide, Takuji Kinkyo, Shigeyuki Hamori. 2019. ANALYZING INDUSTRY-LEVEL VULNERABILITY BY PREDICTING FINANCIAL BANKRUPTCY. *Economic Inquiry* **57**:4, 2017-2034. [Crossref]
452. Liqian Cai, Arnab Bhattacharjee, Roger Calantone, Taps Maiti. 2019. Variable Selection with Spatially Autoregressive Errors: A Generalized Moments LASSO Estimator. *Sankhya B* **81**:S1, 146-200. [Crossref]
453. Christian Lessmann, Arne Steinkraus. 2019. The geography of natural resources, ethnic inequality and civil conflicts. *European Journal of Political Economy* **59**, 33-51. [Crossref]
454. Clint L.P. Pennings, Jan van Dalen, Laurens Rook. 2019. Coordinating judgmental forecasting: Coping with intentional biases. *Omega* **87**, 46-56. [Crossref]

455. Willem Boshoff, Rossouw van Jaarsveld. 2019. Market Definition Using Consumer Characteristics and Cluster Analysis. *South African Journal of Economics* **87**:3, 302-325. [Crossref]
456. Xia Li, Ruibin Bai, Peer-Olaf Siebers, Christian Wagner. 2019. Travel time prediction in transport and logistics. *VINE Journal of Information and Knowledge Management Systems* **49**:3, 277-306. [Crossref]
457. Susan Athey, Guido W. Imbens. 2019. Machine Learning Methods That Economists Should Know About. *Annual Review of Economics* **11**:1, 685-725. [Crossref]
458. Qiuqin He, Bing Xu. 2019. Determinants of economic growth: A varying-coefficient path identification approach. *Journal of Business Research* **101**, 811-818. [Crossref]
459. Lucy C. Sorensen. 2019. "Big Data" in Educational Administration: An Application for Predicting School Dropout Risk. *Educational Administration Quarterly* **55**:3, 404-446. [Crossref]
460. Mario Molina, Filiz Garip. 2019. Machine Learning for Sociology. *Annual Review of Sociology* **45**:1, 27-45. [Crossref]
461. Yi Ren, Tong Xia, Yong Li, Xiang Chen. 2019. Predicting socio-economic levels of urban regions via offline and online indicators. *PLOS ONE* **14**:7, e0219058. [Crossref]
462. Desamparados Blazquez, Josep Domenech, Jose A. Gil, Ana Pont. 2019. Monitoring e-commerce adoption from online data. *Knowledge and Information Systems* **60**:1, 227-245. [Crossref]
463. Yu-Chien Ko, Yang-Yin Ting, Hamido Fujita. 2019. A visual analytics with evidential inference for big data: case study of chemical vapor deposition in solar company. *Granular Computing* **4**:3, 531-544. [Crossref]
464. Marco Castellani. 2019. Does culture matter for the economic performance of countries? An overview of the literature. *Journal of Policy Modeling* **41**:4, 700-717. [Crossref]
465. Thomas Renault. 2019. L'apport du Big Data (Mégadonnée) et des nouvelles données de la recherche en finance. *Vie & sciences de l'entreprise* **N° 206**:2, 9-19. [Crossref]
466. Thomas Renault. 2019. 2. Données massives et recherche en économie : une (r)évolution ?. *Regards croisés sur l'économie* **n° 23**:2, 32-40. [Crossref]
467. Abigail N. Devereaux. 2019. The nudge wars: A modern socialist calculation debate. *The Review of Austrian Economics* **32**:2, 139-158. [Crossref]
468. Rui Gonçalves, Vitor Miguel Ribeiro, Fernando Lobo Pereira, Ana Paula Rocha. 2019. Deep learning in exchange markets. *Information Economics and Policy* **47**, 38-51. [Crossref]
469. Jeffrey T. Prince. 2019. A paradigm for assessing the scope and performance of predictive analytics. *Information Economics and Policy* **47**, 7-13. [Crossref]
470. Feiyu Hu, Jim Warren, Daniel J. Exeter. 2019. Geography and patient history in long-term lipid lowering medication adherence for primary prevention of cardiovascular disease. *Spatial and Spatio-temporal Epidemiology* **29**, 13-29. [Crossref]
471. Deepak Gupta, Rinkle Rani. 2019. A study of big data evolution and research challenges. *Journal of Information Science* **45**:3, 322-340. [Crossref]
472. Henry E. Brady. 2019. The Challenge of Big Data and Data Science. *Annual Review of Political Science* **22**:1, 297-323. [Crossref]
473. Michael Mayer, Steven C. Bourassa, Martin Hoesli, Donato Scognamiglio. 2019. Estimation and updating methods for hedonic valuation. *Journal of European Real Estate Research* **12**:1, 134-150. [Crossref]
474. Ajay Agrawal, Joshua S. Gans, Avi Goldfarb. 2019. Artificial Intelligence: The Ambiguous Labor Market Impact of Automating Prediction. *Journal of Economic Perspectives* **33**:2, 31-50. [Abstract] [View PDF article] [PDF with links]

475. Susan Athey, Mohsen Bayati, Guido Imbens, Zhaonan Qu. 2019. Ensemble Methods for Causal Effects in Panel Data Settings. *AEA Papers and Proceedings* **109**, 65-70. [Abstract] [View PDF article] [PDF with links]

476. Cetin Ciner. 2019. Do industry returns predict the stock market? A reprise using the random forest. *The Quarterly Review of Economics and Finance* **72**, 152-158. [Crossref]

477. Jong-Min Kim, Hojin Jung. 2019. Predicting bid prices by using machine learning methods. *Applied Economics* **51**:19, 2011-2018. [Crossref]

478. Patrick Dunleavy, Mark Evans. 2019. Australian administrative elites and the challenges of digital-era change. *Journal of Chinese Governance* **4**:2, 181-200. [Crossref]

479. Paolo Brunori, Vito Peragine, Laura Serlenga. 2019. Upward and downward bias when measuring inequality of opportunity. *Social Choice and Welfare* **52**:4, 635-661. [Crossref]

480. Pelin Demirel, Qian Cher Li, Francesco Rentocchini, J. Pawan Tamvada. 2019. Born to be green: new insights into the economics and management of green entrepreneurship. *Small Business Economics* **52**:4, 759-771. [Crossref]

481. O. Rampado, L. Gianusso, C.R. Nava, R. Ropolo. 2019. Analysis of a CT patient dose database with an unsupervised clustering approach. *Physica Medica* **60**, 91-99. [Crossref]

482. Ilias Pasidis. 2019. Congestion by accident? A two-way relationship for highways in England. *Journal of Transport Geography* **76**, 301-314. [Crossref]

483. Susan Athey, Julie Tibshirani, Stefan Wager. 2019. Generalized random forests. *The Annals of Statistics* **47**:2, 1148-1178. [Crossref]

484. Emmanuel Silva, Hossein Hassani, Dag Madsen, Liz Gee. 2019. Googling Fashion: Forecasting Fashion Consumer Behaviour Using Google Trends. *Social Sciences* **8**:4, 111. [Crossref]

485. Colin F. Camerer, Gideon Nave, Alec Smith. 2019. Dynamic Unstructured Bargaining with Private Information: Theory, Experiment, and Outcome Prediction via Machine Learning. *Management Science* **65**:4, 1867-1890. [Crossref]

486. Dario Sansone. 2019. Beyond Early Warning Indicators: High School Dropout and Machine Learning. *Oxford Bulletin of Economics and Statistics* **81**:2, 456-485. [Crossref]

487. Mustafa Yahşi, Ethem Çanakoğlu, Semra Ağralı. 2019. Carbon price forecasting models based on big data analytics. *Carbon Management* **10**:2, 175-187. [Crossref]

488. Wolfram Höpken, Tobias Eberle, Matthias Fuchs, Maria Lexhagen. 2019. Google Trends data for analysing tourists' online search behaviour and improving demand forecasting: the case of Åre, Sweden. *Information Technology & Tourism* **21**:1, 45-62. [Crossref]

