

High-Dimensional Methods and Inference on Structural and Treatment Effects[†]

Alexandre Belloni, Victor Chernozhukov, and Christian Hansen

Data with a large number of variables relative to the sample size—“high-dimensional data”—are readily available and increasingly common in empirical economics. High-dimensional data arise through a combination of two phenomena.

First, the data may be inherently high dimensional in that many different characteristics per observation are available. For example, the US Census, the Current Population Survey, the Survey of Income and Program Participation, the National Longitudinal Survey of Youth, and the American Housing Survey collect information on hundreds of individual characteristics. Economists are also increasingly using scanner datasets that record transaction-level data for households across a wide range of products, or text data where counts of words in documents may be used as variables. In both of these latter examples, there may be thousands or tens of thousands of available variables per observation.

Second, even when the number of available variables is relatively small, researchers rarely know the exact functional form with which the small number of variables enters the model of interest. Researchers are thus faced with a large set of potential variables formed by different ways of interacting and transforming the underlying variables.

■ *Alexandre Belloni is Associate Professor of Decision Sciences, Fuqua School of Business, Duke University, Durham, North Carolina. Victor Chernozhukov is Professor of Economics, Massachusetts Institute of Technology, Cambridge, Massachusetts. Christian Hansen is Wallace W. Booth Professor of Econometrics and Statistics, University of Chicago Booth School of Business, Chicago, Illinois. Their email addresses are abn5@duke.edu, vchern@mit.edu, and chansen1@chicagobooth.edu.*

[†]To access the Appendix, visit <http://dx.doi.org/10.1257/jep.28.2.29>

There are many statistical methods available for constructing prediction models in the presence of high-dimensional data; for example, see Hastie, Tibshirani, and Friedman (2009) for a review. These methods tend to do a good job at prediction, which is what they are designed for, but they can often lead to incorrect conclusions when inference about model parameters such as regression coefficients is the object of interest (Leeb and Pötscher 2008a, b).

The goal of this paper is to provide an overview of how innovations in “data mining” can be adapted and modified to provide high-quality inference about model parameters. These data mining methods are relevant for learning about economic parameters where they are motivated, for example, by a desire to control properly for confounding variables. Note that here we use the term “data mining” in a modern sense which denotes a principled search for “true” predictive power that guards against false discovery and overfitting, does not erroneously equate in-sample fit to out-of-sample predictive ability, and accurately accounts for using the same data to examine many different hypotheses or models.

The key concept underlying the analysis of high-dimensional data is that dimension reduction or “regularization” is necessary to draw meaningful conclusions. The need for regularization can easily be seen when one considers an example where there are exactly as many variables (plus a constant) as there are observations. In this case, the ordinary least squares estimator will fit the data perfectly, returning an R^2 of one. However, using the estimated model is likely to result in very poor forecasting properties out-of-sample because the model estimated by least squares is overfit: the least-squares fit captures not only the signal about how predictor variables may be used to forecast the outcome, but also fits the noise that is present in the given sample, and is not useful for forming out-of-sample predictions. Producing a useful forecasting model in this simple case requires regularization; that is, the estimates must be constrained so that overfitting is avoided and useful out-of-sample forecasts can be obtained.

We begin with a discussion of “approximately sparse” regression models in high-dimensional data. These models are characterized by having many potential predictor/control variables of which only a few are important for predicting the outcome. The challenge in this case is to obtain good out-of-sample forecasts of outcome (and/or treatment) variables without assuming that the researcher knows which of the many available variables actually correspond to the important predictors. We then turn to the issue of model selection with high-dimensional data when the goal is learning about specific model parameters. We show how methods designed for forecasting in approximately sparse regression models can be used in this context. To illustrate these ideas, we apply them to examples from three papers in the empirical literature: estimating the effect of eminent domain on house prices, estimating the effect of abortion on crime, and estimating the effect of institutions on economic output. Our focus is not to rework these studies in any complete way but to show how one can bring high-dimensional analysis into this work and how this introduction can influence the findings of the analysis.

Approximately Sparse Regression Models

To fix ideas, suppose we are interested in forecasting outcome y_i with controls w_i according to the model

$$y_i = g(w_i) + \zeta_i,$$

where the expected value of the error terms ζ_i given w_i is equal to zero. Further, suppose we have a sample of $i = 1, \dots, n$ independent observations. To avoid overfitting and produce useful out-of-sample forecasts, we will generally need to restrict or regularize the function $g(\cdot)$.

There are many regularization approaches that produce the needed dimension reduction. Perhaps the simplest and most widely applied approach is the researcher making an ad hoc decision. Typically, applied researchers assume that they need only a small number of controls, which are chosen based on economic intuition and a modeling framework. Moreover, the researcher often assumes the controls enter the model in a simple fashion, usually linearly, allowing for the usual set of simple transformations and forming a small number of interaction terms. This approach has intuitive appeal and at some level is unavoidable. A researcher will always have to start by imposing some dimension reduction. However, it does leave one wondering whether the correct variables and functional forms were chosen.

Nonparametric methods, such as traditional series/sieve expansions, are also available. In this framework, one assumes that the model depends only on a small number of variables in a smooth but potentially unknown way and then uses a series of transformations of these variables in estimation; for example, see Newey (1997) and Chen (2007). Practical implementation of a nonparametric estimator requires that the researcher has selected an initial set of variables and a pre-specified set of series terms containing transformations of these variables. While more flexible than parametrics, traditional nonparametrics has a number of important limitations. Most importantly, it is again assumed that the most important terms for predicting y_i are contained within a pre-specified set of transformed variables determined by the researcher that is quite small relative to the number of observations.

In this paper, we focus on an approach to regularization that treats $g(w_i)$ as a high-dimensional, approximately linear model.¹ Specifically, we assume that

$$g(w_i) = \sum_{j=1}^p \beta_j x_{i,j} + r_{p,i}.$$

¹ There is also work on high-dimensional nonlinear models. For example, van de Geer (2008) and Belloni, Chernozhukov, and Wei (2013) consider high-dimensional generalized linear regression, and Belloni and Chernozhukov (2011), Belloni, Chernozhukov, and Kato (2013), and Kato (2011) consider quantile regression. We consider only linear models here for simplicity. The basic insights from the high-dimensional linear models extend to nonlinear settings though the theoretical analysis and practical computation is more complicated. High-dimensional linear models can also encompass many interesting settings and can accommodate flexible functional approximation just as nonparametric series estimators can.

The variables $x_i = (x_{i,1}, \dots, x_{i,p})'$ may simply be the elementary regressors w_i or may be made of transformations of these elementary regressors as in series modeling. In contrast to series modeling, we allow the number of these variables p to be larger than the sample size n . The final term $r_{p,i}$ is an approximation error. As with series, it is assumed that $r_{p,i}$ is small enough relative to sampling error in a well-defined sense (Bickel, Ritov, and Tsybakov 2009; Belloni, Chen, Chernozhukov, and Hansen 2012). Without further restrictions on the model, practical inference in this kind of high-dimensional linear model remains impossible since $p \geq n$ is allowed.

A structure that has played an important role in the literature is *approximate sparsity* of the high-dimensional linear model. Approximate sparsity imposes a restriction that only s variables among all of $x_{i,j}$, where s is much smaller than n , have associated coefficients β_j that are different from 0, while permitting a nonzero approximation error $r_{p,i}$. Thus, estimators for this kind of model attempt to learn the identities of the variables with large nonzero coefficients, while simultaneously estimating these coefficients.²

Note that the approximately sparse high-dimensional linear model structure includes as a special case both the traditional parametric and nonparametric model. The approximately sparse high-dimensional model generalizes these approaches by allowing the researcher to consider many explanatory variables and to use the data to learn which of the many variables are the most important. This setting thus encompasses many usual approaches to data analysis and accommodates the realistic scenario where a researcher does not know a priori exactly which variables should be included in a model.

An appealing method for estimating the parameters of sparse high-dimensional linear models is the Least Absolute Shrinkage and Selection Operator (LASSO), introduced by Frank and Friedman (1993) and Tibshirani (1996), where coefficients are chosen to minimize the sum of the squared residuals plus a penalty term that penalizes the size of the model through the sum of absolute values of the coefficients. In our discussion and empirical examples, we use a variant of the LASSO estimator that we proposed in Belloni, Chen, Chernozhukov, and Hansen (2012) defined as

$$\hat{\beta} = \arg \min_b \sum_{i=1}^n \left(y_i - \sum_{j=1}^p x_{i,j} b_j \right)^2 + \lambda \sum_{j=1}^p |b_j| \gamma_j,$$

where $\lambda > 0$ is the “penalty level” and γ_j are the “penalty loadings.” The penalty loadings are chosen to insure basic equivariance of coefficient estimates to rescaling of $x_{i,j}$ and can also be chosen to address heteroskedasticity, clustering, and non-normality in model errors. For more discussion, see Belloni, Chen, Chernozhukov, and Hansen (2012), Belloni, Chernozhukov, Hansen, and Kozbur (2014), and the online appendix for this paper at <http://e-jep.org>.

² Much of the high-dimensional linear model literature assumes that the model is exactly sparse, so the approximation error is identically 0. The approximately sparse model is strictly more general in that it allows for a nonzero approximation error in the analysis.

The penalty level, λ , controls the degree of penalization. Practical choices for λ that provably guard against overfitting are provided in Belloni, Chen, Chernozhukov, and Hansen (2012). (See also Belloni, Chernozhukov, Fernández-Val, and Hansen 2013; Belloni, Chernozhukov, Hansen, and Kozbur 2014.) It is also common to choose λ by cross-validation in prediction contexts, though it is important to note that this choice may not immediately equate to good performance when prediction is not the end goal.

The penalty function in the LASSO is special in that it has a kink at 0, which results in a sparse estimator with many coefficients set exactly to zero. Thus, the LASSO estimator may be used for variable selection by simply selecting the variables with nonzero estimated coefficients. A large part of the appeal of the LASSO estimator relative to other selection methods is that the LASSO problem is a convex optimization problem and highly efficient computational algorithms exist for its solution. LASSO-type estimators have also been shown to have appealing properties under plausible assumptions that allow for approximation errors, heteroskedasticity, clustering and fixed effects, and non-normality (Bickel, Ritov, and Tsybakov 2009; Belloni, Chen, Chernozhukov, and Hansen 2012; Belloni, Chernozhukov, Hansen, and Kozbur 2014; Gautier and Tsybakov 2011).

Finally, it is important to note that the nonzero coefficients that are part of the solution to the LASSO problem tend to be substantially biased towards zero. An appealing method to alleviate this bias is to employ the Post-LASSO estimator as in Belloni and Chernozhukov (2013) and Belloni, Chen, Chernozhukov, and Hansen (2012). The Post-LASSO estimator works in two steps. First, LASSO is applied to determine which variables can be dropped from the standpoint of prediction. Then, coefficients on the remaining variables are estimated via ordinary least squares regression using only the variables with nonzero first-step estimated coefficients. The Post-LASSO estimator is convenient to implement and, as we show in Belloni and Chernozhukov (2013) and Belloni, Chen, Chernozhukov, and Hansen (2012), works as well as and often better than LASSO in terms of rates of convergence and bias.

Model Selection When the Goal is Causal Inference

Using LASSO as a method for penalized estimation of the coefficients of a sparse linear model is useful for obtaining forecasting rules and for estimating which variables have a strong association to an outcome in a sparse framework. However, naively using the results obtained from such a procedure to draw inferences about model parameters can be problematic.

Part of the difficulty in drawing inferences after regularization or model selection is that these procedures are designed for forecasting, not for inference about model parameters. This observation suggests that more desirable inference properties may be obtained if one focuses on model selection over the predictive parts of the economic problem—the reduced forms and first-stages—rather than using model selection in the structural model directly.

The more difficult problem with doing inference following model selection is that model selection mistakes may occur. If one could be sure that the variable selector would always choose exactly all of the variables with nonzero coefficients, one could simply use the data to select this set of variables and then use the selected set coupled with any conventional procedure to do estimation and inference about parameters of interest. The validity of this approach is delicate because it relies on perfect model selection. Once one allows for the realistic scenario where some variables may have small but nonzero effects, it is likely there will be model selection mistakes in which such variables are not selected. The omission of such variables then generally contaminates estimation and inference results based on the selected set of variables. This problem is not restricted to the high-dimensional setting but is present even in low-dimensional settings when model selection is considered. This intuition is formally developed in Leeb and Pötscher (2008a, b). Because model selection mistakes seem inevitable in realistic settings, it is important to develop inference procedures that are robust to such mistakes.

An element in the provision of this robustness that has been employed recently is to focus on a small set of parameters of interest over which no model selection will be done, leaving model selection or regularization to be done only over “nuisance” parts of the problem. The estimation for the main parameters is then carried out using estimating equations that are *orthogonal* or *immune* to small perturbations in the nuisance parts. In Belloni, Chen, Chernozhukov, and Hansen (2012) and Belloni, Chernozhukov, and Hansen (2013, forthcoming), we provide an approach that does this in a canonical instrumental variable model; see also Ng and Bai (2009). In Belloni, Chernozhukov, and Hansen (2013) and Belloni, Chernozhukov, and Hansen (forthcoming), we provide an approach for inference about coefficients in a partially linear model, or about average treatment effects in a heterogeneous treatment effects model with binary treatment; see also Farrell (2013). In addition to showing how to obtain valid inference following model selection in canonical econometric models, these papers develop basic intuition helpful in understanding how inference following regularization may be performed outside of these models. Thus, we outline the approaches of these papers below.

Providing formal results for doing inference about parameters following model selection for other models relevant in applied economics is a topic of ongoing research. For example, in Belloni, Chernozhukov, Fernández-Val, and Hansen (2013), we consider the estimation of heterogeneous treatment effects with endogenous receipt of treatment and present a general framework for econometric models where the orthogonality condition is explained in detail.

Inference with Selection among Many Instruments

Consider the linear instrumental variables model with potentially many instruments

$$y_i = \alpha d_i + \varepsilon_i$$

$$d_i = z_i' \Pi + r_i + v_i,$$

where $E[\varepsilon_i | z_i] = E[v_i | z_i, r_i] = 0$ but $E[\varepsilon_i v_i] \neq 0$, leading to endogeneity. In this setting, d_i is a scalar endogenous variable of interest, z_i is a p -dimensional vector of instruments where the number of instruments p may be much larger than the number of observations,³ and r_i is an approximation error. Allowing for a small number of included exogenous variables is straightforward by defining the variables in the model as residuals after partialing these variables out, and we suppress this case for simplicity. The results in Belloni, Chen, Chernozhukov, and Hansen (2012) also allow for a nonscalar but finite-dimensional treatment vector.

One approach to estimation and inference about α in this context is to select a small number of instruments from z_i to use in a conventional two-stage least squares estimation. In Belloni, Chen, Chernozhukov, and Hansen (2012), we provide a set of formal conditions under which conventional inference from the two-stage least squares estimator based on instruments selected by LASSO or another variable selection procedure is valid for learning about the parameter of interest, α . The key features that allow this can be illustrated by noting that this model cleanly fits into the heuristic outline for doing valid inference after using high-dimensional methods provided above. The parameter of interest, α , is finite-dimensional and there is no selection over whether d_i will be included in the model. The variable selection component of the problem is limited to the first-stage equation relating the endogenous variable to the instruments, which is a pure predictive relationship. Finally, the structure of the problem is such that model selection mistakes in which a valid instrument with a small but nonzero coefficient is left out of the first-stage will not substantially affect the second-stage estimator of α as long as other instruments with large coefficients are selected. In other words, the second-stage instrumental variable estimate is *orthogonal* or *immune* to variable selection errors where instruments with small, nonzero coefficients are mistakenly excluded from estimation.

Inference with Selection among Many Controls

As a more complex example, consider a linear model where a treatment variable, d_i , is taken as exogenous after conditioning on control variables:

$$y_i = \alpha d_i + x_i' \theta_y + r_{yi} + \zeta_i,$$

where $E[\zeta_i | d_i, x_i, r_{yi}] = 0$, x_i is a p -dimensional vector of controls where $p \gg n$ is allowed, r_{yi} is an approximation error, and the parameter of interest is α , the effect of the treatment on the outcome.

Before turning to a procedure that provides high-quality estimates and inferential statements about α , it is useful to discuss some intuitive benchmarks that do

³ In the instrumental variables setting, there are many papers that examine the properties of various instrumental variables estimators under many-instrument asymptotics where the number of instruments p is allowed to increase with the sample size n in such a way that $p < n$ and $p/n \rightarrow \rho < 1$; see, for example, Bekker (1994), Chao and Swanson (2005), Hansen, Hausman, and Newey (2008), and Hausman, Newey, Woutersen, Chao, and Swanson (2012). These approaches do not apply when $p \geq n$ and tend to perform poorly when $p/n \approx 1$.

not work. Considering such cases builds intuition concerning features that must be guarded against in applying high-dimensional methods in this and related contexts.

One naive approach would attempt to select control variables by applying LASSO to the equation above, forcing the treatment variable to remain in the model by excluding α from the LASSO penalty. One could then try to estimate and do inference about α by applying ordinary least squares with y_i as the outcome, and d_i and any selected control variables as regressors. The problem with this approach can be seen by noting that LASSO and any other high-dimensional modeling device targets *prediction*, not learning about specific model parameters. From the standpoint of prediction, any variable that is highly correlated to the treatment variable will tend to be dropped since including such a variable will tend not to add much predictive power for the outcome given that the treatment is already in the model. Of course, the exclusion of a variable that is highly correlated to the treatment will lead to substantial *omitted-variables bias* if the coefficient in θ_y associated with the variable is nonzero. Such omissions will happen routinely in any procedure that looks just at the equation above.

