Research Statement

July 2021

Nobel Prize winner Trygve Haavelmo defined econometrics as "a conjunction of economic theory, and actual measurements, using the theory and technique of statistical inference as a bridge pier" (Haavelmo, 1944). The other founding fathers of econometrics (Frisch, 1933; Tinbergen, 1951) also emphasized that economic theory, data analysis, and statistics were the core constituents of the field.

I think of myself as a classical econometrician, in the broad sense of the definition above. My focus over the last years has been on using an econometric approach to gain a better understanding of the following topics:

1. *The use of (online) randomized experiments to guide innovation.*
2. *The use of econometric models for macroeconomic policy evaluation.*
3. *Machine learning and its relation to economics and econometrics.*
4. *Deadly force use by police in the United States.*

In this statement I provide a brief motivation for the study of each subject, the set of questions I have asked, and the contributions I have made. I conclude by discussing other projects that, while an important part of my research record and intellectual development, do not neatly fit into the categories above.

1 Randomized Experiments and Innovation

1.1 Lean experimentation

Randomized controlled trials are increasingly driving innovation in many fields. In the high-tech sector, internet companies (i.e., companies that present themselves to the public via a website) run thousands of experiments—known as "A/B Tests"—every year to screen innovations to the products and services they offer. In policy and academic circles, randomized experiments are frequently used to evaluate social programs and shape public policy.

Public or private actors that have to choose among several potential innovations while having a limited number of trial participants, can proceed in two different ways. They can either discard some innovations without experimentation, and run only a handful of randomized controlled trials; or they can experiment on all available potential innovations by running several randomized controlled trials, each with possibly only a few experimental subjects ("lean" experiments). Which of these two strategies is more effective?

This question was the subject of the paper **"A/B Testing with Fat Tails"**, published in the ***Journal of Political Economy*** jointly with Eduardo Azevedo, Alex Deng, Justin Rao, and Glen Weyl (Azevedo, Deng, Montiel Olea, Rao, and Weyl, 2020a). The key insight in this paper is that the optimal experimentation strategy depends on whether most gains from running a randomized controlled trial accrue from typical innovations or from rare, large, and unpredictable successes. In particular, if the unobserved distribution of innovation quality has "fat tails" (that is, if the distribution of innovation quality features innovations with large effects), then experimenting on more ideas, each with possibly smaller sample sizes, is preferred. If, however, the tails are not too fat, the more typical approach of using a few randomized controlled trials with large sample sizes becomes optimal.

The main conceptual contribution of this paper is a reassessment of the value of randomized controlled trials with small sample sizes. Bayesian decision theory suggests that if an agent has a strict ordering over two actions (for example, if absent evidence the agent prefers not to implement a product innovation), then a statistically noisy randomized experiment is unlikely to reverse this ranking. This makes the value of randomized experiments with few participants close to zero (Radner and Stiglitz, 1984). Our main results show that this result does not hold in the case where a lean randomized experiment can deliver a very large outcome in favor of implementing the innovation, which can occur when the unobserved distribution of innovation quality has fat tails.

Our main message, which is normative, thus advocate for a more agile pipeline for innovation, where actors can quickly iterate through many ideas, experiment quickly, and pivot from ideas that are not resounding successes (provided there is evidence that innovations with large effects are possible).

As in the classical definition of econometrics, this project builds on a combination of economic theory, data analysis, and statistical and mathematical methodology.

The paper uses statistical decision theory and microeconomic theory to provide a model of a firm that needs to allocate scarce experimental resources (participants in the randomized controlled trial) to screen different potential innovations. The model's specifications were inspired by how modern technology companies use online, large-scale experiments to guide their product innovation decisions. To acquire this knowledge, we collaborated with researchers at Microsoft Bing's EXP platform—one of the largest experimentation platforms in the world—and used their data for our empirical analysis. On the statistical front, we had to think about how to use our model and Microsoft's data to estimate the firm's unobserved distribution of unobserved innovation quality, using observations of past experimental outcomes. This required a careful study of Empirical Bayes methods. These methods were first proposed by the mathematician and statistician Herbert Robbins (1956, 1964) and have only been recently used in econometrics (Gu and Koenker, 2017, 2020; Liu, Moon, and Schorfheide, 2020). A summary of our research on Empirical Bayes methods (and how to use treatment effect estimates from a large number of experiments to improve the estimates of the effects of each experiment) appeared in the paper **"Empirical Bayes Estimation of Treatment Effects with Many A/B Tests: An overview"**, published in the ***American Economic Association Papers and Proceedings***, jointly with Eduardo Azevedo, Alex Deng, and Glen Weyl (Azevedo, Deng, Montiel Olea, and Weyl (2019)).

1.2 Sample size determination for online experiments

The popularity of online experimentation has reignited interest in experimental design questions, such as how best to decide the size of an experiment. A prescient literature in statistical decision theory studied this and other questions in the 1950s and 1960s. For instance, Raiffa and Schlaifer (1961) provide a textbook treatment of the "mathematical analysis of decision making when the state of the world is uncertain but further information about it can be obtained by experimentation".

The paper **"The A/B Testing Problem with Gaussian Priors"**, ***submitted to the Journal of Economic Theory*** jointly with Eduardo Azevedo, David Mao, and Amilcar Velez (Azevedo, Mao, Montiel Olea, and Velez, 2020b), revisits this classic framework. We consider a risk-neutral firm that has an idea of unknown quality, but can perform an experiment to learn about it. We use this framework to provide a simple characterization of the experiment's optimal sample size and to conduct comparative statics.

The main message is that determining an experiment's sample size by maximizing profits is more appealing than the standard and prevailing practice of using power calculations for sample size determination. This point echoes the critiques of Manski and Tetenov (2016, 2019). The implementation of the optimal sample size is especially appealing in data-rich environments, such as experimentation in online firms, where information from past experiments is readily available and can be used to estimate the prior following an Empirical Bayes approach.

2 Macroeconomic Policy Evaluation

I initially became an economist to understand better how to evaluate and recommend policy actions, particularly concerning the macroeconomy. Macroeconomic policy evaluation is a difficult problem. The complexities largely stem from the fact that macro data is not a random sample; sample sizes are small; and policy relevant parameters are typically not identified from data. Econometrics provides a great framework to think about these issues. My work in the area of *macroeconometric* policy evaluation has focused on developing methods to help applied researchers handle some of the aforementioned complexities.