489. Chinmay Kakatkar, Martin Spann. 2019. Marketing analytics using anonymized and fragmented tracking data. *International Journal of Research in Marketing* **36**:1, 117-136. [Crossref]

490. Kohei Kawamura, Yohei Kobashi, Masato Shizume, Kozo Ueda. 2019. Strategic central bank communication: Discourse analysis of the Bank of Japan's Monthly Report. *Journal of Economic Dynamics and Control* **100**, 230-250. [Crossref]

491. Koffi Dumor, Li Yao. 2019. Estimating China's Trade with Its Partner Countries within the Belt and Road Initiative Using Neural Network Analysis. *Sustainability* **11**:5, 1449. [Crossref]

492. Nicholas Berente, Stefan Seidel, Hani Safadi. 2019. Research Commentary—Data-Driven Computationally Intensive Theory Development. *Information Systems Research* **30**:1, 50-64. [Crossref]

493. Jinu Lee. 2019. A Neural Network Method for Nonlinear Time Series Analysis. *Journal of Time Series Econometrics* **11**:1. . [Crossref]

494. Krista L. Uggerslev, Frank Bosco. Raising the Ante 745-760. [Crossref]

495. Erik Nelson, John Fitzgerald, Nathan Tefft. 2019. The distributional impact of a green payment policy for organic fruit. *PLOS ONE* **14**:2, e0211199. [Crossref]

496. Ron S. Jarmin. 2019. Evolving Measurement for an Evolving Economy: Thoughts on 21st Century US Economic Statistics. *Journal of Economic Perspectives* **33**:1, 165-184. [Abstract] [View PDF article] [PDF with links]

497. Michael Friendly, Jürgen Symanzik, Ortac Onder. 2019. Visualising the Titanic Disaster. *Significance* **16**:1, 14-19. [Crossref]

498. Yan Liu, Tian Xie. 2019. Machine learning versus econometrics: prediction of box office. *Applied Economics Letters* **26**:2, 124-130. [Crossref]

499. Eli P Fenichel, Yukiko Hashida. 2019. Choices and the value of natural capital. *Oxford Review of Economic Policy* **35**:1, 120-137. [Crossref]

500. Jorge Iván Pérez-Rave, Juan Carlos Correa-Morales, Favián González-Echavarría. 2019. A machine learning approach to big data regression analysis of real estate prices for inferential and predictive purposes. *Journal of Property Research* **36**:1, 59-96. [Crossref]

501. Cinzia Daraio. Econometric Approaches to the Measurement of Research Productivity 633-666. [Crossref]

502. Evgeniy M. Ozhegov, Daria Teterina. Methods of Machine Learning for Censored Demand Prediction 441-446. [Crossref]

503. Pier Francesco De Maria, Leonardo Tomazeli Duarte, Álvaro de Oliveira D'Antona, Cristiano Torezzan. Digital Humanities and Big Microdata: New Approaches for Demographic Research 217-231. [Crossref]

504. Hossein Hassani, Xu Huang, Emmanuel Sirimal Silva. Big Data and Blockchain 7-48. [Crossref]

505. Giorgio Gnecco, Federico Nutarelli. Optimal Trade-Off Between Sample Size and Precision of Supervision for the Fixed Effects Panel Data Model 531-542. [Crossref]

506. Raffaele Dell'Aversana, Edgardo Bucciarelli. Towards a Natural Experiment Leveraging Big Data to Analyse and Predict Users' Behavioural Patterns Within an Online Consumption Setting 103-113. [Crossref]

507. Andrew Haughwout, Benjamin R. Mandel. Empirical analysis of the US consumer 1-21. [Crossref]

508. Thomas B. Götz, Thomas A. Knetsch. 2019. Google data in bridge equation models for German GDP. *International Journal of Forecasting* **35**:1, 45-66. [Crossref]

509. Yang Xiao, De Wang, Jia Fang. 2019. Exploring the disparities in park access through mobile phone data: Evidence from Shanghai, China. *Landscape and Urban Planning* **181**, 80-91. [Crossref]

510. Jessica Lichy, Maher Kachour. Big Data Perception & Usage 89-94. [Crossref]

511. Andres Algaba, David Ardia, Keven Bluteau, Samuel Borms, Kris Boudt. 2019. Econometrics Meets Sentiment: An Overview of Methodology and Applications. *SSRN Electronic Journal* . [Crossref]

512. Jens Prufer, Patricia Prufer. 2019. Data Science for Entrepreneurship Research: Studying Demand Dynamics for Entrepreneurial Skills in the Netherlands. *SSRN Electronic Journal* . [Crossref]

513. Ajay Agrawal, Joshua S. Gans, Avi Goldfarb. 2019. Artificial Intelligence: The Ambiguous Labor Market Impact of Automating Prediction. *SSRN Electronic Journal* . [Crossref]

514. Bo Cowgill, Catherine E. Tucker. 2019. Economics, Fairness and Algorithmic Bias. *SSRN Electronic Journal* **133**. . [Crossref]

515. Laurent Ferrara, Anna Simoni. 2019. When Are Google Data Useful to Nowcast Gdp? An Approach Via Pre-Selection and Shrinkage. *SSRN Electronic Journal* . [Crossref]

516. Jan Abrell, Mirjam Kosch, Sebastian Rausch. 2019. How Effective Was the UK Carbon Tax?—A Machine Learning Approach to Policy Evaluation. *SSRN Electronic Journal* . [Crossref]

517. George G. Judge. 2019. Combining the Information From Econometrics Learning (EL) and Machine Learning (ML). *SSRN Electronic Journal* . [Crossref]
518. John A. Clithero, Jae Joon Lee, Joshua Tasoff. 2019. Supervised Machine Learning for Eliciting Individual Reservation Values. *SSRN Electronic Journal* . [Crossref]
519. Muhammad Zia Hydari, Idris Adjerid, Aaron Striegel. 2019. Health Wearables, Gamification, and Healthful Activity. *SSRN Electronic Journal* **10**. . [Crossref]
520. Danxia Xie, Longtian Zhang, Ke Tang, Zhen Sun. 2019. Data in Growth Model. *SSRN Electronic Journal* . [Crossref]
521. Jochen Hartmann. 2019. Classification Using Decision Tree Ensembles. *SSRN Electronic Journal* . [Crossref]
522. Kenny Ching, Enrico Forti, Evan Rawley. 2019. Extemporaneous Coordination in Specialist Teams: The Familiarity Complementarity. *SSRN Electronic Journal* . [Crossref]
523. Anna Grodecka-Messi, Isaiah Hull. 2019. The Impact of Local Taxes and Public Services on Property Values. *SSRN Electronic Journal* . [Crossref]
524. Kenneth David Strang, Zhaohao Sun. Managerial Controversies in Artificial Intelligence and Big Data Analytics 55-74. [Crossref]
525. Ruben Xing, Jinluan Ren, Jianghua Sun, Lihua Liu. A Critical Review of the Big-Data Paradigm 75-88. [Crossref]
526. Javier Vidal-García, Marta Vidal, Rafael Hernández Barros. Computational Business Intelligence, Big Data, and Their Role in Business Decisions in the Age of the Internet of Things 1048-1067. [Crossref]
527. Javier Vidal-García, Marta Vidal, Rafael Hernández Barros. Business Applications of Big Data 1346-1367. [Crossref]
528. Marvin N. Wright, Inke R. König. 2019. Splitting on categorical predictors in random forests. *PeerJ* **7**, e6339. [Crossref]
529. Marcos López de Prado. 2019. Beyond Econometrics: A Roadmap Towards Financial Machine Learning. *SSRN Electronic Journal* **45**. . [Crossref]
530. Alexandre Rubesam. 2019. Machine Learning Portfolios with Equal Risk Contributions. *SSRN Electronic Journal* **61**. . [Crossref]
531. Mateus Souza. 2019. Predictive Counterfactuals for Event Studies with Staggered Adoption: Recovering Heterogeneous Effects from a Residential Energy Efficiency Program. *SSRN Electronic Journal* **105**. . [Crossref]
532. F. Douglas Foster, Xue-Zhong 'Tony' He, Junqing Kang, Shen Lin. 2019. The Microstructure of Endogenous Liquidity Provision. *SSRN Electronic Journal* **121**. . [Crossref]
533. Nicolas Pröllochs, Stefan Feuerriegel, Dirk Neumann. 2018. Statistical inferences for polarity identification in natural language. *PLOS ONE* **13**:12, e0209323. [Crossref]
534. Fritz Schiltz, Chiara Masci, Tommaso Agasisti, Daniel Horn. 2018. Using regression tree ensembles to model interaction effects: a graphical approach. *Applied Economics* **50**:58, 6341-6354. [Crossref]
535. Fritz Schiltz, Paolo Sestito, Tommaso Agasisti, Kristof De Witte. 2018. The added value of more accurate predictions for school rankings. *Economics of Education Review* **67**, 207-215. [Crossref]
536. Arthur Dyevre, Nicolas Lampach. 2018. The origins of regional integration: Untangling the effect of trade on judicial cooperation. *International Review of Law and Economics* **56**, 122-133. [Crossref]
537. Monica Andini, Emanuele Ciani, Guido de Blasio, Alessio D'Ignazio, Viola Salvestrini. 2018. Targeting with machine learning: An application to a tax rebate program in Italy. *Journal of Economic Behavior & Organization* **156**, 86-102. [Crossref]