There are two problems with the above naive approach. First, it ignores a key component to understanding omitted-variables bias, the relationship between the treatment variable and the controls. To aid in learning about this relationship, we introduce an additional “reduced form” relation between the treatment and controls:

$$d_i = x_i' \theta_d + r_{di} + v_i,$$

where $E[v_i | x_i, r_{di}] = 0$. The other problem is that the naive approach is based on a “structural” model where the target is to learn the treatment effect given controls, not an equation representing a forecasting rule for y_i given d_i and x_i . It is thus useful to transform the first equation of this section to a reduced form, *predictive* equation by substituting the equation introduced for d_i into the “structural” equation yielding the reduced form system:

$$y_i = x_i' (\alpha \theta_d + \theta_y) + (\alpha r_{di} + r_{yi}) + (\alpha v_i + \zeta_i) = x_i' \pi + r_{ci} + \varepsilon_i$$

$$d_i = x_i' \theta_d + r_{di} + v_i,$$

where $E[\varepsilon_i | x_i, r_{ci}] = 0$, r_{ci} is a composite approximation error, and the second equation is the same as above. Both of these equations represent predictive relationships, which may be estimated using high-dimensional methods.

Before turning to the recommended procedure, let us mention the second set of naive procedures that use only one of these two equations for selection. The problem with working with only one of the above equations is that single-equation approaches rely on there being no errors in variable selection. To see the problem with such an approach, note that applying a variable selection method to say the first equation for forecasting y_i with x_i will tend to select variables with large entries

in coefficient vector π but will tend to miss variables with moderately sized coefficients. However, missing variables that have strong predictive power for d_i , namely variables with large coefficients in θ_a , may lead to substantive *omitted-variables bias* in the estimator of α if the coefficients on these variables in θ_y are moderately sized. Intuitively, any such variable has a moderate direct effect on the outcome that will be incorrectly misattributed to the effect of the treatment when this variable is strongly related to the treatment and the variable is not included in the regression. Similarly, if one applied a variable selection method to only the second equation for predicting d_i , one would potentially miss variables that have moderate-sized coefficients in predicting d_i but large direct effects on y_i . Such an omission may again lead to non-negligible omitted-variables bias.

To guard against such model selection mistakes, it is important to consider *both* equations for selection: we apply variable selection methods to *each* of the two reduced form equations and then use all of the selected controls in estimation of α . Thus, variable selection is used to select a set of variables that are useful for predicting y_i , say x_{y_i} , and a set of variables that are useful for predicting d_i , say x_{d_i} . We then estimate α by ordinary least squares regression of y_i on d_i and the *union* of the variables selected for predicting y_i and d_i , contained in x_{y_i} and x_{d_i} . We thus make sure we use variables that are important for either of the two predictive relationships to guard against omitted-variables bias, discussed above, when estimating α .

Using both variable selection steps immunizes the resulting procedure against the types of model selection mistakes discussed above for single-equation procedures. Specifically, using the variables selected in both reduced form equations ensures that any variables that have large effects in either the “structural” equation for y_i or the reduced form equation for d_i are included in the model. Any excluded variables are therefore at most mildly associated to y_i and d_i , which greatly limits the scope for omitted-variables bias. It is also noteworthy that the “double selection” procedure implicitly estimates the residuals ε_i and v_i and then regresses the estimates of ε_i on the estimates of v_i to construct an estimator of α , thereby providing a selection analog of Robinson’s (1988) method for estimating the parameters of a partially linear model to the high-dimensional case.

In Belloni, Chernozhukov, and Hansen (2013, forthcoming), we provide formal conditions under which this “double selection” procedure will lead to valid inference about α even when selection mistakes are allowed, and provide substantial simulation evidence that the procedure works across a wide variety of models. Using both selection steps also enhances efficiency by finding variables that are strongly predictive of the outcome and may remove residual variance.⁴

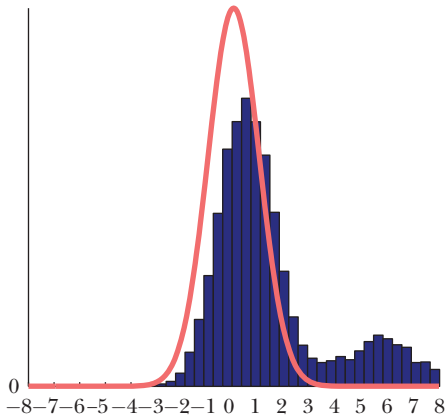
As a concrete illustration of the potential pitfalls of naive procedures and the robustness of the “double selection” approach, we present results from a simulation

⁴That is, standard errors may go down, at least theoretically, after performing the variable selection steps if the selected variables reduce residual variance sufficiently to offset the increased variability due to including more variables. In fact, under homoskedasticity, the estimator is semi-parametrically efficient, achieving the efficiency bound of Robinson (1988).

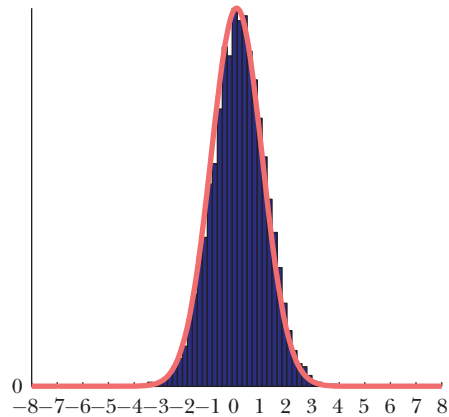
Figure 1

The “Double Selection” Approach to Estimation and Inference versus a Naive Approach: A Simulation from Belloni, Chernozhukov, and Hansen (forthcoming)
(distributions of estimators from each approach)

A: A Naive Post-Model Selection Estimator



B: A Post-Double-Selection Estimator



Source: Belloni, Chernozhukov, and Hansen (forthcoming).

Notes: The left panel shows the sampling distribution of the estimator of α based on the first naive procedure described in this section: applying LASSO to the equation $y_i = d_i + x_i' \theta_y + r_{y_i} + \zeta_i$ while forcing the treatment variable to remain in the model by excluding α from the LASSO penalty. The right panel shows the sampling distribution of the “double selection” estimator (see text for details) as in Belloni, Chernozhukov, and Hansen (forthcoming). The distributions are given for centered and studentized quantities.

exercise in this linear modeling context in Figure 1. Details underlying the simulation are as in Belloni, Chernozhukov, and Hansen (2013). The left panel shows the sampling distribution of the estimator of α based on the first naive procedure discussed in this section, while the right panel shows the sampling distribution of the “double selection” estimator. The second mode in the left panel is due to model selection mistakes where important variables are missed leading to badly biased estimates of α . This strong omitted-variables bias is absent from the distribution of the “double selection” estimator, which was specifically designed to reduce the influence of such mistakes as discussed above.

Some Empirical Examples

In this section, we provide three concrete examples of the use of these methods. An online Appendix available with this paper at <http://e-jep.org> provides implementation details.

Estimating the Impact of Eminent Domain on House Prices

We consider instrumental variable estimation of the effects of federal appellate court decisions regarding eminent domain on housing prices. Recall that eminent domain refers to the government's taking of private property. Federal court rulings that a government seizure was unlawful (pro-plaintiff rulings) thus uphold individual property rights and make future exercise of eminent domain more difficult due to the structure of the US legal system. A more detailed discussion of the economics of takings law (or eminent domain) and other institutional and econometric considerations can be found in Belloni, Chen, Chernozhukov, and Hansen (2012) and Chen and Yeh (2012).

The analysis of the effects of takings law is complicated by the possible endogeneity between takings law decisions and economic variables: for example, a taking may be less likely if real estate prices are low and sellers are eager to unload property. To address the potential endogeneity of takings law, we employ an instrumental variables strategy based on the identification argument of Chen and Sethi (2010) and Chen and Yeh (2012) that relies on the random assignment of judges to federal appellate panels. Because judges are randomly assigned to three-judge panels to decide appellate cases, the exact identity of the judges and their demographics are randomly assigned conditional on the distribution of characteristics of federal circuit court judges in a given circuit-year. Under this random assignment, the characteristics of judges serving on federal appellate panels can only be related to property prices through the judges' decisions; thus the judge's characteristics will plausibly satisfy the instrumental variable exclusion restriction.

Following this argument, we try to uncover the effect of takings law by estimating models of the form

$$\log(\text{Case-Shiller}_{ct}) = \alpha \cdot \text{TakingsLaw}_{ct} + \beta_c + \beta_t + \gamma_c t + W'_{ct} \delta + \varepsilon_{ct}$$

using the characteristics of judges actually assigned to cases as instruments for TakingsLaw_{ct} . In this equation, Case-Shiller_{ct} is the average of the Case-Shiller home price index within circuit court c at time t ; TakingsLaw_{ct} represents the number of pro-plaintiff appellate takings decisions in federal circuit court c and year t ; W_{ct} are included exogenous variables that include a dummy variable for whether there were relevant cases in that circuit-year, the number of takings appellate decisions, and controls for the distribution of characteristics of federal circuit court judges in a given circuit-year; and β_c , β_t , and $\gamma_c t$ are respectively circuit-specific effects, time-specific effects, and circuit-specific time trends. An appellate court decision is coded as pro-plaintiff if the court ruled that a taking was unlawful, thus overturning the government's seizure of the property in favor of the private owner. The parameter of interest, α , thus represents the effect of an additional decision upholding individual property rights on a measure of property prices. The sample size in this example is 183.

The argument given above suggests that judges' characteristics satisfy the instrumental variables exclusion restriction. Of course, to be valid instruments, the characteristics must also be useful for predicting judicial decisions. In the

data, we observe a variety of demographic information about each judge, and the basic identification argument suggests that any set of characteristics of the three-judge panel will be unrelated to structural unobservables. Given the large number of instruments that could be constructed by considering all combinations of characteristics of three-judge panels, it is also infeasible to just use all possible instruments.

Thus, a sensible way to proceed is to use variable selection methods to find a set of good instruments from a large set of intuitively chosen potential instruments. Under the exclusion restriction, the ideal set of instruments provides a high-quality prediction of the endogenous variable—judicial decisions in this example. Forming high-quality predictions, which is of course different from obtaining a good in-sample fit, is exactly what LASSO and other data mining procedures are designed to do. Note that using LASSO with proper penalty parameters theoretically guarantees that any instruments selected are not simply spuriously correlated to the endogenous variable but have true predictive power. This guarantee means that LASSO could select no instruments at all as there may be no set of variables with sufficient predictive power to achieve the required standard.

Intuitively, reliably distinguishing true predictive power from spurious association becomes more difficult as more variables are considered. This intuition can be seen in the theory of high-dimensional variable selection methods, and the methods work best in simulations when selection is done over a collection of variables that is not overly extensive. It is therefore important that some persuasive economic intuition exists to produce a carefully chosen, well-targeted set of variables to be selected over even when using automatic variable selection methods.

In this example, we first did dimension reduction by intuitively selecting characteristics thought to have strong signals about judge preferences over government versus individual property rights. We chose to consider only gender, race, religion (Jewish, Catholic, Protestant, evangelical, not-religious), party affiliation, source of academic degrees (bachelor's degree from an in-state university, bachelor's degree from a public university, JD from a public university, has an LLM or SJD), and whether the judge had been elevated from a district court. For each of these baseline variables, we then constructed three new variables, counting the number of panels with one member with each characteristic, two members with each characteristic, and three members with each characteristic. To allow for nonlinearities, we included first-order interactions between all of the previously mentioned variables, a cubic polynomial in the number of panels with at least one Democrat, a cubic polynomial in the number of panels with at least one member with a JD from a public university, and a cubic polynomial in the number of panels with at least one member elevated from within the district. In addition to limiting the selection to be over this set of baseline variables, we did additional pre-processing to remove instruments that we thought likely to be irrelevant based on features of the instrument set alone. We removed any instrument where the standard deviation was extremely small and also removed one instrument from any pair of instruments that had a bivariate correlation exceeding .99 in absolute

value. These instruments were removed as they were highly unlikely to have much power.⁵ After these initial choices, we are left with a total of 147 instruments. The number of instruments plus the number of control variables is greater than the number of observations in this example, so conventional instrumental variables estimators using the full set of variables are not defined.

With this set of 147 instruments, we then estimate the first-stage relationship using LASSO. The estimated coefficients have just one nonzero element, the coefficient on the number of panels with one or more members with JD from a public university squared. Using this instrument gives a first-stage coefficient of 0.4495 with estimated standard error of 0.0511—that is, this variable appears to be a strong instrument. The second stage estimate using the LASSO-selected instrument is then 0.0648 with estimated standard error of 0.0240. This estimate is statistically significant at the usual levels, suggesting that a single additional judicial decision reinforcing individual property rights is associated with between 2 and 11 percent higher property prices with an average number of pro-plaintiff decisions per year of 0.19.

For comparison, we also experimented with an “intuitive” instrumental variable using the number of judicial panels with one or more Democrats. The political affiliation of judges is known to predict judicial decisions in several contexts, so one might hypothesize that this intuition carries over to judicial decisions regarding eminent domain. When we used the number of panels with one or more judges identified as Democrats as the single instrument, we found that one would not reject the hypothesis that this instrument is unrelated to the endogenous variable, the number of pro-plaintiff decisions, at any reasonable confidence level—which in turn suggests that this instrument is too weak to be useful.

We suspect that most analysts, relying on intuition about what might be a useful instrumental variable, would not have intuited that they should use the number of judicial panels with one or more members with JD from a public university squared. However, we find that one obtains a much stronger first-stage relationship in a two-stage least squares approach using this instrument selected by a formal variable selection method, relative to that obtained by an “intuitive” benchmark. This stronger first-stage in turn leads to a corresponding sensible and reasonably precise second-stage estimate. This substantive difference suggests that high-dimensional techniques may usefully complement researchers’ intuition for choosing instruments and strengthen their ability to draw useful conclusions from the data.

Estimating the Effect of Legalized Abortion on Crime

Donohue and Levitt (2001) sought to estimate the effect of abortion on crime rates. Looking at state-level data, they find that higher rates of abortion in the years

⁵ Note that selection based on characteristics of the instruments without reference to the endogenous variable or outcome cannot introduce bias as long as the instruments satisfy the instrumental variable exclusion restriction.

around 1970, as legal restrictions on abortion were eased in a number of states, are associated with lower rates of crime two decades later. However, there is a basic problem in estimating the causal impact of abortion on crime: state-level abortion rates during the earlier time period were not randomly assigned. It seems at least plausible that certain factors may be associated with both state-level abortion rates and state-level crime rates. Failing to control for these factors will then lead to omitted-variables bias in the estimated abortion effect.

To address these potential confounding factors, Donohue and Levitt (2001) estimate a differences-in-differences style model for state-level crime rates running from 1985 to 1997. Their basic specification is

$$y_{cit} = \alpha_c a_{cit} + w_{it}'\beta_c + \delta_{ci} + \gamma_{ct} + \varepsilon_{cit}.$$

The dependent variable, y_{cit} , indexes the crime-rate for crime type c (categorized between *violent*, *property*, and *murder*) in state i in year t . On the right-hand side, the independent variables are w_{it} , a measure of the abortion rate relevant for type of crime c (as determined by the ages of criminals when they tend to commit crimes); w_{it} , a set of variables to control for time-varying confounding state-level factors; δ_{ci} , state-specific effects that control for any time-invariant state-specific characteristics; and γ_{ct} , time-specific effects that control for national aggregate trends. For control variables w_{it} , Donohue and Levitt (2001) include the log of lagged prisoners per capita, the log of lagged police per capita, the unemployment rate, per-capita income, the poverty rate, the generosity of the Aid to Families with Dependent Children (AFDC) welfare program at time $t - 15$, a dummy for having a concealed weapons law, and beer consumption per capita. They present baseline estimation results based on this formulation as well as results from different models which vary the sample and set of controls in their tables IV and V. We refer the reader to the original paper for additional details, data definitions, and institutional background.

In this example, we take first-differences of the basic Donohue-Levitt formulation as our baseline. We use the same state-level data as Donohue and Levitt (2001) but delete Washington, DC,⁶ which gives a sample with 50 cross-sectional observations and 12 time-series observations for a total of 600 observations. With these deletions, our baseline estimates using the same controls are quite similar to those reported in Donohue and Levitt (2001). Estimates of the effect of abortion on crime from this first-difference model are given in the first row of Table 1. These baseline results suggest that increases in abortion rates are strongly associated with decreases in crime rates; for example, an increase in the effective abortion rate of

⁶ Removing Washington DC produces results similar to those in Donohue and Levitt (2001) without the need to introduce the weights used in Donohue and Levitt (2001) and is done for simplicity. This similarity between the weighted results and unweighted results excluding Washington DC is also discussed in Donohue and Levitt (2001).

Table 1
Effect of Abortion on Crime

| Estimator | Type of crime | | | | | |
|------------------|---------------|------------|----------|------------|--------|------------|
| | Violent | | Property | | Murder | |
| | Effect | Std. error | Effect | Std. error | Effect | Std. error |
| First-difference | -.157 | .034 | -.106 | .021 | -.218 | .068 |
| All controls | .071 | .284 | -.161 | .106 | -1.327 | .932 |
| Double selection | -.171 | .117 | -.061 | .057 | -.189 | .177 |

Notes: This table reports results from estimating the effect of abortion on violent crime, property crime, and murder. The row labeled “First-difference” gives baseline first-difference estimates using the controls from Donohue and Levitt (2001). The row labeled “All controls” includes a broad set of controls meant to allow flexible trends that vary with state-level characteristics. The row labeled “Double selection” reports results based on the double selection method outlined in this paper and selecting among the variables used in the “All controls” results.

100 per 1,000 live births is associated with around a 15 percent reduction in violent crime. This association may be taken as causal under the assumption that all potential confounding factors not captured in w_{it} are either time-invariant or captured by a national trend.

Due to the inclusion of state and time effects, the baseline specification will control for any factors related to abortion and crime rates that are either time-invariant or vary only at the national level. While this formulation is fairly flexible, it produces valid estimates of the causal effect of abortion on crime rates *only if* time-varying state-specific factors that are correlated to both abortion and crime rates are captured by a small set of characteristics. An approach that is sometimes used to help alleviate such concerns is to include a set of state-specific linear time trends in the model to account for differences in state-specific trends that may be related to both the outcome and treatment variable of interest. However, this approach introduces many additional variables. Perhaps more importantly, the assumption of a linear state-specific trend is questionable in many circumstances as an approximation and certainly cannot capture the evolution of variables such as the crime rate or the abortion rate over any long time horizon.