2.1 External Instruments with an application to tax policy

The impact of income tax changes on economic activity is a fundamental question in empirical macroeconomics, and a policy issue of great importance. Economic theory suggests—based on substitution effects in the labor-consumption choices of households—that an increase in labor income taxes reduces the compensation obtained from working, thereby reducing the number of hours of work supplied by households, reducing, in turn, overall economic activity.

Substitution effects that influence labor supply decisions are not the only forces at play when analyzing the effects of tax changes. As articulated clearly by Romer and Romer (2010): "Because the factors that give rise to tax changes are often correlated with other developments in the economy, disentangling the effects of the tax changes from the effects of these underlying factors is difficult".

A very exciting development in macroeconometrics—spearheaded by Romer and Romer (1989, 2010)—proposed an insighful solution to the problem: exploiting the narrative record describing the history and motivation of tax policy changes to find sources of "as-if" random variation. This "narrative approach" identifies tax changes that are unlikely to be contaminated by other developments affecting output (for example, tax changes motivated by beliefs about fairness). It took some time to realize that the constructed series of "as-if" random reforms could be viewed as an *instrumental variable* in the econometric sense of the word (Stock, 2008)

The paper with James Stock and Mark Watson, **"Inference in Structural Vector Autoregressions Identified with an External Instrument"**, forthcoming in the ***Journal of Econometrics*** (Montiel Olea, Stock, and Watson, 2021c), shows how these "external" sources of variation—termed *external instruments*—could be used to construct confidence intervals for dynamic policy effects (*impulse response coefficients*) using Structural Vector Autoregressions. A central theme is to develop a procedure not affected by the strength of the correlation between the policy variable and the external instrument. A large literature within microeconometrics has developed both diagnostics for "weak instruments" and "weak-instrument robust" procedures (including two of the chapters in my doctoral dissertation, which I will discuss in the last section of this statement). The extension of these procedures to the macroeconometric framework was not obvious.

A substantial application of the methods developed in the paper with Stock and Watson appeared in **"Marginal Tax Rates and Income: New Time Series Evidence"**, published in the ***Quarterly Journal of Economics***, jointly with Karel Mertens (Mertens and Montiel Olea, 2018). The paper uses new measures of exogenous variation in marginal tax rates associated with postwar tax reforms in the United States to estimate the dynamic effects of changes in average marginal tax rates on different variables, such as reported income (across different groups of the income distribution), Gross Domestic Product (GDP), and employment. The dynamic effects are estimated using external instruments in Structural Vector Autoregressions, but also using the popular "local projection" approach of Jordà (2005), estimated with instrumental variables as explained in Ramey (2016).

Our main empirical findings are the following. First, incomes in the top 1% of the income distribution displayed the strongest short-run responses to tax rates—roughly, a 1% decrease in the marginal tax rate increases reported income by 1.2%—consistent with the notion that high-income taxpayers engage in more avoidance behavior. Contrary to prior time series studies of tax return data, we also find statistically significant elasticities for lower income groups. Regarding economic activity, our estimates suggest that marginal rate cuts lead to increases in real GDP and declines in unemployment. However, counterfactual tax reforms cutting marginal tax rates for only the top 1% or the bottom 99% in the income distribution have very different effects. In the short-run, a top marginal rate cut is estimated to raise real GDP, and lower aggregate unemployment, but lead to greater income inequality. Targeted cuts for the bottom 99% also generate positive effects on reported incomes and aggregate economic activity, but with a delay of several years.

2.2 Local Projection Inference for impulse responses

Talking to applied macroeconomists about the work in Mertens and Montiel Olea (2018) led to the realization that the "local projections" of Jordà (2005)—direct linear regressions of future outcomes on covariates—were increasingly viewed as the default approach to estimate dynamic policy effects. This was a bit surprising: if the process generating the data is a (stationary) Vector Autoregression, then the confidence intervals for dynamic policy effects based on local projections estimators may be wider than necessary. Why would empirical researchers accept this inefficiency?

The paper with Mikkel Plagborg-Møller, **"Local Projection Inference is Simpler and More**

Robust Than You Think", forthcoming in ***Econometrica*** (Montiel Olea and Plagborg-Møller, 2021) provides theoretical results justifying the use of local projection inference. We show that—in addition to its intuitive appeal—local projection inference is robust to two common features of macroeconomic applications: highly persistent data and the estimation of impulse responses at long horizons.

Key to our result is that we consider *lag-augmented* local projections, which use lags of the variables in the regression as controls. Surprisingly, lag-augmenting local projections not only makes them more robust, but also make them simpler—by obviating the need to correct for serial correlation in the regression residuals.

Our paper builds on an important body of work in time series econometrics. Formally, our main theorem shows that standard confidence intervals based on lag-augmented projections have correct asymptotic coverage *uniformly* over the persistence in the data generating process. Our idea for using lag-augmentation came from the insightful paper of Inoue and Kilian (2020). Our study of uniform validity was inspired by the seminal work of Mikusheva (2007, 2012).

2.3 Identification and Inference in SVARs

My interest in macroeconomic policy evaluation led to the study of identification in Structural Vector Autoregressions (SVARs). It is easy to calculate the correlation between policy changes (such as changes in the average marginal tax rate) and different macroeconomic variables (such as GDP). The problem of identification can be viewed as the quest of transforming these sample correlations into estimates of dynamic policy effects (represented in the model by impulse response coefficients). This typically requires that the researcher imposes additional structure.

In the paper **"Delta-Method Inference for a Class of Set-Identified SVARs"**, published in the ***Journal of Econometrics***, jointly with Bulat Gafarov and Matthias Meier (Gafarov, Meier, and Montiel Olea, 2018), we studied SVARs that impose equality and/or inequality restrictions to set-identify dynamic policy effects. We contributed to the literature by proposing a novel delta-method confidence interval for the coefficients of the impulse response function. The paper's main observation is that the largest and smallest values of an impulse response coefficient compatible with the data distribution can be viewed as value functions of a nonlinear optimization problem. In SVARs, the data distribution is typically parameterized by a finite-dimensional parameter vector: the autoregressive coefficients and the covariance matrix of residuals. Thus, it was possible to use generalizations of the envelope theorem to characterize the directional derivatives of the largest and smallest impulse response coefficients (with respect to the VAR parameters), and then use the delta method for directionally differentiable functions as in Fang and

Santos (2019) to characterize the asymptotic distribution of plug-in estimators. Conducting inference was challenging, as it was not clear how to construct standard errors to guarantee pointwise validity. We used our results to assess the effects of the announcement of the Quantitative Easing program in August 2010.