538. Eder Johnson de Area Leão Pereira, Marcus Fernandes da Silva, I.C. da Cunha Lima, H.B.B. Pereira. 2018. Trump's Effect on stock markets: A multiscale approach. *Physica A: Statistical Mechanics and its Applications* **512**, 241-247. [Crossref]

539. Otto Kässi, Vili Lehdonvirta. 2018. Online labour index: Measuring the online gig economy for policy and research. *Technological Forecasting and Social Change* **137**, 241-248. [Crossref]

540. Marco Pangallo, Michele Loberto. 2018. Home is where the ad is: online interest proxies housing demand. *EPJ Data Science* **7**:1. . [Crossref]

541. Gary Smith. 2018. Step away from stepwise. *Journal of Big Data* **5**:1. . [Crossref]

542. Alex Coad, Dominik Janzing, Paul Nightingale. 2018. Tools for causal inference from cross-sectional innovation surveys with continuous or discrete variables: Theory and applications. *Cuadernos de Economía* **37**:75, 779-808. [Crossref]

543. Misheck Mutize, Sean Joss Gossel. 2018. Do sovereign credit rating announcements influence excess bond and equity returns in Africa?. *International Journal of Emerging Markets* **13**:6, 1522-1537. [Crossref]

544. Atin Basuchoudhary, James T. Bang. 2018. Predicting Terrorism with Machine Learning: Lessons from "Predicting Terrorism: A Machine Learning Approach". *Peace Economics, Peace Science and Public Policy* **24**:4. . [Crossref]

545. Martin Prause, Jurgen Weigand. 2018. Market Model Benchmark Suite for Machine Learning Techniques. *IEEE Computational Intelligence Magazine* **13**:4, 14-24. [Crossref]

546. Maddalena Cavicchioli, Angeliki Papana, Ariadni Papana Dagiasis, Barbara Pistoresi. 2018. Maximum Likelihood Estimation for the Generalized Pareto Distribution and Goodness-of-Fit Test with Censored Data. *Journal of Modern Applied Statistical Methods* **17**:2. . [Crossref]

547. Bo Xiong, Yuhe Song. 2018. Big Data and Dietary Trend: The Case of Avocado Imports in China. *Journal of International Food & Agribusiness Marketing* **30**:4, 343-354. [Crossref]

548. Ranjith Vijayakumar, Mike W.-L. Cheung. 2018. Replicability of Machine Learning Models in the Social Sciences. *Zeitschrift für Psychologie* **226**:4, 259-273. [Crossref]

549. Gilles Bastin, Paola Tubaro. 2018. Le moment big data des sciences sociales. *Revue française de sociologie* **Vol. 59**:3, 375-394. [Crossref]

550. Julien Boelaert, Étienne Ollion. 2018. The Great Regression. *Revue française de sociologie* **Vol. 59**:3, 475-506. [Crossref]

551. Chiara Masci, Geraint Johnes, Tommaso Agasisti. 2018. Student and school performance across countries: A machine learning approach. *European Journal of Operational Research* **269**:3, 1072-1085. [Crossref]

552. Stelios Michalopoulos, Elias Papaioannou. 2018. Spatial Patterns of Development: A Meso Approach. *Annual Review of Economics* **10**:1, 383-410. [Crossref]

553. Matheus Albergaria, Maria Sylvia Saes. 2018. Measuring externalities in an information commons: the case of libraries. *Journal of Cleaner Production* **192**, 855-863. [Crossref]

554. Carsten Fink, Christian Helmers, Carlos J. Ponce. 2018. Trademark squatters: Theory and evidence from Chile. *International Journal of Industrial Organization* **59**, 340-371. [Crossref]

555. Jiaying Kou, Xiaoming Fu, Jiahua Du, Hua Wang, Geordie Z. Zhang. Understanding Housing Market Behaviour from a Microscopic Perspective 1-9. [Crossref]

556. Ozalp Babaoglu, Alina Sirbu. Cognified Distributed Computing 1180-1191. [Crossref]

557. Nan-Chen Chen, Margaret Drouhard, Rafal Kocielnik, Jina Suh, Cecilia R. Aragon. 2018. Using Machine Learning to Support Qualitative Coding in Social Science. *ACM Transactions on Interactive Intelligent Systems* **8**:2, 1-20. [Crossref]

558. Diego Aparicio, Marcos López de Prado. 2018. How hard is it to pick the right model? MCS and backtest overfitting. *Algorithmic Finance* **7**:1-2, 53-61. [Crossref]
559. Olga Takács, János Vincze. 2018. Bérelőrejelzések – prediktorok és tanulságok. *Közgazdasági Szemle* **65**:6, 592-618. [Crossref]
560. Vincenzo Butticè, Carlotta Orsenigo, Mike Wright. 2018. The effect of information asymmetries on serial crowdfunding and campaign success. *Economia e Politica Industriale* **45**:2, 143-173. [Crossref]
561. Guy David, Philip A. Saynisch, Aaron Smith-McLallen. 2018. The economics of patient-centered care. *Journal of Health Economics* **59**, 60-77. [Crossref]
562. Desamparados Blazquez, Josep Domenech. 2018. Big Data sources and methods for social and economic analyses. *Technological Forecasting and Social Change* **130**, 99-113. [Crossref]
563. Katsuyuki Tanaka, Takuji Kinkyo, Shigeyuki Hamori. 2018. Financial Hazard Map: Financial Vulnerability Predicted by a Random Forests Classification Model. *Sustainability* **10**:5, 1530. [Crossref]
564. Baban Hasnat. 2018. Big Data: An Institutional Perspective on Opportunities and Challenges. *Journal of Economic Issues* **52**:2, 580-588. [Crossref]
565. Benjamin Seligman, Shripad Tuljapurkar, David Rehkopf. 2018. Machine learning approaches to the social determinants of health in the health and retirement study. *SSM - Population Health* **4**, 95-99. [Crossref]
566. Patrick Mikalef, Michail N. Giannakos, Ilias O. Pappas, John Krogstie. The human side of big data: Understanding the skills of the data scientist in education and industry 503-512. [Crossref]
567. Jessica M. Franklin, Chandrasekar Gopalakrishnan, Alexis A. Krumme, Karandeep Singh, James R. Rogers, Joe Kimura, Caroline McKay, Newell E. McElwee, Niteesh K. Choudhry. 2018. The relative benefits of claims and electronic health record data for predicting medication adherence trajectory. *American Heart Journal* **197**, 153-162. [Crossref]
568. Myron P. Gutmann, Emily Klancher Merchant, Evan Roberts. 2018. "Big Data" in Economic History. *The Journal of Economic History* **78**:1, 268-299. [Crossref]
569. Gregorio Caetano, Vikram Maheshri. 2018. Identifying dynamic spillovers of crime with a causal approach to model selection. *Quantitative Economics* **9**:1, 343-394. [Crossref]
570. Mochen Yang, Gediminas Adomavicius, Gordon Burtch, Yuqing Ren. 2018. Mind the Gap: Accounting for Measurement Error and Misclassification in Variables Generated via Data Mining. *Information Systems Research* **29**:1, 4-24. [Crossref]
571. Pilsun Choi, Insik Min. 2018. A Predictive Model for the Employment of College Graduates Using a Machine Learning Approach. *Journal of Vocational Education & Training* **21**:1, 31-54. [Crossref]
572. Keith H Coble, Ashok K Mishra, Shannon Ferrell, Terry Griffin. 2018. Big Data in Agriculture: A Challenge for the Future. *Applied Economic Perspectives and Policy* **40**:1, 79-96. [Crossref]
573. Stephan D. Whitaker. 2018. Big Data versus a survey. *The Quarterly Review of Economics and Finance* **67**, 285-296. [Crossref]
574. Patrick Zschech, Vera Fleißner, Nicole Baumgärtel, Andreas Hilbert. 2018. Data Science Skills and Enabling Enterprise Systems. *HMD Praxis der Wirtschaftsinformatik* **55**:1, 163-181. [Crossref]
575. Desamparados BLAZQUEZ, Josep DOMENECH. 2018. WEB DATA MINING FOR MONITORING BUSINESS EXPORT ORIENTATION. *Technological and Economic Development of Economy* **24**:2, 406-428. [Crossref]
576. Gilbert Saporta. From Conventional Data Analysis Methods to Big Data Analytics 27-41. [Crossref]
577. Rimvydas Skyrius, Gintarė Giriūnienė, Igor Katin, Michail Kazimianec, Raimundas Žilinskas. The Potential of Big Data in Banking 451-486. [Crossref]
578. Chaitanya Baru. Data in the 21st Century 3-17. [Crossref]

