UNDERSTANDING NONCOMPETITION AGREEMENTS: THE 2014 NONCOMPETE SURVEY PROJECT

J.J. Prescott,[*] Norman D. Bishara[**] & Evan Starr[***]

ABSTRACT

In recent years, scholars and policymakers have devoted considerable attention to the potential consequences of employment noncompetition agreements and to whether legislatures ought to reform the laws that govern the enforcement of these controversial contractual provisions. Unfortunately, much of this interest—and the content of proposed reforms—derives from anecdotal tales of burdensome noncompetes among low-wage workers and from scholarship that is either limited to slivers of the population (across all studies, less than 1%) or relies on strong assumptions about the incidence of noncompetition agreements. Better understanding of the use of noncompetes and effective noncompetition law reform requires a more complete picture of the frequency and distribution of

    [*]   Professor of Law, University of Michigan Law School; Professor of Economics (courtesy), University of Michigan; jprescott@umich.edu

    [**]   Associate Professor of Business Law and Ethics, Stephen M. Ross School of Business at the University of Michigan; nbishara@umich.edu

    [***]   Assistant Professor, Robert H. Smith School of Business, University of Maryland; estarr@rhsmith.umd.edu

noncompetes at the individual employee level. Accordingly, in 2014, we administered a nationwide survey of individuals in the labor force to ask them about their employment status, history, and future expectations—including their experience with and understanding of noncompetition agreements. In this Article, we describe the methods we used to carry out this survey and refine the data for analysis in hopes of encouraging other researchers to use survey approaches to fill other, similarly important gaps in our knowledge. To illustrate the value of the survey project, we present a surprising empirical finding from our data, one that raises questions about existing scholarship and theories about why employers use noncompetes: We find little evidence that the incidence of noncompetition agreements in a state (after controlling for potentially confounding factors) has any relationship to the level of enforcement of such agreements in that state. In other words, an employee in California (where noncompetes are prohibited) appears to be just as likely to labor under a noncompete as an employee in Florida (where noncompetes are much more likely to be enforced).

TABLE OF CONTENTS

INTRODUCTION

Restrictive covenants such as covenants not to compete (CNCs) ("noncompetition agreements" or, simply, "noncompetes"),[1] nondisclosure agreements (NDAs) (also called "confidentiality agreements"),[2] and nonsolicitation agreements (NSAs)[3] are undoubtedly a regular facet of employment relationships in the United States. But just how common a feature of employment are agreements like noncompetes, exactly? Which kinds of workers in the U.S. labor force have agreed to these restrictions—and do they bargain over them or even understand the terms to which they agree to adhere? Which employers in which industries across the United States use these agreements—and why do they use them? What connections are there between the use of these agreements and underlying state law? Is there a need for legal or policy reform—a

---

1. For our purposes, we define a noncompete as a contract between an employer and an employee that limits the employee's ability to engage in future competitive activities—i.e., either working for a competitor or starting a competing enterprise. In this Article, we focus on the employment restrictions that may affect the post-employment mobility of an employee, not the class of noncompetes related to the sale of the goodwill of a business. *See, e.g.*, Rachel S. Arnow-Richman, *Bargaining for Loyalty in the Information Age: A Reconsideration of the Role of Substantive Fairness in Enforcing Employee Noncompetes*, 80 OR. L. REV. 1163, 1169 (2001); *see also, e.g.*, Norman D. Bishara et al., *An Empirical Analysis of Noncompetition Clauses and Other Restrictive Postemployment Covenants*, 68 VAND. L. REV. 1, 6 n.8 (2015).

2. In the employment context, these agreements prohibit the disclosure of information of the employer that the employee acquired as part of the employment relationship. This information need not qualify as a trade secret, and the term of the restriction is generally indefinite. *See* Terry Morehead Dworkin & Elletta Sangrey Callahan, *Buying Silence*, 36 AM. BUS. L.J. 151, 155 (1998).

3. These agreements temporarily prohibit former employees from recruiting former co-workers or pursuing business with the clients of their former employers. *See* Bishara et al., *supra* note 1, at 7; *see also* David L. Johnson, *The Parameters of "Solicitation" in an Era of Non-Solicitation Covenants*, 28 A.B.A. J. LAB. & EMP. L. 99 (2012) (critiquing nonsolicitation provisions).

question many state legislators have begun pondering in recent years—and if so, what changes are needed to achieve specific outcomes, such as fairness for employees, innovation, training and investment, and knowledge protection?

The answers to these empirical questions and many others are critical to understanding the behavior of employers and employees, the causes and consequences of noncompetition agreements, and whether and how to reform the law of restrictive covenants in this country. Yet even a rough answer to the seemingly foundational first question—how common are noncompetition agreements in the United States?—remains to date conspicuously absent.