In different seminars where we presented our delta-method paper, we were asked if a Bayesian econometrician would care at all about the irregularity (directional differentiability, and not full differentiability) of the parameters of interest. The paper, **"Posterior Distribution of Nondifferentiable Functions"**, published in the ***Journal of Econometrics***, jointly with Toru Kitagawa, Jonathan Payne, and Amilcar Velez (Kitagawa, Montiel Olea, Payne, and Velez, 2020), provides an answer to this question. The paper examines the asymptotic behavior of the posterior distribution of a possibly nondifferentiable function $g(\theta)$, where θ is a finite-dimensional parameter. The main assumption is that the distribution of a suitable estimator $\widehat{\theta}_n$, its bootstrap approximation, and the Bayesian posterior for θ are all equivalent asymptotically. We show that for transformations g that are locally Lipschitz, though not necessarily differentiable, the posterior distribution of $g(\theta)$ and the bootstrap distribution of $g(\widehat{\theta}_n)$ coincide asymptotically. One of the main implications is that credible intervals for a nondifferentiable parameter $g(\theta)$ cannot be presumed to be approximately valid confidence intervals (even if this relation holds true for θ). Our result implies that even though it is true that a Bayesian need not care about the "irregularity" of the parameters of interest, the asymptotic relation between Bayesian and frequentist procedures breaks down in the presence of such irregularities.

The two papers described above contributed to the literature on set-identified SVARs, but had two important limitations. First, Gafarov et al. (2018) was only applicable to SVARs where there were restrictions on only one of the model's structural shocks (for example, a monetary shock). This is an important limitation, because it is sometimes desirable to impose restrictions on more than one structural shock (a fiscal shock in addition to a monetary shock). Second, the inference procedure was designed for impulse responses at specific horizons; it was not clear how to conduct inference simultaneously over different horizons. This is also an important limitation, because understanding the shape of the impulse response function, and not only about its values at a particular horizon, is relevant in many applications.

The working paper **"Projection Inference for Set-Identified SVARs"**, joint with Bulat Gafarov and Matthias Meier (Gafarov, Meier, and Montiel Olea, 2016), addresses these issues by using the projection method. The idea is that the bounds of the identified set for the parameter of interest are functions of a finite-dimensional parameter (the VAR slope coefficients and covariance matrix of residuals) with an asymptotically normal distribution. It is then possible to "project" a confidence set for these parameters, by collecting the largest and smallest values of the endpoints of the identified set as the VAR parameters range in their confidence set. A well-known drawback of this projection approach is that it can be quite conservative: both the frequentist coverage, and the robust Bayes credibility criterion of Giacomini and Kitagawa (2020) can be above their intended level. Our paper suggests an approach to make the resulting confidence regions tighter, using the insights of Kaido, Molinari, and Stoye (2019) to remove projection bias by calibrating the size of the original confidence set.

Our paper is still a work in progress for two reasons. First, there have been significant theoretical advances on the problem of how to conduct inference on "subvectors" (or more generally smooth functions) of set-identified parameters (Kaido et al., 2019; Chen, Christensen, and Tamer, 2018; Bugni, Canay, and Shi, 2017). Mapping the general assumptions in these papers to the simpler set-up of SVARs has been challenging, mainly because our work focused on constraint qualifications in nonlinear programs. The connection between these types of constraints and the diverse high-level assumptions that appear in the moment-inequality literature have only been recently and lucidly analyzed in Kaido, Molinari, and Stoye (2021).

Second, it is not entirely clear to us what was the best way to conduct simultaneous inference even in point-identified SVARs. In the paper **"Simultaneous Confidence Bands: Theory, Implementation, and an Application to SVARs"**, published in the ***Journal of Applied Econometrics***, jointly with Mikkel Plagborg-Møller (Montiel Olea and Plagborg-Møller (2019)), we provide an answer to this question. We show that the popular "sup-t" band for simultaneous inference has a number of optimality properties and suggest a computationally convenient Bayesian "sup-t" band with exact finite-sample simultaneous credibility, and also asymptotically correct frequentist coverage under standard regularity conditions. Part of the work that is left to be done in Gafarov et al. (2016) is to adapt some the ideas in Montiel Olea and Plagborg-Møller (2019) on how to calibrate the credibility of the "sup-t' band, to the context of set-identified SVARs (with the target of simultaneous Robust Bayesian credibility).

3 Econometrics and Machine Learning

Haavelmo's definition of econometrics does not mention the role that *algorithms* and *computation* play in fulfilling the goals of econometric research. Both of these elements—which are the central components of the field of machine learning—are necessary to transform data and statistical methods into a concrete measurement. Thus, research aiming to connect econometrics and machine learning seems a natural continuation of the original goals of econometrics.

Since I arrived to Columbia in 2016, I have been studying several machine learning tools that I believe could be useful for economists: text analysis, variational inference, statistical learning, and dropout training (all of which I will explain below). My interest in ML arose from conversations I had with Guido Imbens in

2014, and his view on the impact that this field was bound to have in Economics. Like Athey and Imbens (2019), I believe that "being familiar with these methods will allow researchers to do more sophisticated empirical work and to communicate more effectively with researchers in other fields."

3.1 The analysis of text

Text is becoming an increasingly popular input in empirical economics research. The work I did on external instruments (Mertens and Montiel Olea, 2018; Montiel Olea et al., 2021c) allowed me to appreciate the role that textual records of macroeconomic policy—such as the transcripts of the Federal Open Market Committee (FOMC) or congressional reports on tax bills—plays in macroeconomic policy analysis.

In the paper **"Robust Machine Learning Algorithms for Text Analysis"**, ***Revise and Resubmit at Quantitative Economics***, jointly with Shikun Ke and James Nesbit (Ke, Montiel Olea, and Nesbit, 2021), we study the Latent Dirichlet Allocation (LDA) model of Blei, Ng, and Jordan (2003), a popular Bayesian tool for the analysis of text data with recent applications in economics. We study the extent to which the LDA output is determined by the choice of prior.

Our main result is that the parameters of the model are only set-identified, and so the choice of prior matters even asymptotically. We also characterize—using the theoretical Robust Bayes framework of Giacomini and Kitagawa (2020)—the upper and lower values of the posterior mean of a given (continuous) function over a reasonable class of priors (namely, those that are consistent with some fixed distribution over the population probabilities that each term has of appearing in each document). Our theoretical results suggest that the set of posterior means can be approximated by optimizing the function of interest over all possible *Nonnegative Matrix Factorizations* (which is a tool analogous to the Singular Value Decomposition, but with positivity constraints on the factors) of either the population term-document frequency matrix or its sample analogue.