Instead of using state-specific linear trends, we consider a generalization of the baseline model that allows for nonlinear trends interacted with observed state-specific characteristics and then use variable selection methods to find potentially important confounding variables. This approach allows us to consider quite flexible models without including so many additional variables that it becomes mechanically impossible to learn about the abortion effect. A key choice in using high-dimensional variable selection methods is the set of candidate variables to consider. For this example, our choice of these variables was motivated by our desire to accommodate a flexible trend that might offer a sensible model of the evolution of abortion or

crime rates over a 12-year period. To accomplish this, we use the double-selection procedure outlined in the previous section with models of the form

$$\Delta y_{cit} = \alpha_c \Delta a_{cit} + z_{cit}' \beta_c + \tilde{\gamma}_{ct} + \Delta \varepsilon_{cit}$$

$$\Delta a_{cit} = z_{cit}' \Pi_c + \tilde{\kappa}_{ct} + \Delta v_{cit}.$$

In this formulation, $\Delta y_{cit} = y_{cit} - y_{cit-1}$ and Δa_{cit} , $\Delta \varepsilon_{cit}$, and Δv_{cit} are defined similarly; $\tilde{\gamma}_{ct}$ and $\tilde{\kappa}_{ct}$ are time effects; z_{cit} is a large set of controls; and we have introduced an equation for the abortion rate to make the relation to the earlier discussion clear. z_{itc} consists of 284 variables made up of the levels, differences, initial level, initial difference, and within-state average of the eight state-specific time-varying observables, the initial level and initial difference of the abortion rate relevant for crime type c , quadratics in each of the preceding variables, interactions of all the aforementioned variables with t and t^2 , and the main effects t and t^2 . This set of variables corresponds to a cubic trend for the level of the crime rate and abortion rate that is allowed to depend on observed state-level characteristics.

Because the set of variables we consider has fewer elements than there are observations, we can estimate the abortion effect after controlling for the full set of variables. Results from ordinary least squares regression of the differenced crime rate on the differenced abortion rate, a full set of time dummies, and the full set of variables in z_{itc} are given in the second row of Table 1. The estimated abortion effects are extremely imprecise with confidence intervals at the usual levels including implausibly large negative and implausibly large positive values for the abortion effect across all three outcomes. Of course, very few researchers would consider using 284 controls with only 600 observations because of exactly this issue.

The final row of Table 1 provides the estimated abortion effects based on the double-selection method of Belloni, Chernozhukov, and Hansen (forthcoming). At each stage of the process, we include the full set of time dummies without penalizing the parameters on these variables, which results in their selection in all cases, as we wish to allow for a flexible aggregate trend. In this example, we use LASSO to select variables from z_{cit} that are useful for predicting the change in crime rate Δy_{cit} and the change in the associated abortion rate. We then use the union of the set of selected variables, including time effects, as controls in a final ordinary least squares regression of Δy_{cit} on Δa_{cit} . In all equations, the selected variables suggest the presence of a nonlinear trend that depends on state-specific characteristics.⁷ Looking at

⁷ For violent crime, lagged prisoners, lagged police, lagged police $\times t$, the initial income difference, the initial income difference $\times t$, the initial beer consumption difference $\times t$, average income, average income $\times t$, and the initial abortion rate are selected in the abortion equation; and no variables are selected in the crime equation. For property crime, lagged prisoners, lagged police, lagged income, the initial income difference, the initial income difference $\times t$, average income, and the initial abortion rate are selected in the abortion equation; and initial income squared $\times t^2$ and average income squared $\times t^2$ are selected in the crime equation. For the murder rate, lagged prisoners, lagged prisoners $\times t$, lagged police $\times t$, the initial income difference $\times t$, average income $\times t$, the initial abortion

the results, we see that estimated abortion effects are much more precise than the “kitchen sink” results that include all controls. However, the double-selection estimates for the effect of abortion on crime rates still produce 95 percent confidence intervals that encompass large positive and negative values.

It is interesting that one would draw qualitatively different conclusions from the estimates obtained using formal variable selection than from the estimates obtained using a small set of intuitively selected controls. Specifically, one would conclude that increases in abortion have a strong negative effect on crime rates using a small set of intuitively selected controls but would fail to reject the hypothesis that abortion is unrelated to crime rates at usual significance levels using estimates obtained using formal variable-selection. Of course, this comparison does not mean that the effect of the abortion rate provided in the first row of Table 1 is inaccurate for measuring the causal effect of abortion on crime. It does, however, imply that this conclusion is not robust to the presence of fairly parsimonious nonlinear trends. Foote and Goetz (2008) reach a similar conclusion based on an intuitive argument.⁸

Estimating the Effect of Institutions on Output

For our final example, we consider estimation of the effect of institutions on aggregate output following the work of Acemoglu, Johnson, and Robinson (2001). Estimating the effect of institutions on output is complicated by the clear potential for simultaneity between institutions and output: specifically, better institutions may lead to higher incomes, but higher incomes may also lead to the development of better institutions. To help overcome this simultaneity, Acemoglu, Johnson, and Robinson (2001) use mortality rates for early European settlers as an instrument for institution quality. The validity of this instrument hinges on the argument that settlers set up better institutions in places where they are more likely to establish long-term settlements; that where they are likely to settle for the long term is related to settler mortality at the time of initial colonization; and that institutions are highly persistent. The exclusion restriction for the instrumental variable is then motivated by the argument that GDP, while persistent, is unlikely to be strongly influenced by mortality in the previous century, or earlier, except through institutions.

In their paper, Acemoglu, Johnson, and Robinson (2001) note that their instrumental variable strategy will be invalidated if other factors are also highly persistent and related to the development of institutions within a country and to the country’s GDP. A leading candidate for such a factor, as they discuss, is geography. Thus, they control for the distance from the equator in their baseline specifications and also consider specifications with different sets of geographic controls such as dummy variables for continents; see their table 4.

rate, and the initial abortion rate $\times t$ are selected in the abortion equation; and no variables are selected in the crime equation.

⁸ See also Donohue and Levitt’s (2008) response to Foote and Goetz (2008), which considers the same problem using a longer panel and finds similar results to Donohue and Levitt (2001) including state-specific linear time trends.

As a complement to these results, we consider using high-dimensional methods to aid in estimating the model

$$\log(\text{GDPpercapita}_i) = \alpha \cdot \text{ProtectionfromExpropriation}_i + x_i' \beta + \varepsilon_i.$$

We use the same set of 64 country-level observations as Acemoglu, Johnson, and Robinson (2001). *ProtectionfromExpropriation* is a measure of the strength of individual property rights that is used as a proxy for the strength of institutions, and x_i is a set of variables that are meant to control for geography. The underlying identifying assumption is the same as that employed in Acemoglu, Johnson, and Robinson (2001), which is that mortality risk is a valid instrument after controlling for geography. Acemoglu, Johnson, and Robinson (2001) address this by assuming that the confounding effect of geography is adequately captured by a linear term in distance from the equator or a set of dummy variables. The use of high-dimensional methods allow us to replace this assumption by the assumption that geography can be sufficiently controlled for by a small number of variables constructed from geographic information whose identities will be learned from the data.

To make use of high-dimensional methods, we note that the model in this example is equivalent to the three-equation system

$$\log(\text{GDPpercapita}_i) = \alpha \cdot \text{ProtectionfromExpropriation}_i + x_i' \beta + \varepsilon_i$$

$$\text{ProtectionfromExpropriation}_i = \pi_1 \cdot \text{SettlerMortality}_i + x_i' \Pi_2 + v_i$$

$$\text{SettlerMortality}_i = x_i' \gamma + u_i,$$

which yields three reduced form equations relating the structural variables to the controls:

$$\log(\text{GDPpercapita}_i) = x_i' \tilde{\beta} + \tilde{\varepsilon}_i$$

$$\text{ProtectionfromExpropriation}_i = x_i' \tilde{\Pi}_2 + \tilde{v}_i$$

$$\text{SettlerMortality}_i = x_i' \gamma + u_i.$$

We can thus select a set of control terms by carrying out variable selection for each of these three reduced form equations using the essential idea outlined in the discussion of selecting control variables. Valid estimation and inference for the parameter α can then proceed by conventional instrumental variable estimation using *SettlerMortality*_{*i*} as an instrument for *ProtectionfromExpropriation*_{*i*}, with the union of variables selected from each reduced form as included control variables.

It is important that a set of baseline variables be selected before variable selection methods are applied. Our target is to control for geography, so we consider

Table 2
Effect of Institutions on Output

| | <i>Latitude</i> | <i>All controls</i> | <i>Double selection</i> |
|--------------|---------------------|---------------------|-------------------------|
| First stage | -0.5372 (0.1545) | -0.2182 (0.2011) | -0.5429 (0.1719) |
| Second stage | 0.9692 (0.2128) | 0.9891 (0.8005) | 0.7710 (0.1971) |

Notes: In an exercise that follows the work of Acemoglu, Johnson, and Robinson (2001), this table reports results from estimating the effect of institutions, using settler mortality as an instrument. The row “First Stage” gives the first-stage estimate of the coefficient on settler mortality obtained by regressing “*ProtectionfromExpropriation_i*” on “*SettlerMortality_i*” and the set of control variables indicated in the column heading. The row “Second stage” gives the estimate of the structural effect of institutions on $\log(\text{GDP per capita})$ using “*SettlerMortality_i*” as the instrument and controlling for variables as indicated in the column heading (see text for details). Each column reports the results for different sets of control variables. The column “Latitude” controls linearly for distance from the equator. The column “All controls” includes 16 controls defined in the main text and in footnote 9, and the column “Double selection” uses the union of the set of controls selected by LASSO for predicting GDP per capita, for predicting institutions, and for predicting settler mortality. Standard errors are in parentheses.

a flexible but still parsimonious set of variables constructed from geography. Specifically, we set x_i equal to the dummy variables for Africa, Asia, North America, and South America plus a cubic-spline in latitude (altogether, twelve variables for latitude).⁹

We report estimation results in Table 2. The first row of the table, labeled “First stage,” gives the estimate of the coefficient on “*SettlerMortality*” from the first stage regression of “*ProtectionfromExpropriation*” on “*SettlerMortality*” and a set of control variables defined by the column headings, with the corresponding estimated standard error provided in parentheses below the coefficient estimate. The second row of the table, labeled “Second stage,” gives the estimate of the structural effect of institutions on “ $\log(\text{GDPpercapita})$ ” obtained by instrumental variables estimation of “ $\log(\text{GDPpercapita})$ ” on “*ProtectionfromExpropriation*” using “*SettlerMortality*” as the instrument and controlling for variables as indicated in the column heading, with the estimated standard error again provided below the coefficient estimate in parentheses.

The first column of the table labeled “Latitude” gives baseline results that control linearly for latitude. These results correspond to the findings of Acemoglu, Johnson, and Robinson (2001), suggesting a strong positive effect of improved institutions on output with an underlying reasonably strong first-stage. This contrasts

⁹ Specifically, we include latitude, latitude², latitude³, (latitude−.08)₊, (latitude−.16)₊, (latitude−.24)₊, ((latitude−.08)₊)², ((latitude−.16)₊)², ((latitude−.24)₊)², ((latitude−.08)₊)³, ((latitude−.16)₊)³, and ((latitude−.24)₊)³ where latitude denotes the distance of a country from the equator normalized to be between 0 and 1, the breakpoints in the latitude function were chosen by taking round numbers near the quartiles of latitude, and $f(a) = (a)_+$ is shorthand notation for $f(a) = (a)1(a > 0)$ where $1(\cdot)$ is the indicator function that returns 1 when the expression inside the parentheses is true and 0 otherwise.

strongly with the second column of the table, which gives results controlling for all 16 of the variables defined above and in footnote 9. Controlling for the full set of terms results in a very imprecisely estimated first-stage. The estimate of the effect of institutions is then unreliable given the weak first-stage.

The variable selection methods discussed in this paper are constructed to produce a reasonable trade-off between this perhaps overly flexible second case and the parsimonious first case by allowing flexible functions to be considered but only using terms which are useful for understanding the underlying reduced form relationships. The final column of Table 2 labeled “Double selection” controls for the union of the set of variables selected by running LASSO on each of the three reduced form equations. The same single variable, the dummy variable for Africa, is selected in the reduced form equations for GDP and mortality, and no variables are selected in the reduced form for the expropriation variable. Thus, the final column is simply the conventional instrumental variable estimate with the Africa dummy included as the single control variable. The results are qualitatively similar to the baseline results, though the first-stage is somewhat weaker and the estimated effect of institutions is slightly attenuated though still very strong and positive. The slightly weaker first-stage suggests that the intuitive baseline obtained by controlling linearly for latitude may be inadequate, though the results are not substantively altered in this case. Again, we believe these results suggest that high-dimensional techniques may usefully complement the sets of sensitivity analyses that researchers are already doing (such as those underlying table 4 of Acemoglu, Johnson, and Robinson 2001). High-dimensional techniques can add rigor to these exercises and thus potentially strengthen the plausibility of conclusions drawn in applied economic papers.

Conclusion

The high-dimensional methods discussed in this paper provide a useful addition to the standard tools used in applied economic research. They allow researchers to perform inference about economically interesting model parameters in settings with rich confounding information. In these settings, dimension reduction is important if one hopes to learn from the data. We have emphasized variable selection methods for achieving this dimension reduction and outlined procedures that provide valid inference in simple, canonical models allowing for inevitable variable selection mistakes. In Belloni, Chernozhukov, Fernández-Val, and Hansen (2013), we provide an extended treatment showing that valid post-selection inference is generally available when estimation is based on orthogonal estimating equations.

We hope readers will see that data mining done correctly is the opposite of “bad practice”: it is an extremely useful tool that opens many doors in the analysis of interesting economic data. These tools allow researchers to add rigor and robustness to the “art” of variable or model selection in data analyses where the aim is to draw inferences about economically meaningful parameters. We have only skimmed

the surface of the growing set of statistical methods appropriate to this setting and anticipate that there will be many developments that further extend these tools to cover an increasing set of economically relevant settings. Ultimately, the practice of considering high-dimensional data more openly coupled with appropriate methodology should strengthen the plausibility of inferences drawn from data and allow a deeper exploration of the structures underlying economic data.

References

- Acemoglu, Daron, Simon Johnson, and James A. Robinson.** 2001. "The Colonial Origins of Comparative Development: An Empirical Investigation." *American Economic Review* 91(5): 1369–1401.
- Bekker, Paul A.** 1994. "Alternative Approximations to the Distributions of Instrumental Variable Estimators." *Econometrica* 63(3): 657–81.
- Belloni, Alexandre, Daniel Chen, Victor Chernozhukov, and Christian Hansen.** 2012. "Sparse Models and Methods for Optimal Instruments with an Application to Eminent Domain." *Econometrica* 80(6): 2369–2429.
- Belloni, Alexandre, and Victor Chernozhukov.** 2011. " ℓ_1 -Penalized Quantile Regression for High Dimensional Sparse Models." *Annals of Statistics* 39(1): 82–130.
- Belloni, Alexandre, and Victor Chernozhukov.** 2013. "Least Squares after Model Selection in High-Dimensional Sparse Models." *Bernoulli* 19(2): 521–47.
- Belloni, Alexandre, Victor Chernozhukov, Ivan Fernández-Val, and Christian Hansen.** 2013. "Program Evaluation with High-dimensional Data." arXiv e-print, arXiv:1311.2645.
- Belloni, Alexandre, Victor Chernozhukov, and Christian B. Hansen.** 2013. "Inference for High-Dimensional Sparse Econometric Models." *Advances in Economics and Econometrics. 10th World Congress, Vol. 3*, edited by Daron Acemoglu, Manuel Arellano, and Eddie Dekel, 245–95. Cambridge University Press.
- Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen.** Forthcoming. "Inference on Treatment Effects after Selection amongst High-Dimensional Controls." *Review of Economic Studies*.
- Belloni, Alexandre, Victor Chernozhukov, Christian Hansen, and Damian Kozbur.** 2014. "Inference in High-Dimensional Panel Data." Unpublished paper; available from the authors by request.
- Belloni, Alexandre, Victor Chernozhukov, and Kengo Kato.** 2013. "Uniform Post Selection Inference for LAD Regression and Other Z-Estimation Problems." arXiv e-print, arXiv:1304.0282.
- Belloni, Alexandre, Victor Chernozhukov, and Ying Wei.** 2013. "Honest Confidence Regions for Logistic Regression with a Large Number of Controls." arXiv e-print arXiv:1304.3969.
- Bickel, Peter J., Ya'acov Ritov, and Alexandre B. Tsybakov.** 2009. "Simultaneous Analysis of Lasso and Dantzig Selector." *Annals of Statistics* 37(4): 1705–32.
- Chao, John C., and Norman R. Swanson.** 2005. "Consistent Estimation with a Large Number of Weak Instruments." *Econometrica* 73(5): 1673–92.
- Chen, Daniel L., and Jasmin Sethi.** 2010. "Insiders and Outsiders: Does Forbidding Sexual Harassment Exacerbate Gender Inequality?" Available at SSRN: <http://ssrn.com/abstract=1119968>.
- Chen, Daniel L., and Susan Yeh.** 2012. "Growth under the Shadow of Expropriation? The Economic Impacts of Eminent Domain." <http://www.sole-jole.org/13463.pdf>.
- Chen, Xiaohong.** 2007. "Large Sample Sieve Estimation of Semi-Nonparametric Models." Chap. 76 in *Handbook of Econometrics*, Vol. 6B, edited by James J. Heckman and Edward E. Leamer, 5559–5632. Elsevier.
- Donohue, John J., III, and Steven D. Levitt.** 2001. "The Impact of Legalized Abortion on Crime." *Quarterly Journal of Economics* 116(2): 379–420.
- Donohue, John J., III, and Steven D. Levitt.** 2008. "Measurement Error, Legalized Abortion, and the Decline in Crime: A Response to Foote and Goetz." *Quarterly Journal of Economics* 123(1): 425–40.