579. Yong Yoon. Spatial Choice Modeling Using the Support Vector Machine (SVM): Characterization and Prediction 767-778. [Crossref]
580. Wolfram Höpken, Tobias Eberle, Matthias Fuchs, Maria Lexhagen. Search Engine Traffic as Input for Predicting Tourist Arrivals 381-393. [Crossref]
581. Shu-Heng Chen, Ye-Rong Du, Ying-Fang Kao, Ragupathy Venkatachalam, Tina Yu. On Complex Economic Dynamics: Agent-Based Computational Modeling and Beyond 1-14. [Crossref]
582. Thomas K. Bauer, Phillip Breidenbach, Sandra Schaffner. Big Data in der wirtschaftswissenschaftlichen Forschung 129-148. [Crossref]
583. Cinzia Daraio. Nonparametric Methods and Higher Education 1-7. [Crossref]
584. Peng Ye, Julian Qian, Jieying Chen, Chen-hung Wu, Yitong Zhou, Spencer De Mars, Frank Yang, Li Zhang. Customized Regression Model for Airbnb Dynamic Pricing 932-940. [Crossref]
585. Thiago Gonçalves dos Santos Martins, Ana Luiza Fontes de Azevedo Costa, Thomaz Gonçalves dos Santos Martins. 2018. Big Data use in medical research. *Einstein (São Paulo)* **16**:3. . [Crossref]
586. Andre Boik. 2018. Prediction and Identification in Two-Sided Markets. *SSRN Electronic Journal* . [Crossref]
587. Lucie Martin-Bonnel de Longchamp, Nicolas Lampach, Ludovic Parisot. 2018. How Cognitive Biases Affect Energy Savings in Low Energy Buildings. *SSRN Electronic Journal* . [Crossref]
588. Jens Prufer, Patricia Prufer. 2018. Data Science for Institutional and Organizational Economics. *SSRN Electronic Journal* . [Crossref]
589. Phanish Puranam, Yash Raj Shrestha, Vivianna Fang He, Georg von Krogh. 2018. Algorithmic Induction Through Machine Learning: Opportunities for Management and Organization Research. *SSRN Electronic Journal* . [Crossref]
590. Akos Lada, Diego Aparicio, Michael Bailey. 2018. Predicting Heterogeneous Treatment Effects in Ranking Systems. *SSRN Electronic Journal* . [Crossref]
591. Silvia Emili, Attilio Gardini. 2018. High Spatial and Temporal Detail in Timely Prediction of Tourism Demand. *SSRN Electronic Journal* . [Crossref]
592. Santiago Carbo-Valverde, Pedro Cuadros-Solas, Francisco Rodriguez-Fernandez. 2018. How Do Bank Customers Go Digital? A Random Forest Approach. *SSRN Electronic Journal* . [Crossref]
593. Jeffrey Prince. 2018. A Paradigm for Assessing the Scope and Performance of Predictive Analytics. *SSRN Electronic Journal* . [Crossref]
594. Matthew Grennan, Kyle Myers, Ashley Teres Swanson, Aaron Chatterji. 2018. Physician-Industry Interactions: Persuasion and Welfare. *SSRN Electronic Journal* **106**. . [Crossref]
595. Monika Glavina. 2018. 'To Submit or Not to Submit – That Is the (Preliminary) Question': Explaining National Judges' Reluctance to Participate in the Preliminary Ruling Procedure. *SSRN Electronic Journal* . [Crossref]
596. Roberto Casarin, Fausto Corradin, Francesco Ravazzolo, Domenico Sartore. 2018. A Scoring Rule for Factor and Autoregressive Models Under Misspecification. *SSRN Electronic Journal* . [Crossref]
597. Benjamin Bluhm. 2018. Time Series Econometrics at Scale: A Practical Guide to Parallel Computing in (Py)Spark. *SSRN Electronic Journal* . [Crossref]
598. Marco Pangallo, Michele Loberto. 2018. Home Is Where the Ad Is: Online Interest Proxies Housing Demand. *SSRN Electronic Journal* . [Crossref]
599. Evgeniy Ozhegov, Daria Teterina. 2018. The Ensemble Method for Censored Demand Prediction. *SSRN Electronic Journal* . [Crossref]
600. Julian TszKin Chan, Weifeng Zhong. 2018. Reading China: Predicting Policy Change with Machine Learning. *SSRN Electronic Journal* **130**. . [Crossref]