Our paper provides a warning for applied researchers using the off-the-shelf implementations of the LDA: the output of the algorithm can be sensitive to the choice of prior. However, we argue this is ultimately not an issue, since it is possible to acknowledge the lack of point-identification and report identified sets for the parameters of interest (as is done, for example, in SVAR analysis).

3.2 Learning parameter regions

One challenging computational aspect of working with set-identified models is the evaluation of the identified set (broadly speaking, the set of parameter values that are compatible with the data). One early suggestion is the use of random sampling from the identified set to approximate its shape (Bar and Molinari, 2013; Horowitz, Manski, Ponomareva, and Stoye, 2003; Horowitz and Manski, 2006). Similar ideas appear in the evaluation of confidence sets, or highest-posterior density sets. But how many random points from an identified set, a confidence set, or a highest posterior density set suffice to describe them?

In the paper **"(Machine) Learning Parameter Regions"**, published in the ***Journal of Econometrics***, jointly with James Nesbit, we argue that taking random draws from a parameter region in order to approximate its shape is a supervised learning problem (analogous to sampling pixels of an image to recognize it). Misclassification error—a common criterion in machine learning—provides an off-the-shelf tool to assess the quality of a given approximation. We then say—following the literature on Probably Approximately Correct learning (Blumer, Ehrenfeucht, Haussler, and Warmuth, 1989; Valiant, 1984) —that a parameter region can be "learned" if there is an algorithm that yields a misclassification error of at most ϵ with probability at least $1 - \delta$, regardless of the sampling distribution. We show that learning a parameter region is possible if and only if its potential shapes are not too complex as in the Fundamental Theorem of Statistical Learning (Blumer et al. (1989)). Moreover, the tightest band that contains a d-dimensional parameter region is always learnable from the inside (in a sense we make precise). We provide lower and upper bounds on the number of draws required for learning that grow linearly in the dimension of the parameter region, and are uniform with respect to its true shape.

3.3 Robustness of α-posteriors and their variational approximations

As with many other probabilistic machine learning models, the parameter space in the LDA model is of a high dimension. By way of illustration, consider an LDA model with 1000 unique terms, 10 topics, with 100 different documents available for estimation. This admittedly small model has 1 million parameters! Bayesian inference with such a large number of parameters is challenging, to say the least.

A popular approach to conduct inference in these type of models is Variational Inference; see Blei, Kucukelbir, and McAuliffe (2017) for the best review of this subject. Instead of focusing on obtaining the posterior distribution of the model's parameters, Variational Inference focuses on obtaining the best approximation for either the posterior or the α-posterior (which is proportional to the product of the prior and the α-power of the likelihood), within a tractable subclass (for example, distributions with independent marginals).

In the paper **"On the Robustness to Misspecification of α-Posteriors and Their Variational Approximations"**, submitted to the ***Journal of Machine Learning Research***, jointly with Amilcar Velez and my colleagues at the Statistics Department at Columbia, Marco Avella Medina and Cynthia Rush (Avella Medina, Montiel Olea, Rush, and Velez, 2021), we study α-posteriors and their variational approximations. Our goal is to understand whether these alternatives to Bayesian Inference have any desirable theoretical properties, or if they are only useful because of their computational tractability.

We show that even in low dimensional models—where computational tractability is not an issue—α-posteriors and their variational approximations offer some *robustness-to-misspecification of the likelihood*, when tuned appropriately. In particular, they reduce the Kullback-Leibler (KL) divergence from the true, but perhaps infeasible, posterior distribution when there is potential parametric model misspecification. Also, the optimized KL divergence increases logarithmically in the degree of misspecification and not linearly as with the usual posterior.

Our analysis complements the recent contributions of Wang and Blei (2019) in the literature on Variational Inference in misspecified models. Our work is also part of the literature on Bayesian Inference under model misspecification in Econometrics, for example, the work of Müller (2013).

3.4 Dropout Training in Generalized Linear Models

Dropout training is an increasingly popular estimation method in machine learning (in particular, in the context of neural networks). The general idea consists of ignoring some dimensions of the covariate vector at random while estimating the parameters of a statistical model. A common motivation for dropout training is that the random feature selection implicitly performs *model averaging*, potentially improving out-of-sample prediction error and thus mitigating overfitting.

In the paper **"Dropout Training is Distributionally Robust Optimal"**, submitted to the ***Journal of Machine Learning Research***, jointly with José Blanchet, Yang Kang, Viet Nguyen, and Xuhui Zhang (Blanchet, Kang, Montiel Olea, Nguyen, and Zhang, 2021), we show that dropout training in Generalized Linear Models is the minimax solution of a two-player, zero-sum game where an adversarial nature corrupts a statistician's covariates using a multiplicative nonparametric errors-in-variables model. In this game, nature's *least favorable distribution* is *dropout noise*, where nature independently deletes entries of the covariate vector with some fixed probability δ. This result implies that dropout training indeed provides out-of-sample expected loss guarantees for distributions that arise from multiplicative perturbations of in-sample data.

In addition to the decision-theoretic analysis, the paper makes two more contributions. First, there is a concrete recommendation on how to select the tuning parameter δ to guarantee that, as the sample size grows large, the in-sample loss after dropout training exceeds the true population loss with some pre-specified probability. Second, the paper provides a novel, parallelizable, Unbiased Multi-Level Monte Carlo algorithm to speed up the implementation of dropout training. Our algorithm has a much smaller computational cost compared to the naive implementation of dropout based on Stochastic Gradient Descent or Monte-Carlo approximations. Our paper contributes to the growing literature explaining the success of dropout training in mitigating overfitting; e.g., Wager, Wang, and Liang (2013), Wei, Kakade, and Ma (2020).

José Blanchet and I plan to continue working in this area. We are interested in analyzing the extent to which some Machine Learning procedures (such as dropout training or the square-root LASSO of Belloni, Chernozhukov, and Wang (2011)) can be expressed as distributionally robust optimal procedures (for different types of perturbations of a baseline model), and use this representation to think about choosing the tuning parameters (such as the size of the neighborhood).

4 Police Use of Deadly Force

The use of deadly force by police officers is an issue of paramount social importance in the United States, where over a thousand civilians die at the hands of police officers every year. Public interest in police killings of civilians has increased dramatically following the fatal shooting of Michael Brown in Ferguson, Missouri in 2014. I was introduced to the intricacies of this problem through conversations with my colleague Dan O'Flaherty, and in 2019 I obtained support from the Provost's Grant Program for Junior Faculty who Contribute to the Diversity Goals of Columbia University to study this problem. I believe that police use of deadly force is a topic where the conjunction of economic theory, measurement, and statistical analysis can be extremely useful.