- Farrell, Max H.** 2013. "Robust Inference on Average Treatment Effects with Possibly More Covariates than Observations." arXiv e-print, arXiv:1309.4686.
- Foote, Christopher L., and Christopher F. Goetz.** 2008. "The Impact of Legalized Abortion on Crime: Comment." *Quarterly Journal of Economics* 123(1): 407–23.
- Frank, Ildiko E., and Jerome H. Friedman.** 1993. "A Statistical View of Some Chemometrics Regression Tools." *Technometrics* 35(2): 109–135.
- Gautier, Eric, and Alexander B. Tsybakov.** 2011. "High-Dimensional Instrumental Variables Regression and Confidence Sets." Available at arXiv.
- Hansen, Christian, Jerry Hausman, and Whitney K. Newey.** 2008. "Estimation with Many Instrumental Variables." *Journal of Business and Economic Statistics* 26(4): 398–422.
- Hastie, Trevor, Robert Tibshirani, and Jerome Friedman.** 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer: New York, NY.
- Hausman, Jerry, Whitney K. Newey, Tiemen Woutersen, John Chao, and Norman Swanson.** 2012. "Instrumental Variable Estimation with Heteroskedasticity and Many Instruments." *Quantitative Economics* 3(2): 211–55.
- Kato, Kengo.** 2011. "Group Lasso for High Dimensional Sparse Quantile Regression Models." arXiv e-print, arXiv:1103.1458.
- Leeb, Hannes, and Benedikt M. Pötscher.** 2008a. "Can One Estimate the Unconditional Distribution of Post-Model-Selection Estimators?" *Econometric Theory* 24(2): 338–76.
- Leeb, Hannes and Benedikt M. Pötscher.** 2008b. "Recent Developments in Model Selection and Related Areas." *Econometric Theory* 24(2): 319–22.
- Newey, Whitney K.** 1997. "Convergence Rates and Asymptotic Normality for Series Estimators." *Journal of Econometrics* 79(1): 147–68.
- Robinson, Peter M.** 1988. "Root-N-Consistent Semiparametric Regression." *Econometrica* 56(4): 931–54.
- Serena Ng, and Jushan Bai.** 2009. "Selecting Instrumental Variables in a Data Rich Environment." *Journal of Time Series Econometrics* 1(1).
- Tibshirani, Rob.** 1996. "Regression Shrinkage and Selection via the Lasso." *Journal of the Royal Statistical Society Series B* 58(1): 267–88.
- van de Geer, Sara A.** 2008. "High-Dimensional Generalized Linear Models and the LASSO." *Annals of Statistics* 36(2): 614–45.

This article has been cited by:

1. Jorge M. Uribe. 2025. Investment in intangible assets and economic complexity. *Research Policy* **54**:1, 105133. [[Crossref](#)]
2. Denvil R. Duncan, Justin M. Ross, John L. Mikesell. 2025. Can the Federal Budget Process Promote Fiscal Sustainability? Evidence from the Budget Enforcement Act of 1990. *Public Finance Review* **53**:1, 29-61. [[Crossref](#)]
3. Milos Ciganovic, Federico D'Amario. 2024. Forecasting cryptocurrencies log-returns: a LASSO-VAR and sentiment approach. *Applied Economics* **56**:58, 8112-8138. [[Crossref](#)]
4. Berardino Barile, Marco Forti, Alessia Marrocco, Angelo Castaldo. 2024. Causal impact evaluation of occupational safety policies on firms' default using machine learning uplift modelling. *Scientific Reports* **14**:1. . [[Crossref](#)]
5. Asror Nigmonov, Syed Shams, Povilas Urbonas. 2024. Estimating probability of default via delinquencies? Evidence from European P2P lending market. *Global Finance Journal* **63**, 101050. [[Crossref](#)]
6. Alessandro V.M. Oliveira, Luca J. Santos, Dante Mendes Aldrighi. 2024. A hybrid econometric-machine learning framework to support market development in intercity passenger transport: the causal and predictive analytics of economic mobility features. *Journal of the Air Transport Research Society* **3**, 100043. [[Crossref](#)]
7. Anastasiya-Mariya Asanov, Igor Asanov, Guido Buenstorf. 2024. A low-cost digital first aid tool to reduce psychological distress in refugees: A multi-country randomized controlled trial of self-help online in the first months after the invasion of Ukraine. *Social Science & Medicine* **362**, 117442. [[Crossref](#)]
8. Qingliang Fan, Yaqian Wu. 2024. Endogenous Treatment Effect Estimation with a Large and Mixed Set of Instruments and Control Variables. *Review of Economics and Statistics* **106**:6, 1655-1674. [[Crossref](#)]
9. Sara B. Heller, Judd B. Kessler. 2024. Information Frictions and Skill Signaling in the Youth Labor Market. *American Economic Journal: Economic Policy* **16**:4, 1-33. [[Abstract](#)] [[View PDF article](#)] [[PDF with links](#)]
10. Gyuwon Jung, Sangjun Park, Eun-Yeol Ma, Heeyoung Kim, Uichin Lee. 2024. Tutorial on Matching-based Causal Analysis of Human Behaviors Using Smartphone Sensor Data. *ACM Computing Surveys* **56**:9, 1-33. [[Crossref](#)]
11. Wenchao Li. 2024. Do surging house prices discourage fertility? Global evidence, 1870-2012. *Labour Economics* **90**, 102572. [[Crossref](#)]
12. Anna Massfeller, Hugo Storm. 2024. Field observation and verbal exchange as different peer effects in farmers' technology adoption decisions. *Agricultural Economics* **55**:5, 739-757. [[Crossref](#)]
13. Xiaotong Sun, Mi Zhou, Li Huang. 2024. The impact of family urban integration on migrant worker mental health in China. *Frontiers in Public Health* **12**. . [[Crossref](#)]
14. Cigdem Kosar Tas, Huseyin Guler, Yeliz Yalcin. 2024. The usage of bridge estimator to determine the order of integration for possibly integrated series as an alternative to Dickey-Pantula unit root test. *Communications in Statistics - Simulation and Computation* **53**:8, 3916-3932. [[Crossref](#)]
15. Clarissa Lotti, Arieda Muço, Giancarlo Spagnolo, Tommaso Valletti. 2024. Indirect Savings from Public Procurement Centralization. *American Economic Journal: Economic Policy* **16**:3, 347-366. [[Abstract](#)] [[View PDF article](#)] [[PDF with links](#)]

16. Nils Bohr, Tim Deisemann, Douglas Gollin, Frederic Kosmowski, Travis J. Lybbert. 2024. The seeds of misallocation: Fertilizer use and maize varietal misidentification in Ethiopia. *Journal of Development Economics* 7, 103349. [[Crossref](#)]
17. Bo Cowgill, Patryk Perkowski. 2024. Delegation in Hiring: Evidence from a Two-Sided Audit. *Journal of Political Economy Microeconomics* 279. . [[Crossref](#)]
18. Andreas Ferrara, Price Fishback. 2024. Discrimination, Migration, and Economic Outcomes: Evidence from World War I. *Review of Economics and Statistics* 1-19. [[Crossref](#)]
19. Dominick Bartelme, Ting Lan, Andrei A. Levchenko. 2024. Specialization, market access and real income. *Journal of International Economics* 150, 103923. [[Crossref](#)]
20. Girum Abebe, Mulu Gebreyesus. 2024. How Well Do Worker Cooperatives Perform in Manufacturing? Evidence from a Large Low-Income Country. *Economic Development and Cultural Change* 72:4, 1993-2029. [[Crossref](#)]
21. Jorge Guzman. 2024. Go West Young Firm: The Impact of Startup Migration on the Performance of Migrants. *Management Science* 70:7, 4824-4846. [[Crossref](#)]
22. Issa J. Dahabreh, Kirsten Bibbins-Domingo. 2024. Causal Inference About the Effects of Interventions From Observational Studies in Medical Journals. *JAMA* 331:21, 1845. [[Crossref](#)]
23. Marco Testoni. 2024. Transportation networks and competition in the market for corporate control. *Strategic Management Journal* 45:6, 1180-1208. [[Crossref](#)]
24. Yangjie Wang, Junyi Feng, Riazullah Shinwari, Elie Bouri. 2024. Do green finance and green innovation affect corporate credit rating performance? Evidence from machine learning approach. *Journal of Environmental Management* 360, 121212. [[Crossref](#)]
25. Jia Li, Jinming Shi, Ruihan Cao, Jingyi Wu, Jianxu Liu. 2024. How e-commerce can boost China's high-quality agricultural exports. *Frontiers in Sustainable Food Systems* 8. . [[Crossref](#)]
26. Jason Rathje, Riitta Katila, Philipp Reineke. 2024. Making the most of AI and machine learning in organizations and strategy research: Supervised machine learning, causal inference, and matching models. *Strategic Management Journal* . [[Crossref](#)]
27. Abhijit Banerjee, Emily Breza, Arun G Chandrasekhar, Esther Duflo, Matthew O Jackson, Cynthia Kinnan. 2024. Changes in Social Network Structure in Response to Exposure to Formal Credit Markets. *Review of Economic Studies* 91:3, 1331-1372. [[Crossref](#)]
28. Wei Jin, Frederick van der Ploeg, Lin Zhang. 2024. How clean capital slows down disinvestment of carbon-intensive capital in the low-carbon transition. *Journal of Economic Dynamics and Control* 162, 104857. [[Crossref](#)]
29. Shilan Feng, Ya Tan, Zhi Wang, Qinghua Zhang. 2024. Spatial Structure of Peer Networks and Academic Achievement Based on a Random Control Trial Experiment. *China & World Economy* 32:3, 67-97. [[Crossref](#)]
30. Paul Bose, Lorenzo Lupo, Mahyar Habibi, Dirk Hovy, Carlo Schwarz. 2024. Beyond the Stats: Realities, Perception, and Social Media Discourse on Poverty. *AEA Papers and Proceedings* 114, 690-694. [[Abstract](#)] [[View PDF article](#)] [[PDF with links](#)]
31. Lorena Cecilia López Steinmetz, Margarita Sison, Rustam Zhumagambetov, Juan Carlos Godoy, Stefan Haufe. 2024. Machine learning models predict the emergence of depression in Argentinean college students during periods of COVID-19 quarantine. *Frontiers in Psychiatry* 15. . [[Crossref](#)]
32. Carolina Lopez, Anja Sautmann, Simone Schaner. Do Patients Value High-Quality Medical Care? Experimental Evidence from Malaria Diagnosis and Treatment 53, . [[Crossref](#)]
33. David E. Broockman, Joshua L. Kalla. 2024. Consuming cross-cutting media causes learning and moderates attitudes: A field experiment with Fox News viewers. *The Journal of Politics* 279. . [[Crossref](#)]

34. Sergio Perilla, Mounu Prem, Miguel E. Purroy, Juan F. Vargas. 2024. How peace saves lives: Evidence from Colombia. *World Development* **176**, 106529. [[Crossref](#)]
35. Tobias Renkin, Gabriel Züllig. 2024. Credit Supply Shocks and Prices: Evidence from Danish Firms. *American Economic Journal: Macroeconomics* **16**:2, 1-28. [[Abstract](#)] [[View PDF article](#)] [[PDF with links](#)]
36. Jana Bilger, Mark Pletscher, Tobias Müller. 2024. Separating the wheat from the chaff: How to measure hospital quality in routine data?. *Health Services Research* **59**:2. . [[Crossref](#)]
37. Jens Ludwig, Sendhil Mullainathan. 2024. Machine Learning as a Tool for Hypothesis Generation. *The Quarterly Journal of Economics* **139**:2, 751-827. [[Crossref](#)]
38. Vivi Alatas, Arun G Chandrasekhar, Markus Mobius, Benjamin A Olken, Cindy Paladines. 2024. Do Celebrity Endorsements Matter? A Twitter Experiment Promoting Vaccination in Indonesia. *The Economic Journal* **134**:659, 913-933. [[Crossref](#)]
39. Hazel H. Dadanlar, Manisha M. Vaswani, Marwan Al-Shammari, Soumendra Nath Banerjee. 2024. Firm performance feedback and organizational impression management: The moderating role of CEO overconfidence. *Journal of Management & Organization* **38**, 1-24. [[Crossref](#)]
40. Vignesh Yoganathan, Victoria-Sophie Osburg. 2024. The mind in the machine: Estimating mind perception's effect on user satisfaction with voice-based conversational agents. *Journal of Business Research* **175**, 114573. [[Crossref](#)]
41. Isaac K. Ofori, Camara K. Obeng, Simplicie A. Asongu. 2024. What Really Drives Economic Growth in Sub-Saharan Africa? Evidence from the Lasso Regularization and Inferential Techniques. *Journal of the Knowledge Economy* **15**:1, 144-179. [[Crossref](#)]
42. Jonathan B. Hill. 2024. Testing Many Zero Restrictions in a High Dimensional Linear Regression Setting. *Journal of Business & Economic Statistics* 1-29. [[Crossref](#)]
43. Ricardo Masini. 2024. Distributional counterfactual analysis in high-dimensional setup. *Journal of Econometrics* **59**, 105675. [[Crossref](#)]
44. Jan Gromadzki, Katarzyna Sałach, Michał Brzeziński. 2024. When populists deliver on their promises: the electoral effects of a large cash transfer programme in Poland. *Economica* **91**:361, 320-345. [[Crossref](#)]
45. Asror Nigmonov, Syed Shams, Khorshed Alam. 2024. Liquidity risk in FinTech lending: Early impact of the COVID-19 pandemic on the P2P lending market. *Emerging Markets Review* **58**, 101084. [[Crossref](#)]
46. Emil Kupek, Rafaela Liberali. 2024. Food patterns associated with overweight in 7-11-year old children: machine-learning approach. *Ciência & Saúde Coletiva* **29**:1. . [[Crossref](#)]
47. Karolis Matikonis, Matthew Gobey. 2024. Small Business Property Tax Reductions and Firm Productivity. *Small Business Economics* **62**:1, 307-324. [[Crossref](#)]
48. Patricia Kenny, Dan Liu, Denzil Fiebig, Jane Hall, Jared Millican, Sanchia Aranda, Kees van Gool, Philip Haywood. 2024. Specialist Palliative Care and Health Care Costs at the End of Life. *Pharmacoeconomics - Open* **8**:1, 31-47. [[Crossref](#)]
49. Nils Bohr, Tim Deisemann, Douglas Gollin, Frederic Kosmowski, Travis J. Lybbert. 2024. The Seeds of Misallocation: Fertilizer Use and Maize Varietal Misidentification in Ethiopia. *SSRN Electronic Journal* **7**. . [[Crossref](#)]
50. Yuxin Li, Xiaoyin Ma, Luc Renneboog. 2024. In Art We Trust. *Management Science* **70**:1, 98-127. [[Crossref](#)]
51. Masahiro Kato. 2024. Triple/Debiased Lasso for Statistical Inference of Conditional Average Treatment Effects. *SSRN Electronic Journal* **33**. . [[Crossref](#)]

52. Ines Chaieb, Vihang R. Errunza, Lucie Yiliu Lu. 2024. Sustainable Investing Home and Abroad. *SSRN Electronic Journal* **74**. . [[Crossref](#)]
53. Christine L Exley, Raymond Fisman, Judd B. Kessler, Louis Pierre Lepage, Xiaomeng Li, Corinne Low, Xiaoyue Shan, Mattie Toma, Basit Zafar. 2024. Information-Optional Policies and the Gender Concealment Gap. *SSRN Electronic Journal* **125**. . [[Crossref](#)]
54. Ozge Demirci. 2024. Can Gender-Blind Algorithmic Pricing Eliminate the Gender Gap?. *SSRN Electronic Journal* **133**. . [[Crossref](#)]
55. Ahmed Elnahas. 2024. Can Machines Better Predict Insider Trading?. *SSRN Electronic Journal* **55**. . [[Crossref](#)]
56. Nicholas Otis, Rowan Clarke, Solène Delecourt, David Holtz, Rembrand Koning. 2024. The Uneven Impact of Generative AI on Entrepreneurial Performance. *SSRN Electronic Journal* **381**. . [[Crossref](#)]
57. Felipe González, Francisco Muñoz, Cristine von Dessauer. 2024. Empowerment or Indoctrination? Female Training Programs Under Dictatorship. *SSRN Electronic Journal* **112**. . [[Crossref](#)]
58. Ines Chaieb, Vihang R. Errunza, Lucie Yiliu Lu. 2024. Who Invests in What? Public Firms Ownership Around the World. *SSRN Electronic Journal* . [[Crossref](#)]
59. Alexander Koplenig, Sascha Wolfer. 2023. Languages with more speakers tend to be harder to (machine-)learn. *Scientific Reports* **13**:1. . [[Crossref](#)]
60. Alessandro V.M. Oliveira, Bruno F. Oliveira, Moisés D. Vassallo. 2023. Airport service quality perception and flight delays: Examining the influence of psychosituational latent traits of respondents in passenger satisfaction surveys. *Research in Transportation Economics* **102**, 101371. [[Crossref](#)]
61. Hoa Luong, Mehdi Khedmati, Lan Anh Nguyen, Asror Nigmonov, Nafisa Zabeen Ovi, Syed Shams. 2023. CEO-director ties and board gender diversity: US evidence. *Journal of Behavioral and Experimental Finance* **40**, 100861. [[Crossref](#)]
62. Anabela M. Santos, Alex Coad. 2023. Monitoring and evaluation of transformative innovation policy: Suggestions for Improvement. *Socio-Economic Planning Sciences* **90**, 101714. [[Crossref](#)]
63. Aleksandr Tomtosov. 2023. Overlapping portfolio holdings and unique sources of emerging market risk. *Borsa Istanbul Review* **68**. . [[Crossref](#)]
64. Heejun Shin, Joseph Antonelli. 2023. Improved Inference for Doubly Robust Estimators of Heterogeneous Treatment Effects. *Biometrics* **79**:4, 3140-3152. [[Crossref](#)]
65. Dahai Yue, Ninez A. Ponce, Jack Needleman, Susan L. Ettner, Adriana Lleras-Muney. 2023. Increased Schooling Reduces Hospitalization Later in Life: New Evidence with Optimal Instruments from the United States. *American Journal of Health Economics* **113**. . [[Crossref](#)]
66. Fabiano Santos, Thiago Moreira. 2023. Agribusiness, Legislative Elections and Party Fragmentation in Brazil. *Bulletin of Latin American Research* **42**:5, 678-691. [[Crossref](#)]
67. Anastasiya-Mariya Asanov, Igor Asanov, Guido Buenstorf. 2023. Mental health and stress level of Ukrainians seeking psychological help online. *Heliyon* **9**:11, e21933. [[Crossref](#)]
68. Jacopo Bonan, Giovanna d'Adda, Mahreen Mahmud, Farah Said. 2023. Nudging Payment Behavior: Evidence from a Field Experiment on Pay-as-You-Go Off-Grid Electricity. *The World Bank Economic Review* **37**:4, 620-639. [[Crossref](#)]
69. Michael Lokshin, Martin Ravallion, Iván Torre. 2023. The Social Protection Engel Curve. *The World Bank Economic Review* **37**:4, 519-548. [[Crossref](#)]
70. Stefano Caria, Simon Franklin, Marc Witte. 2023. Searching with Friends. *Journal of Labor Economics* **41**:4, 887-922. [[Crossref](#)]
71. Jianan Chen, Binyan Jiang, Jialiang Li. 2023. Nonparametric instrument model averaging. *Journal of Nonparametric Statistics* **35**:4, 905-926. [[Crossref](#)]