601. Chinmay Kakatkar, Volker Bilgram, Johann Füller. 2018. Innovation Analytics: Leveraging Artificial Intelligence in the Innovation Process. *SSRN Electronic Journal* . [Crossref]
602. Michael Mayer, Steven C. Bourassa, Martin Edward Ralph Hoesli, Donato Flavio Scognamiglio. 2018. Estimation and Updating Methods for Hedonic Valuation. *SSRN Electronic Journal* . [Crossref]
603. Roberto Moro Visconti, Giuseppe Montesi, Giovanni Papiro. 2018. Big data-driven stochastic business planning and corporate valuation. *Corporate Ownership and Control* **15**:3-1, 189-204. [Crossref]
604. Nicholas Hallman, Antonis Kartapanis, Jaime J. Schmidt. 2018. Using Sec Edgar Views to Measure Competition Among Big 4 Auditors. *SSRN Electronic Journal* **21**. . [Crossref]
605. Xuezhong He, Shen Lin. 2018. Rational Learning and Trading Behavior in Limit Order Markets. *SSRN Electronic Journal* **104**. . [Crossref]
606. Oz Shy. 2018. Alternative Methods for Studying Consumer Payment Choice. *SSRN Electronic Journal* **105**. . [Crossref]
607. Sumit Agarwal, Long Wang, Yang Yang. 2018. Blessing in Disguise? Environmental Shocks and Performance Enhancement. *SSRN Electronic Journal* **85**. . [Crossref]
608. Tommaso Tani. L'incidenza dei big data e del machine learning sui principi alla base del Regolamento Europeo per la tutela dei dati personali (2016/679/UE) e proposte per una nuova normativa in tema di privacy 35-65. [Crossref]
609. Raghavendra Rau. 2017. Social networks and financial outcomes. *Current Opinion in Behavioral Sciences* **18**, 75-78. [Crossref]
610. Chris Schilling, Josh Knight, Duncan Mortimer, Dennis Petrie, Philip Clarke, John Chalmers, Andrew Kerr, Rod Jackson. 2017. Australian general practitioners initiate statin therapy primarily on the basis of lipid levels; New Zealand general practitioners use absolute risk. *Health Policy* **121**:12, 1233-1239. [Crossref]
611. Lei Dong, Sicong Chen, Yunsheng Cheng, Zhengwei Wu, Chao Li, Haishan Wu. 2017. Measuring economic activity in China with mobile big data. *EPJ Data Science* **6**:1. . [Crossref]
612. Thiago Gonçalves dos Santos Martins, Ana Luiza Fontes de Azevedo Costa. 2017. A new way to communicate science in the era of Big Data and citizen science. *Einstein (São Paulo)* **15**:4, 523-523. [Crossref]
613. Eric Zheng, Yong Tan, Paulo Goes, Ramnath Chellappa, D.J. Wu, Michael Shaw, Olivia Sheng, Alok Gupta. 2017. When Econometrics Meets Machine Learning. *Data and Information Management* **1**:2, 75-83. [Crossref]
614. János Vincze. 2017. Információ és tudás. A big data egyes hatásai a közgazdaságtanra. *Közgazdasági Szemle* **64**:11, 1148-1159. [Crossref]
615. Paola D'Orazio. 2017. Big data and complexity: Is macroeconomics heading toward a new paradigm?. *Journal of Economic Methodology* **24**:4, 410-429. [Crossref]
616. Shu-Heng Chen, Ragupathy Venkatachalam. 2017. Agent-based modelling as a foundation for big data. *Journal of Economic Methodology* **24**:4, 362-383. [Crossref]
617. Teck-Hua Ho, Noah Lim, Sadat Reza, Xiaoyu Xia. 2017. OM Forum—Causal Inference Models in Operations Management. *Manufacturing & Service Operations Management* **19**:4, 509-525. [Crossref]
618. Ernesto D'Avanzo, Giovanni Pilato, Miltiadis Lytras. 2017. Using Twitter sentiment and emotions analysis of Google Trends for decisions making. *Program* **51**:3, 322-350. [Crossref]
619. Takayuki Morimoto, Yoshinori Kawasaki. 2017. Forecasting Financial Market Volatility Using a Dynamic Topic Model. *Asia-Pacific Financial Markets* **24**:3, 149-167. [Crossref]
620. Derek Messacar. 2017. Big Tax Data and Economic Analysis: Effects of Personal Income Tax Reassessments and Delayed Tax Filing. *Canadian Public Policy* **43**:3, 261-283. [Crossref]

621. Rajesh Chandy, Magda Hassan, Prokriti Mukherji. 2017. Big Data for Good: Insights from Emerging Markets*. *Journal of Product Innovation Management* **34**:5, 703-713. [Crossref]
622. Silvia Mendolia, Peter Siminski. 2017. Is education the mechanism through which family background affects economic outcomes? A generalised approach to mediation analysis. *Economics of Education Review* **59**, 1-12. [Crossref]
623. Yuh-Jong Hu, Shu-Wei Huang. Challenges of automated machine learning on causal impact analytics for policy evaluation 1-6. [Crossref]
624. Carlos Tapia, Beñat Abajo, Efren Feliu, Maddalen Mendizabal, José Antonio Martinez, J. German Fernández, Txomin Laburu, Adelaida Lejarazu. 2017. Profiling urban vulnerabilities to climate change: An indicator-based vulnerability assessment for European cities. *Ecological Indicators* **78**, 142-155. [Crossref]
625. Dong-Jin Pyo. 2017. Can Big Data Help Predict Financial Market Dynamics?: Evidence from the Korean Stock Market. *East Asian Economic Review* **21**:2, 147-165. [Crossref]
626. Jessica Lichy, Maher Kachour, Tatiana Khvatova. 2017. Big Data is watching YOU: opportunities and challenges from the perspective of young adult consumers in Russia. *Journal of Marketing Management* **33**:9-10, 719-741. [Crossref]
627. Adam Nowak, Patrick Smith. 2017. Textual Analysis in Real Estate. *Journal of Applied Econometrics* **32**:4, 896-918. [Crossref]
628. Christopher Krauss, Xuan Anh Do, Nicolas Huck. 2017. Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500. *European Journal of Operational Research* **259**:2, 689-702. [Crossref]
629. Sendhil Mullainathan, Jann Spiess. 2017. Machine Learning: An Applied Econometric Approach. *Journal of Economic Perspectives* **31**:2, 87-106. [Abstract] [View PDF article] [PDF with links]
630. Aloisio Dourado, Rommel N. Carvalho, Gustavo C. G. van Erven. Brazil's Bolsa Familia and young adult workers: A parallel RDD approach to large datasets 17-24. [Crossref]
631. Michael Creel. 2017. Neural nets for indirect inference. *Econometrics and Statistics* **2**, 36-49. [Crossref]
632. Xin Li, Bing Pan, Rob Law, Xiankai Huang. 2017. Forecasting tourism demand with composite search index. *Tourism Management* **59**, 57-66. [Crossref]
633. Patricia Kuzmenko FURLAN, Fernando José Barbin LAURINDO. 2017. Agrupamentos epistemológicos de artigos publicados sobre big data analytics. *Transinformação* **29**:1, 91-100. [Crossref]
634. Yael Grushka-Cockayne, Victor Richmond R. Jose, Kenneth C. Lichtendahl. 2017. Ensembles of Overfit and Overconfident Forecasts. *Management Science* **63**:4, 1110-1130. [Crossref]
635. Benjamin David. 2017. Computer technology and probable job destructions in Japan: An evaluation. *Journal of the Japanese and International Economies* **43**, 77-87. [Crossref]
636. Sebastian Tillmanns, Frenkel Ter Hofstede, Manfred Krafft, Oliver Goetz. 2017. How to Separate the Wheat from the Chaff: Improved Variable Selection for New Customer Acquisition. *Journal of Marketing* **81**:2, 99-113. [Crossref]
637. Greg Distelhorst, Jens Hainmueller, Richard M. Locke. 2017. Does Lean Improve Labor Standards? Management and Social Performance in the Nike Supply Chain. *Management Science* **63**:3, 707-728. [Crossref]
638. Johan L. Perols, Robert M. Bowen, Carsten Zimmermann, Basamba Samba. 2017. Finding Needles in a Haystack: Using Data Analytics to Improve Fraud Prediction. *The Accounting Review* **92**:2, 221-245. [Crossref]
639. Felix Ward. 2017. Spotting the Danger Zone: Forecasting Financial Crises With Classification Tree Ensembles and Many Predictors. *Journal of Applied Econometrics* **32**:2, 359-378. [Crossref]

640. Jayson L. Lusk. 2017. Consumer Research with Big Data: Applications from the Food Demand Survey (FooDS). *American Journal of Agricultural Economics* **99**:2, 303-320. [Crossref]

641. Omar A. Guerrero, Eduardo López. 2017. Understanding Unemployment in the Era of Big Data: Policy Informed by Data-Driven Theory. *Policy & Internet* **9**:1, 28-54. [Crossref]

642. Thomas Pave Sohnesen, Niels Stender. 2017. Is Random Forest a Superior Methodology for Predicting Poverty? An Empirical Assessment. *Poverty & Public Policy* **9**:1, 118-133. [Crossref]

643. Zaheer Khan, Tim Vorley. 2017. Big data text analytics: an enabler of knowledge management. *Journal of Knowledge Management* **21**:1, 18-34. [Crossref]

644. Chris Schilling, Duncan Mortimer, Kim Dalziel. 2017. Using CART to Identify Thresholds and Hierarchies in the Determinants of Funding Decisions. *Medical Decision Making* **37**:2, 173-182. [Crossref]