Meaningful discussions about police reform are challenging because law enforcement in the United States is highly decentralized, with thousands of law enforcement agencies operating separately. These agencies differ along a number of important dimensions. Some of these characteristics are observed (such as number of officers, size of population served, gun death rates) and some of them are unobserved (such as selection practices, training, promotion standards, etc). Questions of interest typically involve the construction of *counterfactuals* between pairs of agencies. For example, what would happen to the expected number of lethal encounters involving the Phoenix Police Department if this department were to adopt the selection practices, training, and organizational culture of the New York Police Department? Or what would happen to the expected number of homicides by the New York Police Department if the size of its force were reduced?

In the working paper **"Empirical Bayes Counterfactuals using Poisson Regression with an Application to Police Use of Deadly Force"**, ***submitted to the American Economic Review***, jointly with Brendan O'Flaherty and Rajiv Sethi (Montiel Olea, O'Flaherty, and Sethi, 2021a), we propose an econometric framework—based on Empirical Bayes methods (Robbins, 1956, 1964) in the context of Poisson Regression—to answer these questions.

To understand our approach consider the case in which there is only one unobserved covariate that enters into the conditional mean of the Poisson model multiplicatively. Suppose the econometrician knows i) the joint distribution of the observed and unobserved covariates, and ii) the slope coefficients of the observed covariates in the conditional mean. Under quadratic loss and some standard conditions, the best estimator of the counterfactual involving the unobserved covariate of unit j uses its *posterior mean* as a surrogate. That is, instead of evaluating the conditional mean function at the unobserved covariate of unit j, the conditional mean function is evaluated at the expected value of the unobserved covariate given the available data for unit j.

In practice the joint distribution of observed and unobserved characteristics is unknown. We exploit the Poisson likelihood function to derive a panel data analog of the celebrated formula of Robbins (1956), where the posterior mean is expressed in terms of the cross-sectional distributions of outcomes and covariates, and also on the slope coefficients of the observed variables. Under assumptions we spell out clearly—and that closely follow the work of Honoré and Kesina (2017)—the parameters that enter this formula can be estimated consistently. Plugging in these estimated parameters transforms Robbins' formula into an Empirical Bayes estimator. It is then possible to quantify the sampling uncertainty in the estimation of the counterfactuals by using delta-method/bootstrap arguments. Similar applications of Empirical Bayes methods that motivated the analysis in this paper, but in the context of normally distributed errors, appear in Liu, Moon, and Schorfheide (2020), and Gu and Koenker (2020).

Using our statistical framework—and a novel yearly panel data set of police departments with their corresponding lethal encounters for the period 2013-2018—we report counterfactual values of lethal encounters for the nation's ten largest police departments by population served, obtained by evaluating the conditional mean function at counterfactual values of unobserved and observed covariates. Our results suggest that unobserved characteristics (which include training and culture) are vitally important in determining the lethality of a police department. By way of illustration, we estimate with 90% confidence that lethal encounters in the ten largest police departments from 2013-2018 (which account for 548 killings in the data) would number between 170 and 316 if all departments had the unobserved characteristics of the New York Police Department (the lowest) and between 824 and 1384 if all agencies had the unobserved characteristics of Phoenix (the highest). This suggests there is potential for significant reduction in police killings if certain unobserved characteristics are transferable across agencies.

Our paper does not address the salient issue of racial and ethnic differences in the rate at which civilians are killed by police. This is clearly a matter of public interest. This work is in progress and will be included in a companion paper. We have recently received a Presidential Authority Grant from the Russell Sage foundation to support our research on this.

5 Other Work

5.1 Doctoral Dissertation: "Essays in Econometrics and Decision Theory (2013)"

My doctoral dissertation contained three papers: Montiel Olea and Pflueger (2013); Montiel Olea and Strzalecki (2014); Montiel Olea (2020), all of which are now published. These papers spanned the topics of weak- and set-identification, axiomatic decision theory, and statistical decision theory. These papers do not fit neatly into any of the categories I described at the beginning of this research statement. However, the tools used in these papers have been important in determining my toolkit as an econometrician, as I will explain below.

The first chapter of my dissertation, **"A Robust Test for Weak Instruments"**, published in the ***Journal of Business & Economic Statistics***, jointly with Carolin Pflueger (Montiel Olea and Pflueger, 2013), proposes a simple test for weak instruments that is robust to heteroscedasticity, serial correlation, and clustering. Our paper generalizes the "$F > 10$" rule of thumb in Staiger and Stock (1997) and Stock, Yogo et al. (2005).

Our proposed statistic, the effective F statistic, is a scaled version of the nonrobust first-stage F statistic. The null hypothesis for weak instruments is rejected for large values of the statistic. We considered two different testing procedures: generalized (valid for both TSLS and LIML) and simplified (valid only for TSLS). While the former has increased power, it is computationally more demanding. We showed that the simplified procedure for TSLS has critical values between 11 and 23.1 that depend only on the covariance matrix of the first-stage reduced form coefficients (and we provided code in STATA to compute these critical values). Thus, a simple, asymptotically valid rule of thumb is available for TSLS that rejects when the effective F is greater than 23.1. This paper has over 500 citations in Google Scholar. It is also discussed in detail in the excellent review of Andrews, Stock, and Sun (2019) on weak instruments in instrumental variables regression.

This paper was very useful to understand the role of asymptotic approximations—and the use of nonstandard asymptotics—in econometric analysis. I have continued to use nonstandard asymptotics in other recent projects. For example, in Azevedo, Deng, Montiel Olea, Rao, and Weyl (2020a) we use nonstandard asymptotics to approximate the value of information in randomized experiments with small samples. In Montiel Olea, Ortoleva, Pai, and Prat (2021b), we used it to capture the relevance of prior information in a potentially misspecified Bayesian Linear Regression. And in Avella Medina, Montiel Olea, Rush, and Velez (2021) to analyze the difference between reported and true posteriors in Bayesian models with a misspecified likelihood. At the time I was also really influenced by the work of Müller (2011), which provided a nice framework to, first, state high-level assumptions on some statistics of interests (typically under some nonstandard asymptotics); then, to treat these limiting distributions as a parametric model; and finally to solve any problem of interest in the limit (as if in Le Cam's theory of limiting experiments).