72. Raf Van Gestel, Niels Broekman, Tobias Müller. 2023. Surgeon supply and healthcare quality: Are revision rates for hip and knee replacements lower in hospitals that employ more surgeons?. *Health Economics* 32:10, 2298-2321. [[Crossref](#)]
73. Travis Baseler. 2023. Hidden Income and the Perceived Returns to Migration. *American Economic Journal: Applied Economics* 15:4, 321-352. [[Abstract](#)] [[View PDF article](#)] [[PDF with links](#)]
74. Leander Andres, Marc Fabel, Helmut Rainer. 2023. How much violence does football hooliganism cause?. *Journal of Public Economics* 225, 104970. [[Crossref](#)]
75. Lelys Dinarte-Diaz, Maria Marta Ferreyra, Sergio Urzua, Marina Bassi. 2023. What makes a program good? Evidence from short-cycle higher education programs in five developing countries. *World Development* 169, 106294. [[Crossref](#)]
76. John Tazare, Liam Smeeth, Stephen J. W. Evans, Ian J. Douglas, Elizabeth J. Williamson. 2023. hdsps: A suite of commands for applying high-dimensional propensity-score approaches. *The Stata Journal: Promoting communications on statistics and Stata* 23:3, 683-708. [[Crossref](#)]
77. Jad Beyhum. 2023. High-dimensional inference robust to outliers with #1-norm penalization. *Communications in Statistics - Theory and Methods* 52:16, 5866-5876. [[Crossref](#)]
78. Erkan Gören, Adalbert Winkler. 2023. Statistical Capacity Matters: The Long-Term Effects of Africa's Slave Trade on Development Reflected by Nighttime Light Intensity. *Journal of African Economics* 32:4, 383-414. [[Crossref](#)]
79. Benjamin L. Hallen, Susan L. Cohen, Sung Ho Park. 2023. Are seed accelerators status springboards for startups? Or sand traps?. *Strategic Management Journal* 44:8, 2060-2096. [[Crossref](#)]
80. Jennie E. Brand, Xiang Zhou, Yu Xie. 2023. Recent Developments in Causal Inference and Machine Learning. *Annual Review of Sociology* 49:1, 81-110. [[Crossref](#)]
81. Zijian Guo. 2023. Statistical Inference for Maximin Effects: Identifying Stable Associations across Multiple Studies. *Journal of the American Statistical Association* 1-32. [[Crossref](#)]
82. Sylvia Omulo, Teresa Ita, Robert Mugoh, Charchil Ayodo, Ulzii Luvsansharav, Susan Bollinger, Ashley Styczynski, Brooke M Ramay, Mark A Caudell, Guy H Palmer, Samuel Kariuki, Douglas R Call, Rachel M Smith. 2023. Risk Factors for Colonization With Extended-Spectrum Cephalosporin-Resistant and Carbapenem-Resistant Enterobacterales Among Hospitalized Patients in Kenya: An Antibiotic Resistance in Communities and Hospitals (ARCH) Study. *Clinical Infectious Diseases* 77:Supplement_1, S97-S103. [[Crossref](#)]
83. Román Andrés Zárate. 2023. Uncovering Peer Effects in Social and Academic Skills. *American Economic Journal: Applied Economics* 15:3, 35-79. [[Abstract](#)] [[View PDF article](#)] [[PDF with links](#)]
84. Clément de Chaisemartin, Nicolás Navarrete H.. 2023. The Direct and Spillover Effects of a Nationwide Socioemotional Learning Program for Disruptive Students. *Journal of Labor Economics* 41:3, 729-769. [[Crossref](#)]
85. 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](#)]
86. Shishir Shakya, Jane E. Ruseski. 2023. The effect of Prescription Drug Monitoring Programs on county-level opioid prescribing practices and spillovers. *Contemporary Economic Policy* 41:3, 435-454. [[Crossref](#)]
87. Francesca Micocci, Armando Rungi. 2023. Predicting Exporters with Machine Learning. *World Trade Review* 29, 1-24. [[Crossref](#)]
88. Alain Hecq, Luca Margaritella, Stephan Smeekes. 2023. Granger Causality Testing in High-Dimensional VARs: A Post-Double-Selection Procedure. *Journal of Financial Econometrics* 21:3, 915-958. [[Crossref](#)]

89. Pablo J Ordóñez, Kathy Baylis, M Isabel Ramírez. 2023. Land cover change effects from community forest management in Michoacán, Mexico. *Environmental Research Letters* **18**:6, 065008. [[Crossref](#)]
90. Masayuki Hirukawa, Di Liu, Irina Murtazashvili, Artem Prokhorov. 2023. DS-HECK: double-lasso estimation of Heckman selection model. *Empirical Economics* **64**:6, 3167-3195. [[Crossref](#)]
91. Lena Chan, Courtney Simmons, Scott Tillem, May Conley, Inti A. Brazil, Arielle Baskin-Sommers. 2023. Classifying Conduct Disorder Using a Biopsychosocial Model and Machine Learning Method. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* **8**:6, 599-608. [[Crossref](#)]
92. Andre Horovitz. 2023. Ascertaining the Inference of Bank Internal Default Probabilities Variations on Variable Rate Institutional Loan Prepayments. *The Quarterly Journal of Finance* **13**:02. . [[Crossref](#)]
93. Stephen Martin, Karl Claxton, James Lomas, Francesco Longo. 2023. The impact of different types of NHS expenditure on health: Marginal cost per QALY estimates for England for 2016/17. *Health Policy* **132**, 104800. [[Crossref](#)]
94. Yuya Sasaki, Takuya Ura. 2023. Estimation and inference for policy relevant treatment effects. *Journal of Econometrics* **234**:2, 394-450. [[Crossref](#)]
95. Carolina Bernal, Mounu Prem, Juan F. Vargas, Mónica Ortiz. 2023. Peaceful entry: Entrepreneurship dynamics during Colombia's peace agreement. *Journal of Development Economics* **72**, 103119. [[Crossref](#)]
96. Paul Hünermund, Beyers Louw, Itamar Caspi. 2023. Double machine learning and automated confounder selection: A cautionary tale. *Journal of Causal Inference* **11**:1. . [[Crossref](#)]
97. Annamaria Conti, Jorge A. Guzman. 2023. What Is the US Comparative Advantage in Entrepreneurship? Evidence from Israeli Migration to the United States. *Review of Economics and Statistics* **105**:3, 528-544. [[Crossref](#)]
98. Nils A. Breitmar, Matthew Harding, Carlos Lamarche. 2023. Using Grouped Data to Estimate Revenue Heterogeneity in Online Advertising Auctions. *AEA Papers and Proceedings* **113**, 161-165. [[Abstract](#)] [[View PDF article](#)] [[PDF with links](#)]
99. Nebojsa Dimic, John W. Goodell, Vanja Piljak, Milos Vulcanovic. 2023. Acquisition determinants of energy SPACs: Reflecting a closed group?. *Finance Research Letters* **50**, 104001. [[Crossref](#)]
100. Yuqian Zhang, Abhishek Chakraborty, Jelena Bradic. 2023. Double robust semi-supervised inference for the mean: selection bias under MAR labeling with decaying overlap. *Information and Inference: A Journal of the IMA* **12**:3, 2066-2159. [[Crossref](#)]
101. Xiaolan Yang, Yidong Huang, Wenchao Li. 2023. Economic preferences and state-level corruption: global evidence and mechanism. *Applied Economics Letters* **30**:7, 892-912. [[Crossref](#)]
102. Zhewen Pan, Jianhui Xie. 2023. -Penalized Pairwise Difference Estimation for a High-Dimensional Censored Regression Model. *Journal of Business & Economic Statistics* **41**:2, 283-297. [[Crossref](#)]
103. Pierre C. Bellec, Cun-Hui Zhang. 2023. Debiasing convex regularized estimators and interval estimation in linear models. *The Annals of Statistics* **51**:2. . [[Crossref](#)]
104. Barbara Felderer, Jannis Kueck, Martin Spindler. 2023. Using Double Machine Learning to Understand Nonresponse in the Recruitment of a Mixed-Mode Online Panel. *Social Science Computer Review* **41**:2, 461-481. [[Crossref](#)]
105. Hannah Uckat. Leaning in at Home: Women's Promotions and Intra-Household Bargaining in Bangladesh **57**, . [[Crossref](#)]
106. Wenchao Li, Zhiming Leng, Junjian Yi, Songfa Zhong. 2023. A multifaceted poverty reduction program has economic and behavioral consequences. *Proceedings of the National Academy of Sciences* **120**:10. . [[Crossref](#)]
107. Hui-Ching Chuang, Jau-er Chen. 2023. Exploring Industry-Distress Effects on Loan Recovery: A Double Machine Learning Approach for Quantiles. *Econometrics* **11**:1, 6. [[Crossref](#)]

108. Maria P. Roche. 2023. Academic Entrepreneurship: Entrepreneurial Advisors and Their Advisees' Outcomes. *Organization Science* 34:2, 959-986. [[Crossref](#)]
109. Oliver Buchholz. 2023. The deep neural network approach to the reference class problem. *Synthese* 201:3. . [[Crossref](#)]
110. Jacob H. Hansen, Mathias V. Siggaard. 2023. Double Machine Learning: Explaining the Post-Earnings Announcement Drift. *Journal of Financial and Quantitative Analysis* 75, 1-28. [[Crossref](#)]
111. Sauveur Giannoni, Sylvain Petit, Louinord Voltaire, Philippe Chagnon, Jean Ralph Marvensky Paul. 2023. Adoption of Covid-19 safety certification and pricing strategy in the hotel industry. *Applied Economics* 12, 1-25. [[Crossref](#)]
112. Chaohua Dong, Jiti Gao, Oliver Linton. 2023. High dimensional semiparametric moment restriction models. *Journal of Econometrics* 232:2, 320-345. [[Crossref](#)]
113. Dirk Witteveen. 2023. The Effects of Undergraduate Financing on Advanced Degree Attainment. *Social Forces* 101:3, 1258-1287. [[Crossref](#)]
114. Shutter Zor. 2023. Conservation or revolution? The sustainable transition of textile and apparel firms under the environmental regulation: Evidence from China. *Journal of Cleaner Production* 382, 135339. [[Crossref](#)]
115. Isaac Kwesi Ofori, Camara Kwasi Obeng. What Really Drives Financial Sector Development in Ghana? 13-53. [[Crossref](#)]
116. Matthew C. Dickson, Michael M. Nguyen, Charmi Patel, Shannon C. Grabich, Carmela Benson, Terry Cothran, Grant H. Skrepnek. 2023. Adherence, Persistence, Readmissions, and Costs in Medicaid Members with Schizophrenia or Schizoaffective Disorder Initiating Paliperidone Palmitate Versus Switching Oral Antipsychotics: A Real-World Retrospective Investigation. *Advances in Therapy* 40:1, 349-366. [[Crossref](#)]
117. Yunwei Chen, Yuju Wu, Sarah-Eve Dill, Yian Guo, Christopher Michael Westgard, Alexis Medina, Ann M Weber, Gary L Darmstadt, Huan Zhou, Scott Rozelle, Sean Sylvia. 2023. Effect of the mHealth-supported Healthy Future programme delivered by community health workers on maternal and child health in rural China: study protocol for a cluster randomised controlled trial. *BMJ Open* 13:1, e065403. [[Crossref](#)]
118. Jessica Ordemann, Frauke Peter, Sandra Buchholz. Vielfalt von hochschulischen Bildungsverläufen – neue Aspekte, Erkenntnisse und Entwicklungen 1-20. [[Crossref](#)]
119. Gunther Dahm, Frauke Peter. Einfach anders oder vielfältig verschieden? Ein differenzierter Blick auf Hochschulabsolvent*innen mit beruflicher Vorqualifikation 223-262. [[Crossref](#)]
120. Brett R. Gordon, Mitchell J. Lovett, Bowen Luo, James C. Reeder. 2023. Disentangling the Effects of Ad Tone on Voter Turnout and Candidate Choice in Presidential Elections. *Management Science* 69:1, 220-243. [[Crossref](#)]
121. Matthias Schonlau. Lasso and Friends 73-95. [[Crossref](#)]
122. Serkan Aglasan, Barry K. Goodwin, Roderick M. Rejesus. 2023. Risk effects of GM corn: Evidence from crop insurance outcomes and high-dimensional methods. *Agricultural Economics* 54:1, 110-126. [[Crossref](#)]
123. Ágnes Szabó-Morvai, Daniel Horn, Anna Lovász, Kristof De Witte. 2023. Universal preschool and cognitive skills – the role of school starting age as a moderating factor. *Early Childhood Research Quarterly* 64, 278-289. [[Crossref](#)]
124. Giovanni Cerulli. Model Selection and Regularization 59-146. [[Crossref](#)]
125. Felipe González, Mounu Prem, Cristine von Dessauer. 2023. Empowerment or Indoctrination? Women Centers Under Dictatorship. *SSRN Electronic Journal* 112. . [[Crossref](#)]

126. Berkay Akyapı. 2023. Machine learning and feature selection: Applications in economics and climate change. *Environmental Data Science* 2. . [[Crossref](#)]
127. Thiago Fonseca Morello. 2023. Hospitalization due to fire-induced pollution in the Brazilian Amazon: A causal inference analysis with an assessment of policy trade-offs. *World Development* 161, 106123. [[Crossref](#)]
128. Jiaming Zeng, Michael F. Gensheimer, Daniel L. Rubin, Susan Athey, Ross D. Shachter. 2022. Uncovering interpretable potential confounders in electronic medical records. *Nature Communications* 13:1. . [[Crossref](#)]
129. Markku Kaustia, Andrew Conlin, Niilo Luotonen. 2022. What drives stock market participation? The role of institutional, traditional, and behavioral factors. *Journal of Banking & Finance* 63, 106743. [[Crossref](#)]
130. Anne C. Wunderlich, Andreas Kohler. 2022. The impact of agricultural subsidies on obesity. *Journal of Agriculture and Food Research* 10, 100439. [[Crossref](#)]
131. Lenard Lieb, Johannes Schuffels. 2022. Inflation expectations and consumer spending: the role of household balance sheets. *Empirical Economics* 63:5, 2479-2512. [[Crossref](#)]
132. Thomas M. Cascino, Jeffrey S. McCullough, Xiaoting Wu, Michael J. Pienta, James W. Stewart, Robert B. Hawkins, Alexander A. Brescia, Ashraf Abou el ala, Min Zhang, Pierre-Emmanuel Noly, Jonathan W. Haft, Jennifer A. Cowger, Monica Colvin, Keith D. Aaronson, Francis D. Pagani, Donald S. Likosky, Khalil M Nassar, Paul Tang, Grace Chung, Tessa Watt, Michael P Thompson, Allison M Janda. 2022. Comparison of Evaluations for Heart Transplant Before Durable Left Ventricular Assist Device and Subsequent Receipt of Transplant at Transplant vs Nontransplant Centers. *JAMA Network Open* 5:11, e2240646. [[Crossref](#)]
133. MARC RATKOVIC. 2022. Relaxing Assumptions, Improving Inference: Integrating Machine Learning and the Linear Regression. *American Political Science Review* 22, 1-17. [[Crossref](#)]
134. Mu Yue, Jialiang Li, Baoluo Sun. 2022. Conditional sparse boosting for high-dimensional instrumental variable estimation. *Journal of Statistical Computation and Simulation* 92:15, 3087-3108. [[Crossref](#)]
135. Andreas Ferrara. 2022. World War II and Black Economic Progress. *Journal of Labor Economics* 40:4, 1053-1091. [[Crossref](#)]
136. G. Puga, K. Anderson, F. Doko Tchatoka. 2022. Impact of growing season temperature on grape prices in Australia. *Australian Journal of Grape and Wine Research* 28:4, 651-657. [[Crossref](#)]
137. Giuseppe De Luca, Jan R. Magnus, Franco Peracchi. 2022. Sampling properties of the Bayesian posterior mean with an application to WALS estimation. *Journal of Econometrics* 230:2, 299-317. [[Crossref](#)]
138. Marina Bonaccolto-Töpfer, Stephanie Briel. 2022. The gender pay gap revisited: Does machine learning offer new insights?. *Labour Economics* 78, 102223. [[Crossref](#)]
139. Randy Carter, Netsanet Michael. 2022. Factor Analysis Regression for Predictive Modeling with High-Dimensional Data. *Journal of Quantitative Economics* 20:S1, 115-132. [[Crossref](#)]
140. J. Wesley Burnett, L. Lynne Kiesling. 2022. How do machines predict energy use? Comparing machine learning approaches for modeling household energy demand in the United States. *Energy Research & Social Science* 91, 102715. [[Crossref](#)]
141. Britta Augsburg, Antonella Bancalari, Zara Durrani, Madhav Vaidyanathan, Zach White. 2022. When nature calls back: Sustaining behavioral change in rural Pakistan. *Journal of Development Economics* 158, 102933. [[Crossref](#)]
142. William F Maloney, Felipe Valencia Caicedo. 2022. Engineering Growth. *Journal of the European Economic Association* 20:4, 1554-1594. [[Crossref](#)]