645. Kenneth David Strang, Zhaohao Sun. 2017. Analyzing Relationships in Terrorism Big Data Using Hadoop and Statistics. *Journal of Computer Information Systems* **57**:1, 67-75. [Crossref]

646. Haiyan Song, Han Liu. Predicting Tourist Demand Using Big Data 13-29. [Crossref]

647. Carlianne Patrick, Amanda Ross, Heather Stephens. Designing Policies to Spur Economic Growth: How Regional Scientists Can Contribute to Future Policy Development and Evaluation 119-133. [Crossref]

648. Neelam Younas, Zahid Asghar, Muhammad Qayyum, Fazlullah Khan. Education and Socio Economic Factors Impact on Earning for Pakistan - A Bigdata Analysis 215-223. [Crossref]

649. Anne Fleur van Veenstra, Bas Kotterink. Data-Driven Policy Making: The Policy Lab Approach 100-111. [Crossref]

650. Atin Basuchoudhary, James T. Bang, Tinni Sen. Why This Book? 1-6. [Crossref]

651. Khyati Ahlawat, Amit Prakash Singh. A Novel Hybrid Technique for Big Data Classification Using Decision Tree Learning 118-128. [Crossref]

652. Xiangjun Meng, Liang Chen, Yidong Li. A Parallel Clustering Algorithm for Power Big Data Analysis 533-540. [Crossref]

653. Alexander Peysakhovich, Jeffrey Naecker. 2017. Using methods from machine learning to evaluate behavioral models of choice under risk and ambiguity. *Journal of Economic Behavior & Organization* **133**, 373-384. [Crossref]

654. D. Daniel Sokol, Roisin Comerford. Does Antitrust Have a Role to Play in Regulating Big Data? 293-316. [Crossref]

655. Yu Hou, Artur Hugon, Matthew R. Lyle, Seth Pruitt. 2017. Macroeconomic News in the Cross Section of Asset Growth. *SSRN Electronic Journal* . [Crossref]

656. Scott Kostyshak. 2017. Non-Parametric Testing of U-Shapes, with an Application to the Midlife Satisfaction Dip. *SSRN Electronic Journal* . [Crossref]

657. Kweku A. Opoku-Agyemang. 2017. Priming Human-Computer Interactions: Experimental Evidence from Economic Development Mobile Surveys. *SSRN Electronic Journal* **47**. . [Crossref]

658. Max Biggs, Rim Hariss. 2017. Optimizing Objective Functions Determined from Random Forests. *SSRN Electronic Journal* . [Crossref]

659. Dong-Jin Pyo. 2017. Can Big Data Help Predict Financial Market Dynamics?: Evidence from the Korean Stock Market. *SSRN Electronic Journal* . [Crossref]

660. Mike Horia Teodorescu. 2017. Machine Learning Methods for Strategy Research. *SSRN Electronic Journal* . [Crossref]

661. Chiranjit Chakraborty, Andreas Joseph. 2017. Machine Learning at Central Banks. *SSRN Electronic Journal* . [Crossref]

662. Nicolas Lampach, Arthur Dyevre. 2017. The Origins of Regional Integration: Untangling the Effect of Trade on Judicial Cooperation. *SSRN Electronic Journal* . [Crossref]
663. Guy David, Phil Saynisch, Aaron Smith-McLallen. 2017. The Economics of Patient-Centered Care. *SSRN Electronic Journal* **27**. . [Crossref]
664. Diego Aparicio, Marcos Lopez de Prado. 2017. How Hard Is It to Pick the Right Model?. *SSRN Electronic Journal* . [Crossref]
665. Daniel Fricke. 2017. Financial Crisis Prediction: A Model Comparison. *SSRN Electronic Journal* . [Crossref]
666. Andreas Fuster, Paul Goldsmith-Pinkham, Tarun Ramadorai, Ansgar Walther. 2017. Predictably Unequal? The Effects of Machine Learning on Credit Markets. *SSRN Electronic Journal* . [Crossref]
667. Monica Andini, Emanuele Ciani, Guido de Blasio, Alessio D'Ignazio, Viola Salvestrini. 2017. Targeting Policy-Compliers with Machine Learning: An Application to a Tax Rebate Programme in Italy. *SSRN Electronic Journal* . [Crossref]
668. Javier Vidal-García, Marta Vidal, Rafael Hernandez Barros. Computational Business Intelligence, Big Data, and Their Role in Business Decisions in the Age of the Internet of Things 249-268. [Crossref]
669. Javier Vidal-García, Marta Vidal, Rafael Hernández Barros. Business Applications of Big Data 104-125. [Crossref]
670. Amy K. Johnson, Tarek Mikati, Supriya D. Mehta. 2016. Examining the themes of STD-related Internet searches to increase specificity of disease forecasting using Internet search terms. *Scientific Reports* **6**:1. . [Crossref]
671. Jacques Bughin. 2016. Reaping the benefits of big data in telecom. *Journal of Big Data* **3**:1. . [Crossref]
672. Dave Donaldson, Adam Storeygard. 2016. The View from Above: Applications of Satellite Data in Economics. *Journal of Economic Perspectives* **30**:4, 171-198. [Abstract] [View PDF article] [PDF with links]
673. R. Rajesh. 2016. Forecasting supply chain resilience performance using grey prediction. *Electronic Commerce Research and Applications* **20**, 42-58. [Crossref]
674. Christian Pierdzioch, Marian Risse, Sebastian Rohloff. 2016. Are precious metals a hedge against exchange-rate movements? An empirical exploration using bayesian additive regression trees. *The North American Journal of Economics and Finance* **38**, 27-38. [Crossref]
675. Nuha Almoqren, Mohammed Altayar. The motivations for big data mining technologies adoption in saudi banks 1-8. [Crossref]
676. Michel Wedel, P.K. Kannan. 2016. Marketing Analytics for Data-Rich Environments. *Journal of Marketing* **80**:6, 97-121. [Crossref]
677. Uwe Deichmann, Aparajita Goyal, Deepak Mishra. 2016. Will digital technologies transform agriculture in developing countries?. *Agricultural Economics* **47**:S1, 21-33. [Crossref]
678. Linden McBride, Austin Nichols. 2016. Retooling Poverty Targeting Using Out-of-Sample Validation and Machine Learning. *The World Bank Economic Review* **2**, lhw056. [Crossref]
679. Ben Vinod. 2016. Big Data in the travel marketplace. *Journal of Revenue and Pricing Management* **15**:5, 352-359. [Crossref]
680. Jaideep Ghosh. 2016. Big Data Analytics: A Field of Opportunities for Information Systems and Technology Researchers. *Journal of Global Information Technology Management* **19**:4, 217-222. [Crossref]
681. Stefan Feuerriegel. 2016. Decision support in healthcare: determining provider influence on treatment outcomes with robust risk adjustment. *Journal of Decision Systems* **25**:4, 371-390. [Crossref]

682. Matthias Duschl. 2016. Firm dynamics and regional resilience: an empirical evolutionary perspective. *Industrial and Corporate Change* **25**:5, 867-883. [Crossref]

683. Reinout Heijungs, Patrik Henriksson, Jeroen Guinée. 2016. Measures of Difference and Significance in the Era of Computer Simulations, Meta-Analysis, and Big Data. *Entropy* **18**:10, 361. [Crossref]

684. Gerard George, Ernst C. Osinga, Dovev Lavie, Brent A. Scott. 2016. Big Data and Data Science Methods for Management Research. *Academy of Management Journal* **59**:5, 1493-1507. [Crossref]

685. Alison L. Bailey, Anne Blackstock-Bernstein, Eve Ryan, Despina Pitsoulakis. DATA MINING WITH NATURAL LANGUAGE PROCESSING AND CORPUS LINGUISTICS 255-275. [Crossref]

686. P. Racca, R. Casarin, F. Squazzoni, P. Dondio. 2016. Resilience of an online financial community to market uncertainty shocks during the recent financial crisis. *Journal of Computational Science* **16**, 190-199. [Crossref]