The second chapter of my dissertation is joint work with one of my Ph.D. advisers, Tomasz Stzalecki. The paper **"Axiomatization and Measurement of Quasi-Hyperbolic Discounting"**, published in the ***Quarterly Journal of Economics*** (Montiel Olea and Strzalecki, 2014), studies the popular "quasi-hyperbolic discounting" model of intertemporal choice, which is used to capture how economic agents trade off costs and benefits that occur at different periods in time. In this model, consumption in earlier periods is preferred to consumption in later periods (because the agent is impatient). However, the agent's willingness to substitute consumption between the present and the future is smaller today than at any other point in the future. Such a feature is called *present bias.*

The goal of the paper is to understand the extent to which we can measure present bias using information on choices between finitely many consumption streams. We show this is indeed possible (by using axiomatic decision theory) and we suggested an experimental design to measure present bias. The experiment consists of a delayed delivery of a preferred consumption good until the agent switches from the "patient" alternative (waiting for the delivery of the preferred consumption good) to the "impatient" choice (taking the preferred consumption good today). This exercise allows us to set-identify the distribution of present bias in the population. This was the first project in which I thought about partial identification.

The third chapter of my dissertation, **"Admissible, Similar Tests: A Characterization"**, published in ***Econometric Theory*** (Montiel Olea, 2020), used Statistical Decision Theory to understand what qualified as a "good" test of hypotheses in the linear instrumental variables model. I focus on tests satisfying two classical finite-sample properties: admissibility and similarity. Admissibility means a test cannot be improved uniformly over the parameter space. Similarity requires the null rejection probability to be unaffected by a nuisance parameter.

I characterize the tests satisfying these two properties. The characterization has two parts. The first part—which is a straightforward extension of results in Chernozhukov, Hansen, and Jansson (2009)—states that maximizing the weighted average power (WAP) subject to a similarity constraint suffices to generate tests that are both admissible and similar. The second part—which is the paper's main contribution—states that every admissible, similar procedure is an extended WAP-similar test (a concept that is made precise in the paper). I also provided a WAP-similar test for the homoskedastic instrumental variables model (that was admissible and similar by construction, and was also shown to have finite- and large-sample properties comparable to those of the conditional likelihood ratio test of Moreira (2003)).

The work I did for this chapter was very important to understand the theory of optimal hypothesis testing (especially in the context of linear instrumental variables regression). There are several papers that shaped my views on this topic, including Andrews, Moreira, and Stock (2006), Cattaneo, Crump, and Jansson (2012), Elliott, Müller, and Watson (2015), Andrews (2016), and Moreira and Moreira (2019).

5.2 Competing Models

The interaction between econometrics and microeconomic theory is quite natural when one thinks of Econometrics as a discipline focused on the use of data to solve decision problems under uncertainty—as advocated, for instance, by Gary Chamberlain's decision-theoretic approach to econometrics (Chamberlain, 2000). An excellent example of work in the intersection of econometrics, decision theory, and microeconomic theory is the positive model of empirical science by Clark Medalist of 2021 Isaiah Andrews (Andrews and Shapiro, 2021).

In the paper "**Competing Models**", ***resubmitted to the Quarterly Journal of Economics***, jointy with Pietro Ortoleva, Mallesh Pai, and Andrea Prat, we contribute to a large and growing body of work in economic theory on agents with misspecified models. Although this literature belongs naturally to microeconomic theory, it has recently seen major contributions coming from work of the econometrician Demian Pouzo (Esponda and Pouzo, 2016; Esponda, Pouzo, and Yamamoto, 2021).

In our paper, we posit a framework where agents compete to acquire an asset whose value depends on how well they can predict an unknown variable. Agents are Bayesian, observe identical data, but have different models: they use different subsets of explanatory variables to make their predictions. The "winning" model in this competition crucially depends on the sample size. With small samples, we present a number of results suggesting it is an agent using a low-dimensional model, in the sense of using a smaller number of variables relative to the true data generating process. With large samples, we show that it is generally an agent with a high-dimensional model, possibly including irrelevant variables, but never excluding relevant ones.

We use our framework to understand the proliferation of "factors" in the asset-pricing literature. We argue that the increase in the number of test portfolios used to compute the popular Fama-French cross-sectional regressions mechanically favors asset-pricing models with several factors.

6 Conclusion

The classical definition of econometrics given in the introduction of this statement has narrowed over the years. It is now customary to accompany the term econometrics with different grammatical modifiers to denote specialization. Some of these modifiers make reference to the type of data sets used for measurement (e.g., econometrics of cross-sectional data, time series, or panel data); the areas of economics from which theory is drawn from (e.g., macroeconometrics, microeconometrics, or financial econometrics); the types of statistical techniques used in the analysis (e.g., Bayesian econometrics); and whether the final goal is to provide new methodology or to use existing techniques to quantify a particular relation of interest (that is, theoretical and applied econometrics).

My research has been quite broad and spans several of these areas: I have worked in both macroeconometrics and microeconometrics; I have studied cross-sectional, time series, and panel data; and I have used both Bayesian and frequentist methods in applied and theoretical work. My research has also been highly collaborative.

I view my breadth and ability to collaborate with researchers in different areas (including micro theorists, labor economists, macroeconomists, and statisticians) as one of my biggest assets. Some of my co-authors are friends from graduate school (Eduardo Azevedo, Mikkel Plagborg-Møller, Carolin Pflueger, Glen Weyl); others are my current or former colleagues (Marco Avella, Dan O'Flaherty, Rajiv Sethi, Pietro Ortoleva, Andrea Prat, Cindy Rush); others have worked with me while they were graduate or undergraduate students (Barry Ke, Bulat Gafarov, David Mao, Matthias Meier, James Nesbit, Jonathan Payne); and some other are people whose work I have always admired and that I always wanted to work with (José Blanchet, Toru Kitagawa, Karel Martens, James Stock, Tomasz Strzalecki, Mark Watson). I really enjoy working in groups. I believe that working with researchers of different backgrounds is a very effective way of writing for a general audience (which increases the likelihood of top publications in economics).

Despite the heterogeneity in my work, all my papers share a common objective. The econometrics chapter of the New Palgrave Dictionary of Economics states it succinctly: "Econometric theory and practice seek to provide information required for informed decision making in public and private economic policy." This has been—and will continue to be—the goal of my research.

References

AVELLA MEDINA, M., J. L. MONTIEL OLEA, C. RUSH, AND A. VELEZ (2021): "On the Robustness to Misspecification of α-Posteriors and Their Variational Approximations," ***Submitted to the Journal of Machine Learning Research***.