143. Maryyam Haroon, Farah Said, Mahniya Zafar. 2022. Fostering non-cognitive skills and academic performance: Experimental evidence from women's-only colleges in Pakistan. *Journal of Asian Economics* **81**, 101499. [[Crossref](#)]
144. Eva O. Arceo-Gomez, Raymundo M. Campos-Vazquez. 2022. Gender Bias in Evaluation Processes. *Economics of Education Review* **89**, 102272. [[Crossref](#)]
145. SASKIA TER ELLEN, VEGARD H. LARSEN, LEIF ANDERS THORSRUD. 2022. Narrative Monetary Policy Surprises and the Media. *Journal of Money, Credit and Banking* **54**:5, 1525-1549. [[Crossref](#)]
146. Akbar Zamanzadeh, Tony Cavoli. 2022. The effect of nonpharmaceutical interventions on COVID-19 infections for lower and middle-income countries: A debiased LASSO approach. *PLOS ONE* **17**:7, e0271586. [[Crossref](#)]
147. 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](#)]
148. Harold D. Chiang, Kengo Kato, Yukun Ma, Yuya Sasaki. 2022. Multiway Cluster Robust Double/Debiased Machine Learning. *Journal of Business & Economic Statistics* **40**:3, 1046-1056. [[Crossref](#)]
149. Elena Esposito. 2022. The Side Effects of Immunity: Malaria and African Slavery in the United States. *American Economic Journal: Applied Economics* **14**:3, 290-328. [[Abstract](#)] [[View PDF article](#)] [[PDF with links](#)]
150. Michael Bailey, Drew Johnston, Theresa Kuchler, Johannes Stroebel, Arlene Wong. 2022. Peer Effects in Product Adoption. *American Economic Journal: Applied Economics* **14**:3, 488-526. [[Abstract](#)] [[View PDF article](#)] [[PDF with links](#)]
151. Anthony Bald, Eric Chyn, Justine Hastings, Margarita Machelett. 2022. The Causal Impact of Removing Children from Abusive and Neglectful Homes. *Journal of Political Economy* **130**:7, 1919-1962. [[Crossref](#)]
152. Bart S. Vanneste, Ranjay Gulati. 2022. Generalized Trust, External Sourcing, and Firm Performance in Economic Downturns. *Organization Science* **33**:4, 1599-1619. [[Crossref](#)]
153. Arthur Acolin, Ari Decter-Frain, Matt Hall. 2022. Small-area estimates from consumer trace data. *Demographic Research* **47**, 843-882. [[Crossref](#)]
154. Marco Letta, Pierluigi Montalbano, Adriana Paolantonio. Understanding the Climate Change-Migration Nexus through the Lens of Household Surveys: An Empirical Review to Assess Data Gaps **14**, . [[Crossref](#)]
155. Natalia Zuniga-Garcia, Mauricio Tec, James G. Scott, Randy B. Machemehl. 2022. Evaluation of e-scooters as transit last-mile solution. *Transportation Research Part C: Emerging Technologies* **139**, 103660. [[Crossref](#)]
156. DEKUWMINI MORNAH, RAYMOND MACDERMOTT. 2022. INSTITUTIONS, CULTURE, AND ENTREPRENEURSHIP: DIRECT, TOTAL, AND COMPLIANCE BURDEN EFFECTS. *Global Economy Journal* **22**:02. . [[Crossref](#)]
157. Alessandro V.M. Oliveira, Thiago Caliari, Rodolfo R. Narcizo. 2022. An empirical model of fleet modernization: On the relationship between market concentration and innovation adoption by airlines. *Research in Transportation Business & Management* **43**, 100704. [[Crossref](#)]
158. Craig McIntosh, Andrew Zeitlin. 2022. Using household grants to benchmark the cost effectiveness of a USAID workforce readiness program. *Journal of Development Economics* **157**, 102875. [[Crossref](#)]

159. Sabrin Beg, Waqas Halim, Adrienne M. Lucas, Umar Saif. 2022. Engaging Teachers with Technology Increased Achievement, Bypassing Teachers Did Not. *American Economic Journal: Economic Policy* 14:2, 61-90. [[Abstract](#)] [[View PDF article](#)] [[PDF with links](#)]
160. Opeoluwa Wonuola Olawale, Ben Gilbert, Janet Reyna. 2022. Residential Demand Flexibility: Modeling Occupant Behavior using Sociodemographic Predictors. *Energy and Buildings* 262, 111973. [[Crossref](#)]
161. Yingchao Lan, Aravind Chandrasekaran, Deepa Goradia, Daniel Walker. 2022. Collaboration Structures in Integrated Healthcare Delivery Systems: An Exploratory Study of Accountable Care Organizations. *Manufacturing & Service Operations Management* 24:3, 1796-1820. [[Crossref](#)]
162. MIAO LIU. 2022. Assessing Human Information Processing in Lending Decisions: A Machine Learning Approach. *Journal of Accounting Research* 60:2, 607-651. [[Crossref](#)]
163. Deborah A. Cobb-Clark, Sarah C. Dahmann, Daniel A. Kamhöfer, Hannah Schildberg-Hörisch. 2022. The Predictive Power of Self-Control for Life Outcomes. *Journal of Economic Behavior & Organization* 197, 725-744. [[Crossref](#)]
164. Alexandre Belloni, Christian Hansen, Whitney Newey. 2022. High-dimensional linear models with many endogenous variables. *Journal of Econometrics* 228:1, 4-26. [[Crossref](#)]
165. Helmut Farbmacher, Raphael Guber, Sven Klaassen. 2022. Instrument Validity Tests With Causal Forests. *Journal of Business & Economic Statistics* 40:2, 605-614. [[Crossref](#)]
166. Loraine Seng, Jialiang Li. 2022. Structural Equation Model Averaging: Methodology and Application. *Journal of Business & Economic Statistics* 40:2, 815-828. [[Crossref](#)]
167. Joshua D. Angrist, Brigham Frandsen. 2022. Machine Labor. *Journal of Labor Economics* 40:S1, S97-S140. [[Crossref](#)]
168. Joram Mayshar, Omer Moav, Luigi Pascali. 2022. The Origin of the State: Land Productivity or Appropriability?. *Journal of Political Economy* 130:4, 1091-1144. [[Crossref](#)]
169. Yan Alperovych, Anantha Divakaruni, Sophie Manigart. 2022. Lending when relationships are scarce: The role of information spread via bank networks. *Journal of Corporate Finance* 73, 102181. [[Crossref](#)]
170. Anuj K. Shah, Michael LaForest. 2022. Knowledge about others reduces one's own sense of anonymity. *Nature* 603:7900, 297-301. [[Crossref](#)]
171. Marco Taboga. 2022. Cross-country differences in the size of venture capital financing rounds: a machine learning approach. *Empirical Economics* 62:3, 991-1012. [[Crossref](#)]
172. Helmut Wasserbacher, Martin Spindler. 2022. Machine learning for financial forecasting, planning and analysis: recent developments and pitfalls. *Digital Finance* 4:1, 63-88. [[Crossref](#)]
173. Kelly P. Nelson, Lee C. Parton, Zachary S. Brown. 2022. Biofuels policy and innovation impacts: Evidence from biofuels and agricultural patent indicators. *Energy Policy* 162, 112767. [[Crossref](#)]
174. Yingheng Zhang, Haojie Li, Gang Ren. 2022. Estimating heterogeneous treatment effects in road safety analysis using generalized random forests. *Accident Analysis & Prevention* 165, 106507. [[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. Qingliang Fan, Yu-Chin Hsu, Robert P. Lieli, Yichong Zhang. 2022. Estimation of Conditional Average Treatment Effects With High-Dimensional Data. *Journal of Business & Economic Statistics* 40:1, 313-327. [[Crossref](#)]
177. Shishir Shakya, Elham Erfanian, Alexandre Scarciuffolo. 2022. The political economy of Vermont's abortion bill. *Applied Economics Letters* 29:1, 17-21. [[Crossref](#)]

178. Carolina Lopez, Anja Sautmann, Simone Schaner. 2022. Does Patient Demand Contribute to the Overuse of Prescription Drugs?. *American Economic Journal: Applied Economics* 14:1, 225-260. [[Abstract](#)] [[View PDF article](#)] [[PDF with links](#)]
179. Stephen J. Anderson, David McKenzie. 2022. Improving Business Practices and the Boundary of the Entrepreneur: A Randomized Experiment Comparing Training, Consulting, Insourcing, and Outsourcing. *Journal of Political Economy* 130:1, 157-209. [[Crossref](#)]
180. Jorge Gallego, Mounu Prem, Juan F. Vargas. 2022. Predicting politicians' misconduct: Evidence from Colombia. *Data & Policy* 4. . [[Crossref](#)]
181. Giuseppe Albanese, Guglielmo Barone, Guido de Blasio. 2022. Populist voting and losers' discontent: Does redistribution matter?. *European Economic Review* 141, 104000. [[Crossref](#)]
182. Petre Caraiani. 2022. Using LASSO-family models to estimate the impact of monetary policy on corporate investments. *Economics Letters* 210, 110182. [[Crossref](#)]
183. Jan Gromadzki, Katarzyna Sałach, Michal Brzezinski. 2022. When populists deliver on their promises: the electoral effects of a large cash transfer program in Poland. *SSRN Electronic Journal* 141. . [[Crossref](#)]
184. Benjamin Collier, Cameron Ellis, Benjamin J. Keys. 2022. The Cost of Consumer Collateral: Evidence from Bunching. *SSRN Electronic Journal* 125. . [[Crossref](#)]
185. Demissew Diro Ejara, Wentworth Boynton, Xinyi Lu. 2022. Nonparametric quantile autoregressions. *SSRN Electronic Journal* 19. . [[Crossref](#)]
186. Pierre Koning, Paul Muller, Roger Prudon. 2022. Why Do Temporary Workers Have Higher Disability Insurance Risks Than Permanent Workers?. *SSRN Electronic Journal* 109. . [[Crossref](#)]
187. Marco Bertoni, Danilo Cavapozzi, Giacomo Pasini, Caterina Pavese. 2022. Remote Working and Mental Health During the First Wave of the COVID-19 Pandemic. *SSRN Electronic Journal* 132. . [[Crossref](#)]
188. Simon Briole, Marc Gurgand, Eric Maurin, Sandra McNally, Jenifer Ruiz-Valenzuela, Daniel Santín. 2022. The Making of Civic Virtues: A School-Based Experiment in Three Countries. *SSRN Electronic Journal* 136. . [[Crossref](#)]
189. Pierre Koning, Paul Muller, Roger Prudon. 2022. Why Do Temporary Workers Have Higher Disability Insurance Risks than Permanent Workers?. *SSRN Electronic Journal* 109. . [[Crossref](#)]
190. Narendra Bosukonda, Sonam Singh, Shrihari Sridhar, Vikas Mittal, Ashwin Malshe. 2022. Return On Strategy-Customer Alignment. *SSRN Electronic Journal* 29. . [[Crossref](#)]
191. Juan F. Vargas, Miguel E. Purroy, Sergio Perilla, Felipe Coy, Mounu Prem. 2022. Do Explosions Shape Voting Behavior?. *SSRN Electronic Journal* 11. . [[Crossref](#)]
192. Giorgia Barboni, Juan-Camilo Cárdenas, Nicolás de Roux. 2022. Behavioral Messages and Debt Repayment. *SSRN Electronic Journal* 130. . [[Crossref](#)]
193. Veli Andirin, Yusuf Negggers, Mehdi Shadmehr, Jesse Shapiro. 2022. Measuring the Tolerance of the State: Theory and Application to Protest. *SSRN Electronic Journal* 96. . [[Crossref](#)]
194. Stephen Michael Impink. 2022. Outsourcing IT and Technological Differentiation: Evidence from Digital Startups. *SSRN Electronic Journal* 72. . [[Crossref](#)]
195. 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](#)]
196. Philipp Kugler. 2022. The role of wage beliefs in the decision to become a nurse. *Health Economics* 31:1, 94-111. [[Crossref](#)]

197. Justine Guillochon. 2022. News-driven household macroeconomic expectations: regional vs. national telecast information. *SSRN Electronic Journal* 131. . [[Crossref](#)]
198. Felipe González, Josepa Miquel-Florensa, Mounu Prem, Stéphane Straub. 2022. The Dark Side of Infrastructure: Roads, Repression, and Land in Authoritarian Paraguay. *SSRN Electronic Journal* 60. . [[Crossref](#)]
199. Sergio Perilla, Mounu Prem, Miguel E. Purroy, Juan F. Vargas. 2022. How Peace Saves Lives: Evidence from Colombia. *SSRN Electronic Journal* 72. . [[Crossref](#)]
200. Filippo De Marco, Nicola Limodio. 2022. The Financial Transmission of a Climate Shock: El Niño and US Banks. *SSRN Electronic Journal* 33. . [[Crossref](#)]
201. Isaac K. Ofori, Christopher Quaidoo, Pamela E. Ofori. 2021. What Drives Financial Sector Development in Africa? Insights from Machine Learning. *Applied Artificial Intelligence* 35:15, 2124-2156. [[Crossref](#)]
202. 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](#)]
203. Milad Parpouchi, Akm Moniruzzaman, Julian M. Somers. 2021. The association between experiencing homelessness in childhood or youth and adult housing stability in Housing First. *BMC Psychiatry* 21:1. . [[Crossref](#)]
204. Prajmitra Bhuyan, Emma J. McCoy, Haojie Li, Daniel J. Graham. 2021. Analysing the causal effect of London cycle superhighways on traffic congestion. *The Annals of Applied Statistics* 15:4. . [[Crossref](#)]
205. Igor R.S. Brito, Alessandro V.M. Oliveira, Martin E. Dresner. 2021. An econometric study of the effects of airport privatization on airfares in Brazil. *Transport Policy* 114, 338-349. [[Crossref](#)]
206. Tobias Heinrich, Yoshiharu Kobayashi, Edward Lawson. 2021. Populism and foreign aid. *European Journal of International Relations* 27:4, 1042-1066. [[Crossref](#)]
207. Curtis Lockwood Reynolds, Amanda L. Weinstein. 2021. Gender differences in quality of life and preferences for location-specific amenities across cities. *Journal of Regional Science* 61:5, 916-943. [[Crossref](#)]
208. Xiajing Gong, Meng Hu, Mahashweta Basu, Liang Zhao. 2021. Heterogeneous treatment effect analysis based on machine-learning methodology. *CPT: Pharmacometrics & Systems Pharmacology* 10:11, 1433-1443. [[Crossref](#)]
209. Matthew C. Harding, Carlos Lamarche. 2021. Small Steps with Big Data: Using Machine Learning in Energy and Environmental Economics. *Annual Review of Resource Economics* 13:1, 469-488. [[Crossref](#)]
210. Lelys Dinarte Diaz, Marina Bassi. Quality Determinants of Short-Cycle Programs in Latin America and the Caribbean 123-159. [[Crossref](#)]
211. Jeanwoo Yoo, Prabhjot Grewal, Jessica Hotelling, Aikaterini Papamanoli, Kerry Cao, Simrat Dhaliwal, Robin Jacob, Azad Mojahedi, Michelle E. Bloom, Luis A. Marcos, Hal A. Skopicki, Andreas P. Kalogeropoulos. 2021. Admission NT-proBNP and outcomes in patients without history of heart failure hospitalized with COVID-19. *ESC Heart Failure* 8:5, 4278-4287. [[Crossref](#)]
212. Giovanni Cerulli. 2021. Improving econometric prediction by machine learning. *Applied Economics Letters* 28:16, 1419-1425. [[Crossref](#)]
213. Santiago Saavedra, Mauricio Romero. 2021. Local incentives and national tax evasion: The response of illegal mining to a tax reform in Colombia. *European Economic Review* 138, 103843. [[Crossref](#)]
214. Akash Malhotra. 2021. A hybrid econometric-machine learning approach for relative importance analysis: prioritizing food policy. *Eurasian Economic Review* 11:3, 549-581. [[Crossref](#)]
215. Emil Kupek. 2021. How many more? Under-reporting of the COVID-19 deaths in Brazil in 2020. *Tropical Medicine & International Health* 26:9, 1019-1028. [[Crossref](#)]

216. Tianxi Cai, T. Tony Cai, Zijian Guo. 2021. Optimal Statistical Inference for Individualized Treatment Effects in High-Dimensional Models. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **83**:4, 669-719. [[Crossref](#)]
217. Shishir Shakya, Kazi Sohag. 2021. The fracking boom and crime rates in rural American states: Some critical insights. *The Extractive Industries and Society* **8**:3, 100948. [[Crossref](#)]
218. Klara Zwickl, Simon Sturn, James K. Boyce. 2021. Effects of Carbon Mitigation on Co-pollutants at Industrial Facilities in Europe. *The Energy Journal* **42**:5, 123-148. [[Crossref](#)]
219. Tiffany Jiang. 2021. Using Machine Learning to Analyze Merger Activity. *Frontiers in Applied Mathematics and Statistics* **7**. . [[Crossref](#)]
220. Michael Danquah, Abdul Malik Iddrisu, Ernest Owusu Boakye, Solomon Owusu. 2021. Do gender wage differences within households influence women's empowerment and welfare? Evidence from Ghana. *Journal of Economic Behavior & Organization* **188**, 916-932. [[Crossref](#)]
221. Caitlin Brown, Rossella Calvi, Jacob Penglase. 2021. Sharing the pie: An analysis of undernutrition and individual consumption in Bangladesh. *Journal of Public Economics* **200**, 104460. [[Crossref](#)]
222. Robert M. Bohler, Dominic Hodgkin, Peter W. Kreiner, Traci C. Green. 2021. Predictors of US states' adoption of naloxone access laws, 2001–2017. *Drug and Alcohol Dependence* **225**, 108772. [[Crossref](#)]
223. Yusuke Narita. 2021. A Theory of Quasi-Experimental Evaluation of School Quality. *Management Science* **67**:8, 4982-5010. [[Crossref](#)]
224. Michael Powell, Allison Koenecke, James Brian Byrd, Akihiko Nishimura, Maximilian F. Konig, Ruoxuan Xiong, Sadiqa Mahmood, Vera Mucaj, Chetan Bettgowda, Liam Rose, Suzanne Tamang, Adam Sacarny, Brian Caffo, Susan Athey, Elizabeth A. Stuart, Joshua T. Vogelstein. 2021. Ten Rules for Conducting Retrospective Pharmacoepidemiological Analyses: Example COVID-19 Study. *Frontiers in Pharmacology* **12**. . [[Crossref](#)]
225. Luca J. Santos, Alessandro V.M. Oliveira, Dante Mendes Aldrighi. 2021. Testing the differentiated impact of the COVID-19 pandemic on air travel demand considering social inclusion. *Journal of Air Transport Management* **94**, 102082. [[Crossref](#)]
226. Vira Semenova, Victor Chernozhukov. 2021. Debaised machine learning of conditional average treatment effects and other causal functions. *The Econometrics Journal* **24**:2, 264-289. [[Crossref](#)]
227. Lelys Dinarte Diaz, Maria Marta Ferreyra, Sergio Urzua, Marina Bassi. What makes a Program Good? Evidence from Short-Cycle Higher Education Programs in Latin America and the Caribbean **33**, . [[Crossref](#)]
228. Girum Abebe, A. Stefano Caria, Esteban Ortiz-Ospina. 2021. The Selection of Talent: Experimental and Structural Evidence from Ethiopia. *American Economic Review* **111**:6, 1757-1806. [[Abstract](#)] [[View PDF article](#)] [[PDF with links](#)]
229. Jan Sviták, Jan Tichem, Stefan Haasbeek. 2021. Price effects of search advertising restrictions. *International Journal of Industrial Organization* **77**, 102736. [[Crossref](#)]
230. Alessandro V.M. Oliveira, Rodolfo R. Narcizo, Thiago Caliari, Maurício A.V. Morales, Rafael Prado. 2021. Estimating fuel-efficiency while accounting for dynamic fleet management: Testing the effects of fuel price signals and fleet rollover. *Transportation Research Part D: Transport and Environment* **95**, 102820. [[Crossref](#)]
231. Wenchao Li. 2021. The “miseries” of sex imbalance: Evidence using subjective well-being data. *Journal of Development Economics* **151**, 102634. [[Crossref](#)]
232. Elisa M. Maffioli. 2021. The political economy of health epidemics: Evidence from the Ebola outbreak. *Journal of Development Economics* **151**, 102651. [[Crossref](#)]