687. Michael Mann, Eli Melaas, Arun Malik. 2016. Using VIIRS Day/Night Band to Measure Electricity Supply Reliability: Preliminary Results from Maharashtra, India. *Remote Sensing* **8**:9, 711. [Crossref]

688. Hu Shuijing. Big Data Analytics: Key Technologies and Challenges 141-145. [Crossref]

689. Benjamin F. Mundell, Hilal Maradit Kremers, Sue Visscher, Kurtis M. Hoppe, Kenton R. Kaufman. 2016. Predictors of Receiving a Prosthesis for Adults With Above-Knee Amputations in a Well-Defined Population. *PM&R* **8**:8, 730-737. [Crossref]

690. Michael Peneder. 2016. Competitiveness and industrial policy: from rationalities of failure towards the ability to evolve. *Cambridge Journal of Economics* **11**, bew025. [Crossref]

691. Gérard Biau, Erwan Scornet. 2016. A random forest guided tour. *TEST* **25**:2, 197-227. [Crossref]

692. Joyce P Jacobsen, Laurence M Levin, Zachary Tausanovitch. 2016. Comparing Standard Regression Modeling to Ensemble Modeling: How Data Mining Software Can Improve Economists' Predictions. *Eastern Economic Journal* **42**:3, 387-398. [Crossref]

693. Dror Etzion, J. Alberto Aragon-Correa. 2016. Big Data, Management, and Sustainability. *Organization & Environment* **29**:2, 147-155. [Crossref]

694. Julia Lane. 2016. BIG DATA FOR PUBLIC POLICY: THE QUADRUPLE HELIX. *Journal of Policy Analysis and Management* **35**:3, 708-715. [Crossref]

695. William G. Bostic Jr., Ron S. Jarmin, Brian Moyer. 2016. Modernizing Federal Economic Statistics. *American Economic Review* **106**:5, 161-164. [Abstract] [View PDF article] [PDF with links]

696. Alberto Cavallo, Roberto Rigobon. 2016. The Billion Prices Project: Using Online Prices for Measurement and Research. *Journal of Economic Perspectives* **30**:2, 151-178. [Abstract] [View PDF article] [PDF with links]

697. Paul Smith. 2016. Google's MIDAS Touch: Predicting UK Unemployment with Internet Search Data. *Journal of Forecasting* **35**:3, 263-284. [Crossref]

698. Nalan Baştürk, Roberto Casarin, Francesco Ravazzolo, Herman van Dijk. 2016. Computational Complexity and Parallelization in Bayesian Econometric Analysis. *Econometrics* **4**:1, 9. [Crossref]

699. Chris Schilling, Duncan Mortimer, Kim Dalziel, Emma Heeley, John Chalmers, Philip Clarke. 2016. Using Classification and Regression Trees (CART) to Identify Prescribing Thresholds for Cardiovascular Disease. *PharmacoEconomics* **34**:2, 195-205. [Crossref]

700. Alexander T. Janke, Daniel L. Overbeek, Keith E. Kocher, Phillip D. Levy. 2016. Exploring the Potential of Predictive Analytics and Big Data in Emergency Care. *Annals of Emergency Medicine* **67**:2, 227-236. [Crossref]

701. Jessica M. Franklin, William H. Shrank, Joyce Lii, Alexis K. Krumme, Olga S. Matlin, Troyen A. Brennan, Niteesh K. Choudhry. 2016. Observing versus Predicting: Initial Patterns of Filling Predict

Long-Term Adherence More Accurately Than High-Dimensional Modeling Techniques. *Health Services Research* **51**:1, 220-239. [Crossref]

702. . Spotlight 1: How the internet promotes development 42-46. [Crossref]
703. Karsten Luebke, Joachim Rojahn. Firm-Specific Determinants on Dividend Changes: Insights from Data Mining 335-344. [Crossref]
704. Ali Emrouznejad, Marianna Marra. Big Data: Who, What and Where? Social, Cognitive and Journals Map of Big Data Publications with Focus on Optimization 1-16. [Crossref]
705. Richard W. Evans, Kenneth L. Judd, Kramer Quist. Big Data Techniques as a Solution to Theory Problems 219-231. [Crossref]
706. Luca Onorante, Adrian E. Raftery. 2016. Dynamic model averaging in large model spaces using dynamic Occam#s window. *European Economic Review* **81**, 2-14. [Crossref]
707. Michael S. Hand, Matthew P. Thompson, David E. Calkin. 2016. Examining heterogeneity and wildfire management expenditures using spatially and temporally descriptive data. *Journal of Forest Economics* **22**, 80-102. [Crossref]
708. Scott McQuade, Claire Monteleoni. Online Learning of Volatility from Multiple Option Term Lengths 1-3. [Crossref]
709. Erik Nelson, Clare Bates Congdon. 2016. Measuring the relative importance of different agricultural inputs to global and regional crop yield growth since 1975. *F1000Research* **5**, 2930. [Crossref]
710. Eike Emrich, Christian Pierdzioch. 2016. Public Goods, Private Consumption, and Human Capital: Using Boosted Regression Trees to Model Volunteer Labour Supply. *Review of Economics* **67**:3. . [Crossref]
711. Leroi Raputsoane. 2016. Real Effective Exchange Rates Comovements, Common Factors and the South African Currency. *SSRN Electronic Journal* . [Crossref]
712. Omar A. Guerrero, Eduardo Lopez. 2016. Understanding Unemployment in the Era of Big Data: Policy Informed by Data-Driven Theory. *SSRN Electronic Journal* . [Crossref]
713. Kohei Kawamura, Yohei Kobashi, Masato Shizume. 2016. Strategic Central Bank Communication: Discourse and Game-Theoretic Analyses of the Bank of Japan's Monthly Report. *SSRN Electronic Journal* . [Crossref]
714. Georg von Graevenitz, Christian Helmers, Valentine Millot, Oliver Turnbull. 2016. Does Online Search Predict Sales? Evidence from Big Data for Car Markets in Germany and the UK. *SSRN Electronic Journal* . [Crossref]
715. Inna Grinis. 2016. The STEM Requirements of 'Non-STEM' Jobs: Evidence from UK Online Vacancy Postings and Implications for Skills & Knowledge Shortages. *SSRN Electronic Journal* . [Crossref]
716. Serena Ng. 2016. Opportunities and Challenges: Lessons from Analyzing Terabytes of Scanner Data. *SSRN Electronic Journal* . [Crossref]
717. Leif Anders Thorsrud. 2016. Nowcasting Using News Topics. Big Data versus Big Bank. *SSRN Electronic Journal* . [Crossref]
718. José Luis Gómez-Barroso, Juan Ángel Ruiz. Behavioural Targeting in the Mobile Ecosystem 44-57. [Crossref]
719. Andrew Tiffin. 2016. Seeing in the Dark: A Machine-Learning Approach to Nowcasting in Lebanon. *IMF Working Papers* **16**:56, 1. [Crossref]
720. Jurgen A. Doornik, David F. Hendry. 2015. Statistical model selection with “Big Data”. *Cogent Economics & Finance* **3**:1. . [Crossref]