AZEVEDO, E. M., A. DENG, J. L. MONTIEL OLEA, J. RAO, AND E. G. WEYL (2020a): "A/B Testing with Fat Tails," ***Journal of Political Economy***, Vol. 128, pp. 4614–000.

AZEVEDO, E. M., A. DENG, J. L. MONTIEL OLEA, AND E. G. WEYL (2019): "Empirical Bayes Estimation of Treatment Effects with Many A/B Tests: An Overview," ***AEA Papers and Proceedings***, Vol. 109, pp. 43–47.

AZEVEDO, E. M., D. MAO, J. L. MONTIEL OLEA, AND A. VELEZ (2020b): "The A/B Testing Problem with Gaussian Priors," ***Submitted to the Journal of Economic Theory***.

BLANCHET, J., Y. KANG, J. L. MONTIEL OLEA, V. A. NGUYEN, AND X. ZHANG (2021): "Dropout Training is Distributionally Robust Optimal," ***Submitted to the Journal of Machine Learning Research***.

GAFAROV, B., M. MEIER, AND J. L. MONTIEL OLEA (2016): "Projection Inference for Set-identified SVARs," ***Manuscript, Columbia University***.

——— (2018): "Delta-Method Inference for a Class of Set-identified SVARs," ***Journal of Econometrics***, Vol. 203, pp. 316–327.

KE, S., J. L. MONTIEL OLEA, AND J. NESBIT (2021): "A Robust Machine Learning Algorithm for Text Analysis," ***Revise and Resubmit at Quantitative Economics***.

KITAGAWA, T., J. L. MONTIEL OLEA, J. PAYNE, AND A. VELEZ (2020): "Posterior Distribution of Nondifferentiable Functions," ***Journal of Econometrics***, Vol. 217, pp. 161–175.

MERTENS, K. AND J. L. MONTIEL OLEA (2018): "Marginal Tax Rates and Income: New Time Series Evidence," ***Quarterly Journal of Economics***, Vol. 133, pp. 1803–1884.

MONTIEL OLEA, J. L. (2020): "Admissible, Similar Tests: A Characterization," ***Econometric Theory***, Vol. 36, pp. 347–366.

MONTIEL OLEA, J. L., B. O'FLAHERTY, AND R. SETHI (2021a): "Empirical Bayes Counterfactuals for Poisson Regression with an Application to Police Use of Deadly Force," ***Submitted to the American Economic Review***.

MONTIEL OLEA, J. L., P. ORTOLEVA, M. PAI, AND A. PRAT (2021b): "Competing Models," ***Resubmitted to the Quarterly Journal of Economics***.

Montiel Olea, J. L. and C. Pflueger (2013): "A Robust Test for Weak Instruments," ***Journal of Business & Economic Statistics***, Vol. 31, pp. 358–369.

Montiel Olea, J. L. and M. Plagborg-Møller (2019): "Simultaneous Confidence Bands: Theory, Implementation, and an Application to SVARs," ***Journal of Applied Econometrics***, Vol. 34, pp. 1–17.

——— (2021): "Local Projection Inference is Simpler and More Robust Than You Think," ***Forthcoming in Econometrica***.

Montiel Olea, J. L., J. H. Stock, and M. W. Watson (2021c): "Inference in Structural Vector Autoregressions Identified with an External Instrument," ***Forthcoming in the Journal of Econometrics***.

Montiel Olea, J. L. and T. Strzalecki (2014): "Axiomatization and Measurement of Quasi-Hyperbolic Discounting," ***Quarterly Journal of Economics***, Vol. 129, pp. 1449–1499.

Other Cited Works

Andrews, D. W., M. J. Moreira, and J. H. Stock (2006): "Optimal two-sided invariant similar tests for instrumental variables regression," *Econometrica*, Vol. 74, pp. 715–752.

Andrews, I. (2016): "Conditional linear combination tests for weakly identified models," *Econometrica*, 84, 2155–2182.

Andrews, I. and J. M. Shapiro (2021): "A model of scientific communication," Tech. rep., Forthcoming in Econometrica.

Andrews, I., J. H. Stock, and L. Sun (2019): "Weak instruments in instrumental variables regression: Theory and practice," *Annual Review of Economics*, Vol. 11, pp. 727–753.

Athey, S. and G. W. Imbens (2019): "Machine Learning Methods that Economists Should Know About," *Annual Review of Economics*, Vol. 11, pp. 685–725.

Bar, H. and F. Molinari (2013): "Computational Methods for Partially Identified Models via Data Augmentation and Support Vector Machines," Tech. rep., Cornell University Working Paper.

Belloni, A., V. Chernozhukov, and L. Wang (2011): "Square-root lasso: pivotal recovery of sparse signals via conic programming," *Biometrika*, Vol. 98, pp. 791–806.

Blei, D. M., A. Kucukelbir, and J. D. McAuliffe (2017): "Variational Inference: A Review for Statisticians," *Journal of the American Statistical Association*, Vol. 112, pp. 859–877.

Blei, D. M., A. Y. Ng, and M. I. Jordan (2003): "Latent Dirichlet Allocation," *Journal of Machine Learning research*, Vol. 3, pp. 993–1022.

Blumer, A., A. Ehrenfeucht, D. Haussler, and M. K. Warmuth (1989): "Learnability and the Vapnik-Chervonenkis dimension," *Journal of the ACM (JACM)*, Vol. 36, pp. 929–965.

Bugni, F. A., I. A. Canay, and X. Shi (2017): "Inference for Subvectors and Other Functions of Partially Identified Parameters in Moment Inequality Models," *Quantitative Economics*, Vol. 8, pp. 1–38.

Cattaneo, M. D., R. K. Crump, and M. Jansson (2012): "Optimal inference for instrumental variables regression with non-Gaussian errors," *Journal of Econometrics*, 167, 1–15.

Chamberlain, G. (2000): "Econometrics and decision theory," *Journal of Econometrics*, Vol. 95, pp. 255–283.

Chen, X., T. M. Christensen, and E. Tamer (2018): "Monte Carlo Confidence Sets for Identified Sets," *Econometrica*, Vol. 86, pp. 1965–2018.

Chernozhukov, V., C. Hansen, and M. Jansson (2009): "Admissible invariant similar tests for instrumental variables regression," *Econometric Theory*, pp. 806–818.

Elliott, G., U. K. Müller, and M. W. Watson (2015): "Nearly optimal tests when a nuisance parameter is present under the null hypothesis," *Econometrica*, Vol. 83, pp. 771–811.