233. Katherine Casey, Abou Bakarr Kamara, Niccoló F. Meriggi. 2021. An Experiment in Candidate Selection. *American Economic Review* 111:5, 1575-1612. [[Abstract](#)] [[View PDF article](#)] [[PDF with links](#)]
234. Ryo Makioka. 2021. The impact of anti-sweatshop activism on employment. *Review of Development Economics* 25:2, 630-653. [[Crossref](#)]
235. Conner Mullally, Mayra Rivas, Travis McArthur. 2021. Using Machine Learning to Estimate the Heterogeneous Effects of Livestock Transfers. *American Journal of Agricultural Economics* 103:3, 1058-1081. [[Crossref](#)]
236. Asad Islam, Wang-Sheng Lee, Aaron Nicholas. 2021. The Effects of Chess Instruction on Academic and Non-cognitive Outcomes: Field Experimental Evidence from a Developing Country. *Journal of Development Economics* 150, 102615. [[Crossref](#)]
237. Olatunji Abdul Shobande. 2021. Decomposing the Persistent and Transitory Effect of Information and Communication Technology on Environmental Impacts Assessment in Africa: Evidence from Mundlak Specification. *Sustainability* 13:9, 4683. [[Crossref](#)]
238. Amy E Lerman, Samuel Trachtman. 2021. Where Policies and Politics Diverge. *Public Opinion Quarterly* 84:2, 419-445. [[Crossref](#)]
239. Sara Lowes, Eduardo Montero. 2021. The Legacy of Colonial Medicine in Central Africa. *American Economic Review* 111:4, 1284-1314. [[Abstract](#)] [[View PDF article](#)] [[PDF with links](#)]
240. Alexandre R. Scarciuffolo, Shishir Shakya, Joshua C. Hall. 2021. The political economy of Vermont's anti-fracking movement. *Journal of Public Finance and Public Choice* 36:1, 49-68. [[Crossref](#)]
241. Giuseppe Albanese, Guido de Blasio, Andrea Locatelli. 2021. Does EU regional policy promote local TFP growth? Evidence from the Italian Mezzogiorno. *Papers in Regional Science* 100:2, 327-349. [[Crossref](#)]
242. Michael C Knaus, Michael Lechner, Anthony Strittmatter. 2021. Machine learning estimation of heterogeneous causal effects: Empirical Monte Carlo evidence. *The Econometrics Journal* 24:1, 134-161. [[Crossref](#)]
243. Pushan Dutt, Ilia Tsetlin. 2021. Income distribution and economic development: Insights from machine learning. *Economics & Politics* 33:1, 1-36. [[Crossref](#)]
244. Yungu Kang, David H. Zhu, Yan Anthea Zhang. 2021. Being extraordinary: How CEOs' uncommon names explain strategic distinctiveness. *Strategic Management Journal* 42:2, 462-488. [[Crossref](#)]
245. Rajeev Dehejia, Cristian Pop-Eleches, Cyrus Samii. 2021. From Local to Global: External Validity in a Fertility Natural Experiment. *Journal of Business & Economic Statistics* 39:1, 217-243. [[Crossref](#)]
246. Achim Ahrens, Christopher Aitken, Mark E. Schaffer. Using Machine Learning Methods to Support Causal Inference in Econometrics 23-52. [[Crossref](#)]
247. Amy W. Ando, Kathy Baylis. Spatial Environmental and Natural Resource Economics 1415-1434. [[Crossref](#)]
248. Felipe Valencia Caicedo. Historical Econometrics: Instrumental Variables and Regression Discontinuity Designs 179-211. [[Crossref](#)]
249. Edward Martey, Prince M. Etwire, David Atinga. 2021. To attend or not to attend: Examining the relationship between food hardship, school attendance and education expenditure. *International Journal of Educational Development* 80, 102304. [[Crossref](#)]
250. Jorge Gallego, Gonzalo Rivero, Juan Martínez. 2021. Preventing rather than punishing: An early warning model of malfeasance in public procurement. *International Journal of Forecasting* 37:1, 360-377. [[Crossref](#)]
251. Christophre Georges, Javier Pereira. 2021. Market stability with machine learning agents. *Journal of Economic Dynamics and Control* 122, 104032. [[Crossref](#)]

252. Vegard H. Larsen, Leif Anders Thorsrud, Julia Zhulanova. 2021. News-driven inflation expectations and information rigidities. *Journal of Monetary Economics* **117**, 507-520. [[Crossref](#)]
253. Giuseppe Albanese, Emanuele Ciani, Guido de Blasio. 2021. Anything new in town? The local effects of urban regeneration policies in Italy. *Regional Science and Urban Economics* **86**, 103623. [[Crossref](#)]
254. Ana Beatriz R. Eufrásio, Rogéria A.G. Eller, Alessandro V.M. Oliveira. 2021. Are on-time performance statistics worthless? An empirical study of the flight scheduling strategies of Brazilian airlines. *Transportation Research Part E: Logistics and Transportation Review* **145**, 102186. [[Crossref](#)]
255. Berkay Akyapi. 2021. How Does Automation Affect Aggregate Labor Share and Firm Level Economic Outcomes?. *SSRN Electronic Journal* **110**. . [[Crossref](#)]
256. Alessandra Mattei, Fabrizia Mealli, Anahita Nodehi. Design and Analysis of Experiments 1-41. [[Crossref](#)]
257. Arben Kita. 2021. Machine Learning Predictions of Credit and Equity Risk Premia. *SSRN Electronic Journal* **74**. . [[Crossref](#)]
258. Christoph Boehm, T. Niklas Kroner. 2021. Beyond the Yield Curve: Understanding the Effect of FOMC Announcements on the Stock Market. *SSRN Electronic Journal* **110**. . [[Crossref](#)]
259. Benjamin Collier, Cameron Ellis. 2021. Credit Demand in a Crisis. *SSRN Electronic Journal* **42**. . [[Crossref](#)]
260. Francesca Micocci, Armando Rungi. 2021. Predicting Exporters with Machine Learning. *SSRN Electronic Journal* **83**. . [[Crossref](#)]
261. Jacob Moscona, Karthik Sastry. 2021. Inappropriate Technology: Evidence from Global Agriculture. *SSRN Electronic Journal* **116**. . [[Crossref](#)]
262. Anthony Bald, Eric Chyn, Justine Hastings, Margarita Machelett. 2021. The Causal Impact of Removing Children from Abusive and Neglectful Homes. *SSRN Electronic Journal* **113**. . [[Crossref](#)]
263. Jorge Guzman. 2021. Treatment Effects in Strategic Management: With an Application to Choosing Early Stage Venture Capital. *SSRN Electronic Journal* **21**. . [[Crossref](#)]
264. Nicolás de Roux, Luis Martinez. 2021. Inversión Perdida: Conflicto Civil y Crédito Agrícola en Colombia (Forgone Investment: Civil Conflict and Agricultural Credit in Colombia). *SSRN Electronic Journal* **93**. . [[Crossref](#)]
265. James Reisinger. 2021. Southernization: The Long-Term Effect of Migration on Racial Prejudice and Political Preferences. *SSRN Electronic Journal* **78**. . [[Crossref](#)]
266. Benjamin Collier, Cameron Ellis, Benjamin J. Keys. 2021. The Cost of Consumer Collateral: Evidence from Bunching. *SSRN Electronic Journal* **125**. . [[Crossref](#)]
267. Maxim Ananyev, Mikhail Poyker. 2021. Prisons and Homophobia. *SSRN Electronic Journal* **133**. . [[Crossref](#)]
268. Michael C. Knaus. 2021. A Double Machine Learning Approach to Estimate the Effects of Musical Practice on Student's Skills. *Journal of the Royal Statistical Society Series A: Statistics in Society* **184**:1, 282-300. [[Crossref](#)]
269. Kathy Baylis, Thomas Heckelei, Hugo Storm. Machine learning in agricultural economics 4551-4612. [[Crossref](#)]
270. Nick Huntington-Klein. 2020. Instruments with Heterogeneous Effects: Bias, Monotonicity, and Localness. *Journal of Causal Inference* **8**:1, 182-208. [[Crossref](#)]
271. Jinyong Kim, Mingook Kim, Yongsik Kim. 2020. Bank Transparency and the Market's Perception of Bank Risk. *Journal of Financial Services Research* **58**:2-3, 115-142. [[Crossref](#)]
272. Wanderson Rocha Bittencourt, Pedro H. M. Albuquerque. 2020. Evaluating company bankruptcies using causal forests. *Revista Contabilidade & Finanças* **31**:84, 542-559. [[Crossref](#)]

273. Carolina Lopez, Anja Sautmann, Simone Schaner. Does Patient Demand Contribute to the Overuse of Prescription Drugs ? 6, . [[Crossref](#)]
274. Carolina Castagnetti, Luisa Rosti, Marina Töpfer. The Age Pay Gap between Young and Older Employees in Italy: Perceived or Real Discrimination against the Young? 195-221. [[Crossref](#)]
275. Alex Armand, Alexander Coutts, Pedro C. Vicente, Inês Vilela. 2020. Does Information Break the Political Resource Curse? Experimental Evidence from Mozambique. *American Economic Review* 110:11, 3431-3453. [[Abstract](#)] [[View PDF article](#)] [[PDF with links](#)]
276. Marius Faber. 2020. Robots and reshoring: Evidence from Mexican labor markets. *Journal of International Economics* 127, 103384. [[Crossref](#)]
277. Hector F. Calvo-Pardo, Tullio Mancini, Jose Olmo. 2020. Neural Network Models for Empirical Finance. *Journal of Risk and Financial Management* 13:11, 265. [[Crossref](#)]
278. Barteld Braaksma, Kees Zeelenberg, Sofie De Broe. Big Data in Official Statistics 303-338. [[Crossref](#)]
279. Alex Coad, Stjepan Srhoj. 2020. Catching Gazelles with a Lasso: Big data techniques for the prediction of high-growth firms. *Small Business Economics* 55:3, 541-565. [[Crossref](#)]
280. 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](#)]
281. Ran Goldblatt, Kilian Heilmann, Yonatan Vaizman. 2020. Can Medium-Resolution Satellite Imagery Measure Economic Activity at Small Geographies? Evidence from Landsat in Vietnam. *The World Bank Economic Review* 34:3, 635-653. [[Crossref](#)]
282. Mikhail Stolbov, Maria Shchepeleva. 2020. What predicts the legal status of cryptocurrencies?. *Economic Analysis and Policy* 67, 273-291. [[Crossref](#)]
283. Pieter Bakx, Bram Wouterse, Eddy van Doorslaer, Albert Wong. 2020. Better off at home? Effects of nursing home eligibility on costs, hospitalizations and survival. *Journal of Health Economics* 73, 102354. [[Crossref](#)]
284. Dalia Ghanem, Shu Shen, Junjie Zhang. 2020. A Censored Maximum Likelihood Approach to Quantifying Manipulation in China's Air Pollution Data. *Journal of the Association of Environmental and Resource Economists* 7:5, 965-1003. [[Crossref](#)]
285. Mitsuru Igami. 2020. Artificial intelligence as structural estimation: Deep Blue, Bonanza, and AlphaGo. *The Econometrics Journal* 23:3, S1-S24. [[Crossref](#)]
286. James M. Carson, Cameron M. Ellis, Robert E. Hoyt, Krzysztof Ostaszewski. 2020. Sunk Costs and Screening: Two-Part Tariffs in Life Insurance. *Journal of Risk and Insurance* 87:3, 689-718. [[Crossref](#)]
287. Khishigsuren Davagdorj, Van Huy Pham, Nipon Theera-Umpon, Keun Ho Ryu. 2020. XGBoost-Based Framework for Smoking-Induced Noncommunicable Disease Prediction. *International Journal of Environmental Research and Public Health* 17:18, 6513. [[Crossref](#)]
288. Alexander D'Amour, Peng Ding, Avi Feller, Lihua Lei, Jasjeet Sekhon. 2020. Overlap in observational studies with high-dimensional covariates. *Journal of Econometrics* 28. . [[Crossref](#)]
289. Xiaoyun Cheng, Kun Huang, Lei Qu, Tianbao Zhang, Li Li. 2020. Effects of Vehicle Restriction Policies on Urban Travel Demand Change from a Built Environment Perspective. *Journal of Advanced Transportation* 2020, 1-13. [[Crossref](#)]
290. Martin Abel, Rulof Burger, Patrizio Piraino. 2020. The Value of Reference Letters: Experimental Evidence from South Africa. *American Economic Journal: Applied Economics* 12:3, 40-71. [[Abstract](#)] [[View PDF article](#)] [[PDF with links](#)]
291. Kelli DePriest, Rita D'Aoust, Laura Samuel, Yvonne Commodore-Mensah, Ginger Hanson, Eric P. Slade. 2020. Nurse practitioners' workforce outcomes under implementation of full practice authority. *Nursing Outlook* 68:4, 459-467. [[Crossref](#)]

292. Hugo Storm, Kathy Baylis, Thomas Heckelei. 2020. Machine learning in agricultural and applied economics. *European Review of Agricultural Economics* 47:3, 849-892. [[Crossref](#)]
293. Daron Acemoglu, Pascual Restrepo. 2020. Robots and Jobs: Evidence from US Labor Markets. *Journal of Political Economy* 128:6, 2188-2244. [[Crossref](#)]
294. GUANHAO FENG, STEFANO GIGLIO, DACHENG XIU. 2020. Taming the Factor Zoo: A Test of New Factors. *The Journal of Finance* 75:3, 1327-1370. [[Crossref](#)]
295. Kellyn F Arnold, Vinny Davies, Marc de Kamps, Peter W G Tennant, John Mbotwa, Mark S Gilthorpe. 2020. Reflections on modern methods: generalized linear models for prognosis and intervention—theory, practice and implications for machine learning. *International Journal of Epidemiology* 32. . [[Crossref](#)]
296. Christophe Croux, Julapa Jagtiani, Tarunsai Korivi, Milos Vulcanovic. 2020. Important factors determining Fintech loan default: Evidence from a lendingclub consumer platform. *Journal of Economic Behavior & Organization* 173, 270-296. [[Crossref](#)]
297. Marta Auricchio, Emanuele Ciani, Alberto Dalmazzo, Guido de Blasio. 2020. Life after public employment retrenchment: evidence from Italian municipalities. *Journal of Economic Geography* 20:3, 733-782. [[Crossref](#)]
298. Lorenzo Camponovo. 2020. Bootstrap inference for penalized GMM estimators with oracle properties. *Econometric Reviews* 39:4, 362-372. [[Crossref](#)]
299. Georges Bresson. Comments on “An Econometrician’s Perspective on Big Data” by Cheng Hsiao 431-443. [[Crossref](#)]
300. Falco J. Bargagli Stoffi, Giorgio Gnecco. 2020. Causal tree with instrumental variable: an extension of the causal tree framework to irregular assignment mechanisms. *International Journal of Data Science and Analytics* 9:3, 315-337. [[Crossref](#)]
301. Luv Sharma, Aravind Chandrasekaran, Elliot Bendoly. 2020. Does the Office of Patient Experience Matter in Improving Delivery of Care?. *Production and Operations Management* 29:4, 833-855. [[Crossref](#)]
302. Mireille E. Schnitzer, Joel Sango, Steve Ferreira Guerra, Mark J. Laan. 2020. Data-adaptive longitudinal model selection in causal inference with collaborative targeted minimum loss-based estimation. *Biometrics* 76:1, 145-157. [[Crossref](#)]
303. Jakob Everding, Jan Marcus. 2020. The effect of unemployment on the smoking behavior of couples. *Health Economics* 29:2, 154-170. [[Crossref](#)]
304. Marta Auricchio, Emanuele Ciani, Alberto Dalmazzo, Guido de Blasio. 2020. Redistributive public employment? A test for the South of Italy. *Economics Letters* 186, 108787. [[Crossref](#)]
305. Abhijit Banerjee, Rema Hanna, Benjamin A. Olken, Sudarno Sumarto. 2020. The (lack of) distortionary effects of proxy-means tests: Results from a nationwide experiment in Indonesia. *Journal of Public Economics Plus* 1, 100001. [[Crossref](#)]
306. 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](#)]
307. Román Andrés Zárate. 2020. More than Friends: Beliefs and Peer Effects in the Formation of Social and Academic Skills. *SSRN Electronic Journal* . [[Crossref](#)]
308. Antonella Ferrara, Lewis Dijkstra, Philip McCann, Rosanna Nisticò. 2020. The Response of Regional Well-Being to EU Cohesion Policy Interventions. *SSRN Electronic Journal* 15. . [[Crossref](#)]
309. Zhipeng Liao, Xiaoxia Shi. 2020. A nondegenerate Vuong test and post selection confidence intervals for semi/nonparametric models. *Quantitative Economics* 11:3, 983-1017. [[Crossref](#)]