721. Qing-Ting Zhang, Yuan Liu, Wen Zhou, Zhou-Wang Yang. 2015. A Sequential Regression Model for Big Data with Attributive Explanatory Variables. *Journal of the Operations Research Society of China* **3**:4, 475-488. [Crossref]
722. robert neumann, peter graeff. 2015. quantitative approaches to comparative analyses: data properties and their implications for theory, measurement and modelling. *European Political Science* **14**:4, 385-393. [Crossref]
723. Ben Vinod. 2015. The expanding role of revenue management in the airline industry. *Journal of Revenue and Pricing Management* **14**:6, 391-399. [Crossref]
724. Alan Schwartz, Robert E. Scott. 2015. Third-Party Beneficiaries and Contractual Networks. *Journal of Legal Analysis* **7**:2, 325-361. [Crossref]
725. Marco Capuccini, Lars Carlsson, Ulf Norinder, Ola Spjuth. Conformal Prediction in Spark: Large-Scale Machine Learning with Confidence 61-67. [Crossref]
726. Julia I. Lane, Jason Owen-Smith, Rebecca F. Rosen, Bruce A. Weinberg. 2015. New linked data on research investments: Scientific workforce, productivity, and public value. *Research Policy* **44**:9, 1659-1671. [Crossref]
727. Max Nathan, Anna Rosso. 2015. Mapping digital businesses with big data: Some early findings from the UK. *Research Policy* **44**:9, 1714-1733. [Crossref]
728. Maryann Feldman, Martin Kenney, Francesco Lissoni. 2015. The New Data Frontier. *Research Policy* **44**:9, 1629-1632. [Crossref]
729. Imanol Arrieta-ibarra, Ignacio N. Lobato. 2015. Testing for Predictability in Financial Returns Using Statistical Learning Procedures. *Journal of Time Series Analysis* **36**:5, 672-686. [Crossref]
730. David H. Autor. 2015. Why Are There Still So Many Jobs? The History and Future of Workplace Automation. *Journal of Economic Perspectives* **29**:3, 3-30. [Abstract] [View PDF article] [PDF with links]
731. Yan Chen, Joseph Konstan. 2015. Online field experiments: a selective survey of methods. *Journal of the Economic Science Association* **1**:1, 29-42. [Crossref]
732. Levi Boxell. 2015. K-fold Cross-Validation and the Gravity Model of Bilateral Trade. *Atlantic Economic Journal* **43**:2, 289-300. [Crossref]
733. Hallie Eakin, Kirsten Appendini, Stuart Sweeney, Hugo Perales. 2015. Correlates of Maize Land and Livelihood Change Among Maize Farming Households in Mexico. *World Development* **70**, 78-91. [Crossref]
734. Patrick Bajari, Denis Nekipelov, Stephen P. Ryan, Miaoyu Yang. 2015. Machine Learning Methods for Demand Estimation. *American Economic Review* **105**:5, 481-485. [Abstract] [View PDF article] [PDF with links]
735. David Bholat. 2015. Big Data and central banks. *Big Data & Society* **2**:1, 205395171557946. [Crossref]
736. Hossein Hassani, Emmanuel Sirimal Silva. 2015. Forecasting with Big Data: A Review. *Annals of Data Science* **2**:1, 5-19. [Crossref]
737. Jorge Guzman, Scott Stern. 2015. Where is Silicon Valley?. *Science* **347**:6222, 606-609. [Crossref]
738. Nicole Ludwig, Stefan Feuerriegel, Dirk Neumann. 2015. Putting Big Data analytics to work: Feature selection for forecasting electricity prices using the LASSO and random forests. *Journal of Decision Systems* **24**:1, 19-36. [Crossref]
739. Barbara Dinter, David Douglas, Roger H. L. Chiang, Francesco Mari, Sudha Ram, Detlef Schoder. Big Data Panel at SIGDSS Pre-ICIS Conference 2013: A Swiss-Army Knife? The Profile of a Data Scientist 7-11. [Crossref]

740. Thach V. Bui, Thuc D. Nguyen, Noboru Sonehara, Isao Echizen. Tradeoff Between the Price of Distributing a Database and Its Collusion Resistance Based on Concatenated Codes 163-182. [Crossref]
741. Vlad Diaconita. 2015. Processing unstructured documents and social media using Big Data techniques. *Economic Research-Ekonomska Istraživanja* **28**:1, 981-993. [Crossref]
742. Alex Street, Thomas A. Murray, John Blitzer, Rajan S. Patel. 2015. Estimating Voter Registration Deadline Effects with Web Search Data. *Political Analysis* **23**:2, 225-241. [Crossref]
743. Lilli Japec, Frauke Kreuter, Marcus Berg, Paul Biemer, Paul Decker, Cliff Lampe, Julia Lane, Cathy O'Neil, Abe Usher. 2015. Big Data in Survey Research. *Public Opinion Quarterly* **79**:4, 839-880. [Crossref]
744. Kaushik Basu, Andrew Foster. 2015. Development Economics and Method: A Quarter Century of ABCDE. *The World Bank Economic Review* **29**:suppl 1, S2-S8. [Crossref]
745. Alexander Peysakhovich, Jeffrey Naecker. 2015. Machine Learning and Behavioral Economics: Evaluating Models of Choice Under Risk and Ambiguity. *SSRN Electronic Journal* . [Crossref]
746. Johan Perols, Robert M. Bowen, Carsten Zimmermann, Basamba Samba. 2015. Finding Needles in a Haystack: Using Data Analytics to Improve Fraud Prediction. *SSRN Electronic Journal* . [Crossref]
747. Paul Smith. 2015. Predicting UK Unemployment with Internet Search and Survey Data. *SSRN Electronic Journal* . [Crossref]
748. Ananya Sen, Pinar Yildirim. 2015. Clicks and Editorial Decisions: How Does Popularity Shape Online News Coverage?. *SSRN Electronic Journal* . [Crossref]
749. Roberto Casarin, Stefano Grassi, Francesco Ravazzolo, H. K. van Dijk. 2015. Dynamic Predictive Density Combinations for Large Data Sets in Economics and Finance. *SSRN Electronic Journal* . [Crossref]
750. Christian Pierdzioch, Marian Risse, Sebastian Rohloff. 2015. Are Precious Metals a Hedge Against Exchange-Rate Movements? An Empirical Exploration Using Bayesian Additive Regression Trees. *SSRN Electronic Journal* . [Crossref]
751. Anja Lambrecht, Catherine Tucker. 2015. Can Big Data Protect a Firm from Competition?. *SSRN Electronic Journal* . [Crossref]
752. Allison Baker, Timothy Brennan, Jack Erb, Omar Nayeem, Aleksandr Yankelevich. 2014. Economics at the FCC, 2013–2014. *Review of Industrial Organization* **45**:4, 345-378. [Crossref]
753. Liran Einav, Jonathan Levin. 2014. Economics in the age of big data. *Science* **346**:6210. . [Crossref]
754. Sunny L Jardine, Juha V Siikamäki. 2014. A global predictive model of carbon in mangrove soils. *Environmental Research Letters* **9**:10, 104013. [Crossref]
755. Yael Grushka-Cockayne, Victor Richmond R. Jose, Kenneth C. Lichtendahl. 2014. Ensembles of Overfit and Overconfident Forecasts. *SSRN Electronic Journal* . [Crossref]
756. Tadas Bruzikas, Adriaan R. Soetevent. 2014. Detailed Data and Changes in Market Structure: The Move to Unmanned Gasoline Service Stations. *SSRN Electronic Journal* **24**. . [Crossref]
757. Sriganesh Lokanathan, Roshanthi Lucas Gunaratne. 2014. Behavioral Insights for Development from Mobile Network Big Data: Enlightening Policy Makers on the State of the Art. *SSRN Electronic Journal* . [Crossref]
758. Hui Chen, Winston Wei Dou, Leonid Kogan. 2013. Measuring the 'Dark Matter' in Asset Pricing Models. *SSRN Electronic Journal* **71**. . [Crossref]
759. Greg Distelhorst, Jens Hainmueller, Richard M. Locke. 2013. Does Lean Capability Building Improve Labor Standards? Evidence from the Nike Supply Chain. *SSRN Electronic Journal* . [Crossref]

760. Kaito Yamauchi, Takayuki Morimoto. 2013. Forecasting Financial Market Volatility Using a Dynamic Topic Model. *SSRN Electronic Journal* . [Crossref]
761. Marta Vidal, Javier Vidal-García, Rafael Hernandez Barros. Big Data and Business Decision Making 140-157. [Crossref]
762. Javier Vidal-García, Marta Vidal. Big Data Management in Financial Services 217-230. [Crossref]
763. José Luis Gómez-Barroso, Juan Ángel Ruiz. Behavioural Targeting in the Mobile Ecosystem 141-154. [Crossref]
764. Kees Zeelenberg, Barteld Braaksma. Big Data in Official Statistics 274-296. [Crossref]