Esponda, I. and D. Pouzo (2016): "Berk–Nash equilibrium: A framework for modeling agents with misspecified models," *Econometrica*, 84, 1093–1130.

Esponda, I., D. Pouzo, and Y. Yamamoto (2021): "Asymptotic behavior of Bayesian learners with misspecified models," *Journal of Economic Theory*, Vol. 195, pp. 105260.

Fang, Z. and A. Santos (2019): "Inference on Directionally Differentiable Functions," *The Review of Economic Studies*, Vol. 86, pp. 377–412.

Frisch, R. (1933): "Editorial Note," *Econometrica*, Vol. 1, pp. 1–4.

Giacomini, R. and T. Kitagawa (2020): "Robust Bayesian Inference for Set-identified Models," Tech. rep., Forthcoming in Econometrica.

Gu, J. and R. Koenker (2017): "Unobserved Heterogeneity in Income Dynamics: An Empirical Bayes Perspective," *Journal of Business & Economic Statistics*, Vol. 35, pp. 1–16.

——— (2020): "Invidious Comparisons: Ranking and Selection as Compound Decisions," *arXiv preprint arXiv:2012.12550*.

Haavelmo, T. (1944): "The Probability Approach in Econometrics," *Econometrica*, Vol. 12, Supplement, pp. iii–115.

Honoré, B. E. and M. Kesina (2017): "Estimation of Some Nonlinear Panel Data Models With Both Time-Varying and Time-Invariant Explanatory Variables," *Journal of Business & Economic Statistics*, Vol. 35, pp. 543–558.

Horowitz, J. L. and C. F. Manski (2006): "Identification and Estimation of Statistical Functionals using Incomplete Data," *Journal of Econometrics*, Vol. 132, pp. 445–459.

Horowitz, J. L., C. F. Manski, M. Ponomareva, and J. Stoye (2003): "Computation of Bounds on Population Parameters when the Data are Incomplete," *Reliable Computing*, Vol. 9, pp. 419–440.

Inoue, A. and L. Kilian (2020): "The Uniform Validity of Impulse Response Inference in Autoregressions," *Journal of Econometrics*, Vol. 215, pp. 450–472.

Jordà, Ò. (2005): "Estimation and Inference of Impulse Responses by Local Projections," *American Economic Review*, 161–182.

Kaido, H., F. Molinari, and J. Stoye (2019): "Confidence intervals for Projections of Partially Identified Parameters," *Econometrica*, Vol. 87, pp. 1397–1432.

——— (2021): "Constraint Qualifications in Partial Identification," *Econometric Theory*, pp. 1–24.

Liu, L., H. R. Moon, and F. Schorfheide (2020): "Forecasting with Dynamic Panel Data Models," *Econometrica*, Vol. 88, pp. 171–201.

Manski, C. F. and A. Tetenov (2016): "Sufficient Trial Size to Inform Clinical Practice," *Proceedings of the National Academy of Sciences*, Vol. 113, 10518–10523.

——— (2019): "Trial Size for Near-Optimal Choice Between Surveillance and Aggressive Treatment: Reconsidering MSLT-II," *The American Statistician*, Vol. 73, pp. 305–311.

Mikusheva, A. (2007): "Uniform Inference in Autoregressive Models," *Econometrica*, Vol. 75, pp. 1411–1452.

——— (2012): "One-dimensional Inference in Autoregressive Models with the Potential Presence of a Unit Root," *Econometrica*, Vol. 80, pp. 173–212.

Moreira, H. and M. J. Moreira (2019): "Optimal two-sided tests for instrumental variables regression with heteroskedastic and autocorrelated errors," *Journal of Econometrics*, 213, 398–433.

Moreira, M. J. (2003): "A conditional likelihood ratio test for structural models," *Econometrica*, 71, 1027–1048.

Müller, U. K. (2011): "Efficient tests under a weak convergence assumption," *Econometrica*, Vol. 79, pp. 395–435.

——— (2013): "Risk of Bayesian inference in misspecified models, and the sandwich covariance matrix," *Econometrica*, Vol. 81, pp. 1805–1849.

Radner, R. and J. Stiglitz (1984): "A Nonconcavity in the Value of Information," *Bayesian models in economic theory*, Vol. 5, pp. 33–52.

Raiffa, H. and R. Schlaifer (1961): "Applied Statistical Decision Theory." .

Ramey, V. A. (2016): "Macroeconomic Shocks and their Propagation," *Handbook of macroeconomics*, Vol. 2, pp. 71–162.

Robbins, H. (1956): "An Empirical Bayes Approach to Statistics," *Proc. Third Berkeley Symp. on Math. Statist. and Prob*, Vol. 1, pp. 157–163.

——— (1964): "The Empirical Bayes Approach to Statistical Decision Problems," *The Annals of Mathematical Statistics*, pp. 1–20.

Romer, C. D. and D. H. Romer (1989): "Does Monetary Policy Matter? A New Test in the Spirit of Friedman and Schwartz," *NBER Macroeconomics Annual*, Vol. 4, pp. 121–170.

——— (2010): "The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Shocks," *American Economic Review*, Vol. 100, pp. 763–801.

Staiger, D. and J. H. Stock (1997): "Instrumental variables regression with weak instruments," *Econometrica: journal of the Econometric Society*, pp. 557–586.

Stock, J. H. (2008): "NBER Summer Institute Minicourse 2008: What is New in Econometrics–Time Series, Lecture 7: Structural VARs," *Cambridge, Mass.: National Institute for Economic Research. www.nber.org/minicourse_2008. html.*

Stock, J. H., M. Yogo, et al. (2005): "Testing for weak instruments in linear IV regression," *Identification and inference for econometric models: Essays in honor of Thomas Rothenberg*, Vol. 80, pp. 1.

Tinbergen, J. (1951): *Econometrics*, Philadelphia: Blakiston and Co.

Valiant, L. G. (1984): "A Theory of the Learnable," *Communications of the ACM*, Vol. 27, pp. 1134–1142.

Wager, S., S. Wang, and P. S. Liang (2013): "Dropout Training as Adaptive Regularization," in *Advances in Neural Information Processing Systems 26*, pp. 351–359.

Wang, Y. and D. Blei (2019): "Variational Bayes under Model Misspecification," in *Advances in Neural Information Processing Systems*, ed. by H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, Curran Associates, Inc., vol. Vol. 32.

Wei, C., S. Kakade, and T. Ma (2020): "The Implicit and Explicit Regularization Effects of Dropout," in *Proceedings of the International Conference of Machine Learning*.