310. Travis Baseler. 2020. Hidden Income and the Perceived Returns to Migration: Experimental Evidence from Kenya. *SSRN Electronic Journal* 113. . [[Crossref](#)]
311. Yan Alperovych, Anantha Divakaruni, Sophie Manigart. 2020. Lending When Relationships Are Scarce: The Role of Information Spread via Bank Networks. *SSRN Electronic Journal* 64. . [[Crossref](#)]
312. Nicolás de Roux, Luis Martinez. 2020. Conflict Reduces Investment: Evidence from the Demand for Agricultural Credit in Colombia. *SSRN Electronic Journal* 93. . [[Crossref](#)]
313. Elliott Ash, Sergio Galletta, Tommaso Giommoni. 2020. A Machine Learning Approach to Analyzing Corruption in Local Public Finances. *SSRN Electronic Journal* 156. . [[Crossref](#)]
314. Athina Petropoulou, Vasileios Pappas, Steven R. G. Ongena, Dimitrios Gounopoulos, Richard J. Fairchild. 2020. The Efficiency of US Community Banks. *SSRN Electronic Journal* 13. . [[Crossref](#)]
315. Marcella Alsan, Owen Garrick, Grant Graziani. 2019. Does Diversity Matter for Health? Experimental Evidence from Oakland. *American Economic Review* 109:12, 4071-4111. [[Abstract](#)] [[View PDF article](#)] [[PDF with links](#)]
316. Manav Raj, Robert Seamans. 2019. Primer on artificial intelligence and robotics. *Journal of Organization Design* 8:1. . [[Crossref](#)]
317. Paul Wolfson, Dale Belman. 2019. 15 Years of Research on US Employment and the Minimum Wage. *LABOUR* 33:4, 488-506. [[Crossref](#)]
318. 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](#)]
319. Dominik Rehse, Ryan Riordan, Nico Rottke, Joachim Zietz. 2019. The effects of uncertainty on market liquidity: Evidence from Hurricane Sandy. *Journal of Financial Economics* 134:2, 318-332. [[Crossref](#)]
320. Colin Gray. 2019. Leaving benefits on the table: Evidence from SNAP. *Journal of Public Economics* 179, 104054. [[Crossref](#)]
321. Andrew C. Forrester, Benjamin Powell, Alex Nowrasteh, Michelangelo Landgrave. 2019. Do immigrants import terrorism?. *Journal of Economic Behavior & Organization* 166, 529-543. [[Crossref](#)]
322. Liangjun Su, Takuya Ura, Yichong Zhang. 2019. Non-separable models with high-dimensional data. *Journal of Econometrics* 212:2, 646-677. [[Crossref](#)]
323. Milan Miric, Kevin J. Boudreau, Lars Bo Jeppesen. 2019. Protecting their digital assets: The use of formal & informal appropriability strategies by App developers. *Research Policy* 48:8, 103738. [[Crossref](#)]
324. Susan Athey, Guido W. Imbens. 2019. Machine Learning Methods That Economists Should Know About. *Annual Review of Economics* 11:1, 685-725. [[Crossref](#)]
325. Gerald A. Carlino, Albert Saiz. 2019. Beautiful city: Leisure amenities and urban growth. *Journal of Regional Science* 59:3, 369-408. [[Crossref](#)]
326. Zhuan Pei, Jörn-Steffen Pischke, Hannes Schwandt. 2019. Poorly Measured Confounders are More Useful on the Left than on the Right. *Journal of Business & Economic Statistics* 37:2, 205-216. [[Crossref](#)]
327. Hilal Atasoy, Brad N. Greenwood, Jeffrey Scott McCullough. 2019. The Digitization of Patient Care: A Review of the Effects of Electronic Health Records on Health Care Quality and Utilization. *Annual Review of Public Health* 40:1, 487-500. [[Crossref](#)]
328. Brandyn F. Churchill, Joseph J. Sabia. 2019. The Effects of Minimum Wages on Low-Skilled Immigrants' Wages, Employment, and Poverty. *Industrial Relations: A Journal of Economy and Society* 58:2, 275-314. [[Crossref](#)]

329. Dario Sansone. 2019. Beyond Early Warning Indicators: High School Dropout and Machine Learning. *Oxford Bulletin of Economics and Statistics* **81**:2, 456-485. [[Crossref](#)]
330. Jun Sung Kim, Bin Jiang, Chuhui Li, Hee-Seung Yang. 2019. Returns to women's education using optimal IV selection. *Applied Economics* **51**:8, 815-830. [[Crossref](#)]
331. Amy W. Ando, Kathy Baylis. *Spatial Environmental and Natural Resource Economics* 1-20. [[Crossref](#)]
332. Zhaoyuan Li, Jianfeng Yao. 2019. Testing for heteroscedasticity in high-dimensional regressions. *Econometrics and Statistics* **9**, 122-139. [[Crossref](#)]
333. Clément de Chaisemartin, Nicolas Navarrete. 2019. The Direct and Spillover Effects of a Mental Health Program for Disruptive Students. *SSRN Electronic Journal* . [[Crossref](#)]
334. Emanuele Colonnelli, Jorge A. Gallego, Mounu Prem. 2019. What Predicts Corruption?. *SSRN Electronic Journal* . [[Crossref](#)]
335. Lenard Lieb, Johannes Schuffels. 2019. Inflation Expectations and Consumer Spending: The Role of Household Balance Sheets. *SSRN Electronic Journal* . [[Crossref](#)]
336. Lorenzo Camponovo. 2019. Bootstrap Inference for Penalized GMM Estimators with Oracle Properties. *SSRN Electronic Journal* . [[Crossref](#)]
337. Christophre Georges, Javier Pereira. 2019. Market Stability with Machine Learning Agents. *SSRN Electronic Journal* . [[Crossref](#)]
338. Annamaria Conti, Jorge Guzman. 2019. What is the US Comparative Advantage in Entrepreneurship? Evidence from Israeli Migration to the United States. *SSRN Electronic Journal* **62**. . [[Crossref](#)]
339. Kaspar Wuthrich, Ying Zhu. 2019. Omitted Variable Bias of Lasso-Based Inference Methods Under Limited Variability: A Finite Sample Analysis. *SSRN Electronic Journal* . [[Crossref](#)]
340. Olga Namen, Mounu Prem, Juan F. Vargas. 2019. The Human Capital Peace Dividend. *SSRN Electronic Journal* . [[Crossref](#)]
341. Maria Angelica Bautista, Felipe González, Luis Martinez, Pablo Munoz, Mounu Prem. 2019. The Geography of Dictatorship and Support for Democracy. *SSRN Electronic Journal* . [[Crossref](#)]
342. Daniel Mejia, Mounu Prem, Juan F. Vargas. 2019. The Rise and Persistence of Illegal Crops: Evidence from a Naive Policy Announcement. *SSRN Electronic Journal* . [[Crossref](#)]
343. Hui-Ching Chuang, Jau-er Chen. 2019. Industry Distress and Default Recovery Rates: The Unconditional Quantile Regression Approach. *SSRN Electronic Journal* . [[Crossref](#)]
344. Brett R. Gordon, Mitch Lovett, Bowen Luo, James Reeder. 2019. Disentangling the Effects of Ad Tone on Voter Turnout and Candidate Choice in Presidential Elections. *SSRN Electronic Journal* **47**. . [[Crossref](#)]
345. Abdul-Nasah Soale, Emmanuel Tsyawo. 2019. Clustered Covariate Regression. *SSRN Electronic Journal* **31**. . [[Crossref](#)]
346. Khai Xiang Chiong, Matthew Shum. 2019. Random Projection Estimation of Discrete-Choice Models with Large Choice Sets. *Management Science* **65**:1, 256-271. [[Crossref](#)]
347. Denvil Duncan, Justin M. Ross, John L. Mikesell. 2019. Can the Federal Budget Process Promote Fiscal Sustainability? Evidence from Synthetic Control Using Lasso. *SSRN Electronic Journal* **105**. . [[Crossref](#)]
348. Marco Pangallo, Michele Loberto. 2018. Home is where the ad is: online interest proxies housing demand. *EPJ Data Science* **7**:1. . [[Crossref](#)]
349. Zijian Guo, Hyunseung Kang, T. Tony Cai, Dylan S. Small. 2018. Testing endogeneity with high dimensional covariates. *Journal of Econometrics* **207**:1, 175-187. [[Crossref](#)]
350. Ingvild Almås, Orazio Attanasio, Jyotsna Jalan, Francisco Oteiza, Marcella Vigneri. 2018. Using data differently and using different data. *Journal of Development Effectiveness* **10**:4, 462-481. [[Crossref](#)]

351. Joseph J. Sabia, Richard V. Burkhauser, Taylor Mackay. 2018. Minimum Cash Wages, Tipped Restaurant Workers, and Poverty. *Industrial Relations: A Journal of Economy and Society* 57:4, 637-670. [[Crossref](#)]
352. Gilles Bastin, Paola Tubaro. 2018. Le moment big data des sciences sociales. *Revue française de sociologie* Vol. 59:3, 375-394. [[Crossref](#)]
353. Patricia Boyallian, Pablo Ruiz-Verdú. 2018. Leverage, CEO Risk-Taking Incentives, and Bank Failure during the 2007–10 Financial Crisis*. *Review of Finance* 22:5, 1763-1805. [[Crossref](#)]
354. Stephen Hansen, Michael McMahon, Andrea Prat. 2018. Transparency and Deliberation Within the FOMC: A Computational Linguistics Approach*. *The Quarterly Journal of Economics* 133:2, 801-870. [[Crossref](#)]
355. Simon Sturn. 2018. Do Minimum Wages Lead to Job Losses? Evidence from OECD Countries on Low-Skilled and Youth Employment. *ILR Review* 71:3, 647-675. [[Crossref](#)]
356. Conner Mullally, Shourish Chakravarty. 2018. Are matching funds for smallholder irrigation money well spent?. *Food Policy* 76, 70-80. [[Crossref](#)]
357. Jose Fernandez, Stephan Gohmann, Joshua C. Pinkston. 2018. Breaking Bad in Bourbon Country: Does Alcohol Prohibition Encourage Methamphetamine Production?. *Southern Economic Journal* 84:4, 1001-1023. [[Crossref](#)]
358. Francesco Audrino, Lorenzo Camponovo. 2018. Oracle Properties, Bias Correction, and Bootstrap Inference for Adaptive Lasso for Time Series M -Estimators. *Journal of Time Series Analysis* 39:2, 111-128. [[Crossref](#)]
359. Anil Kumar. 2018. Do Restrictions on Home Equity Extraction Contribute to Lower Mortgage Defaults? Evidence from a Policy Discontinuity at the Texas Border. *American Economic Journal: Economic Policy* 10:1, 268-297. [[Abstract](#)] [[View PDF article](#)] [[PDF with links](#)]
360. Matthew J. Baker. 2018. Was Civil War surgery effective?. *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 51:1, 49-61. [[Crossref](#)]
361. Zhenqiu Liu, Shili Lin. Sparse Treatment-Effect Model for Taxon Identification with High-Dimensional Metagenomic Data 309-318. [[Crossref](#)]
362. Damian Kozbur. 2018. Testing-Based Forward Model Selection. *SSRN Electronic Journal* . [[Crossref](#)]
363. Caitlin Brown, Rossella Calvi, Jacob Penglase. 2018. Sharing the Pie: Undernutrition, Intra-Household Allocation, and Poverty. *SSRN Electronic Journal* . [[Crossref](#)]
364. Yingying Fan, Jinchi Lv, Jingbo Wang. 2018. DNN: A Two-Scale Distributional Tale of Heterogeneous Treatment Effect Inference. *SSRN Electronic Journal* . [[Crossref](#)]
365. Maria Angelica Bautista, Felipe González, Luis Martinez, Pablo Munoz, Mounu Prem. 2018. The Geography of Repression and Support for Democracy: Evidence From the Pinochet Dictatorship. *SSRN Electronic Journal* . [[Crossref](#)]
366. Raphael Guber. 2018. Instrument Validity Tests with Causal Trees: With an Application to the Same-sex Instrument. *SSRN Electronic Journal* 88. . [[Crossref](#)]
367. Marco Pangallo, Michele Loberto. 2018. Home Is Where the Ad Is: Online Interest Proxies Housing Demand. *SSRN Electronic Journal* . [[Crossref](#)]
368. Murray Z. Frank, Keer Yang. 2018. Does Finance Flow to High Productivity Firms?. *SSRN Electronic Journal* . [[Crossref](#)]
369. Abhijit V. Banerjee, Arun G. Chandrasekhar, Esther Duflo, Matthew O. Jackson. 2018. Changes in Social Network Structure in Response to Exposure to Formal Credit Markets. *SSRN Electronic Journal* 113. . [[Crossref](#)]

370. Jorge Guzman. 2018. Go West Young Firm: The Value of Entrepreneurial Migration for Startups and Their Founders. *SSRN Electronic Journal* 24. . [[Crossref](#)]
371. Alex Krumer, Michael Lechner. 2018. MIDWEEK EFFECT ON SOCCER PERFORMANCE: EVIDENCE FROM THE GERMAN BUNDESLIGA. *Economic Inquiry* 56:1, 193-207. [[Crossref](#)]
372. Cesi Cruz, Julien Labonne, Pablo Querubín. 2017. Politician Family Networks and Electoral Outcomes: Evidence from the Philippines. *American Economic Review* 107:10, 3006-3037. [[Abstract](#)] [[View PDF article](#)] [[PDF with links](#)]
373. Marcel Fafchamps, Julien Labonne. 2017. Using Split Samples to Improve Inference on Causal Effects. *Political Analysis* 25:4, 465-482. [[Crossref](#)]
374. 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](#)]
375. David C. Phillips. 2017. Stopping on Nine: Evidence of Heuristic Managerial Decision-Making in Major League Baseball Pitcher Substitutions. *Southern Economic Journal* 84:2, 577-599. [[Crossref](#)]
376. Evan Totty. 2017. THE EFFECT OF MINIMUM WAGES ON EMPLOYMENT: A FACTOR MODEL APPROACH. *Economic Inquiry* 55:4, 1712-1737. [[Crossref](#)]
377. Jushan Bai, Yuan Liao. 2017. Inferences in panel data with interactive effects using large covariance matrices. *Journal of Econometrics* 200:1, 59-78. [[Crossref](#)]
378. T. Tony Cai, Wenguang Sun. 2017. Large-Scale Global and Simultaneous Inference: Estimation and Testing in Very High Dimensions. *Annual Review of Economics* 9:1, 411-439. [[Crossref](#)]
379. Sylvia Allegretto, Arindrajit Dube, Michael Reich, Ben Zipperer. 2017. Credible Research Designs for Minimum Wage Studies. *ILR Review* 70:3, 559-592. [[Crossref](#)]
380. 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](#)]
381. 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](#)]
382. Ingrid Gould Ellen, Keren Mertens Horn, Davin Reed. 2017. Has Falling Crime Invited Gentrification?. *SSRN Electronic Journal* . [[Crossref](#)]
383. Yusuke Narita. 2017. (Non)Randomization: A Theory of Quasi-Experimental Evaluation of School Quality. *SSRN Electronic Journal* 93. . [[Crossref](#)]
384. Chaohua Dong, Jiti Gao, Oliver B. Linton. 2017. High Dimensional Semiparametric Moment Restriction Models. *SSRN Electronic Journal* . [[Crossref](#)]
385. Andreas Fuster, Paul Goldsmith-Pinkham, Tarun Ramadorai, Ansgar Walther. 2017. Predictably Unequal? The Effects of Machine Learning on Credit Markets. *SSRN Electronic Journal* . [[Crossref](#)]
386. 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](#)]
387. Orestis P. Panagopoulos, Vijay Pappu, Petros Xanthopoulos, Panos M. Pardalos. 2016. Constrained subspace classifier for high dimensional datasets. *Omega* 59, 40-46. [[Crossref](#)]
388. Matthew Shum, Khai X. Chiong. 2016. Random Projection Estimation of Discrete-Choice Models with Large Choice Sets. *SSRN Electronic Journal* . [[Crossref](#)]
389. Khai Xiang Chiong, Matthew Shum. 2016. Random Projection Estimation of Discrete-Choice Models with Large Choice Sets. *SSRN Electronic Journal* . [[Crossref](#)]
390. Bin Jiang, Jun Sung Kim, Chuhui Li, Hee-Seung Yang, Ou Yang. 2016. Post-Lasso IV Estimation of Returns to Women's Education. *SSRN Electronic Journal* . [[Crossref](#)]

391. Serena Ng. 2016. Opportunities and Challenges: Lessons from Analyzing Terabytes of Scanner Data. *SSRN Electronic Journal* . [[Crossref](#)]
392. Emanuele Ciani, Guido de Blasio. 2015. European structural funds during the crisis: evidence from Southern Italy. *IZA Journal of Labor Policy* 4:1. . [[Crossref](#)]
393. Holger Spamann. 2015. Empirical Comparative Law. *Annual Review of Law and Social Science* 11:1, 131-153. [[Crossref](#)]
394. 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](#)]
395. Rocío Titiunik. 2015. Can Big Data Solve the Fundamental Problem of Causal Inference?. *PS: Political Science & Politics* 48:01, 75-79. [[Crossref](#)]
396. Pushan Dutt, Ilia Tsetlin. 2015. Beyond Gini: Income Distribution and Economic Development. *SSRN Electronic Journal* 108. . [[Crossref](#)]
397. Paul J. Wolfson, Dale Belman. 2015. 15 Years of Research on U.S. Employment and the Minimum Wage. *SSRN Electronic Journal* . [[Crossref](#)]
398. Kees Zeelenberg, Barteld Braaksma. Big Data in Official Statistics 274-296. [[Crossref](#)]