In fact, we know surprisingly little about the frequency, scope, and strength of noncompetition agreements in this country. We know even less about how differences across jurisdictions in the law of noncompetes and in enforcement behavior relate to the prevalence and content of such agreements.[4] Notwithstanding this dearth of basic information, there has been a near explosion in the attention being paid to noncompetes and their effects. Policymakers and commentators have been engrossed. Researchers, for their part, have published many provocative, but ultimately limited, exploratory studies about the many roles that noncompetes play in employment relationships and the economy.[5] Much of this chorus has been fueled by unsupported assumptions and by high-profile anecdotal evidence of purportedly abusive practices involving noncompetes.[6]

Despite a long and storied history, these restrictive covenants remain actively derided by the public, yet widely allowed by courts

---

4.   *See, e.g.*, Toby E. Stuart & Olav Sorenson, *Liquidity Events and the Geographic Distribution of Entrepreneurial Activity*, 48 ADMIN. SCI. Q. 175, 182 (2003) ("The prevalence of non-compete covenants in employment contracts remains unknown, but available data suggest that they may be nearly ubiquitous in employment contracts in high technology businesses.").

5.   *See infra* Part I.

6.   Two recent high-profile examples involve the expansive use of noncompetes by Jimmy John's and Amazon. *See* Dave Jamieson, *Jimmy John's Makes Low-Wage Workers Sign 'Oppressive' Noncompete Agreements*, HUFFINGTON POST (Oct. 13, 2014, 4:03 PM), http://www.huffingtonpost.com/2014 /10/13/jimmy-johns-noncompete_n_5978180.html [https://perma.cc/9PK7-2U4U] (discussing noncompete usage by the fast-food sandwich franchise for its low-wage sandwich makers and delivery drivers); *see also* Spencer Woodman, *Amazon Makes Even Temporary Warehouse Workers Sign 18-Month Non-Competes*, VERGE (Mar. 26, 2015, 11:44 AM), http://www.theverge.com/2015/3/26/8280309/amazon-warehouse-jobs-exclusive-noncompete-contracts [https://perma.cc/4HS9-2JUW] (describing the use by Amazon of noncompete agreements for its seasonal, low-skilled, and low-wage warehouse workers).

in most states.[7] At least some employers embrace noncompetes as a necessary tool for knowledge and human capital investment protection;[8] many employees, by contrast, fear noncompetes as an unfair restriction on their mobility;[9] and certain subject-matter experts cite noncompetes as a harmful constraint on markets and technological innovation.[10]

We find this disconnect between the keen interest and active debate *around* noncompetes and the lack of credible evidence *about* noncompetes both curious and frustrating. Our intuition and experience tell us that noncompetes and other restrictive covenants are likely to affect employee and employer behavior, and may have important knock-on effects on innovation levels, new venture creation, and employee mobility. We also know that many scholars and policymakers have long debated the fairness aspects of restrictive covenants—particularly their supposed tendency to limit employee freedom.[11] While much noncompete research has been descriptive and focused on the common law, cross-disciplinary empirical work has made important headway of late.[12] Nevertheless, it is clear that—for the most part—this literature comprises unconnected, piecemeal, and often abstract articles. These articles do not have a basic, foundational understanding of noncompete contracting behavior "on the ground" or even a full sense of state-by-state noncompete enforcement realities.

---

7. *See generally* Harlan M. Blake, *Employee Agreements Not to Compete*, 73 HARV. L. REV. 625, 629-70 (1960) (providing a historical view of the evolution of common law noncompete enforcement).

8. *See generally* Mark J. Garmaise, *Ties That Truly Bind: Noncompetition Agreements, Executive Compensation, and Firm Investment*, 27 J.L. ECON. & ORG. 376 (2009) (studying the effect of noncompetes on employers' capital expenditures per employee).

9. *See generally* Matt Marx et al., *Mobility, Skills, and the Michigan Non-Compete Experiment*, 55 MGMT. SCI. 875 (2009) (investigating Michigan's 1985 reversal of its noncompete enforcement policy and analyzing the effect of the change in law on employee mobility).

10. *See generally* Sampsa Samila & Olav Sorenson, *Noncompete Covenants: Incentives to Innovate or Impediments to Growth*, 57 MGMT. SCI. 425 (2011) (examining how noncompete enforcement may limit entrepreneurship).

11. *See* Arnow-Richman, *supra* note 1, at 1166 ("A central concern in the law and scholarship regarding noncompete agreements has long been the extent to which enforcement should be constrained to protect the mobility and economic freedom of workers.").

12. *See, e.g.*, Norman D. Bishara & David Orozco, *Using the Resource-Based Theory to Determine Covenant Not to Compete Legitimacy*, 87 IND. L.J. 979 (2012); On Amir & Orly Lobel, *Driving Performance: A Growth Theory of Noncompete Law*, 16 STAN. TECH. L. REV. 833 (2013).

Recent media reports have cataloged a number of egregious-sounding noncompete practices.[13] These disclosures appear to have influenced public perception, generating speculative conclusions that noncompete use is on the rise.[14] News that employers have used noncompete restrictions with hair stylists, yoga instructors, lawn sprayers, teenage camp counselors,[15] low-wage fast-food workers,[16] and temporary warehouse workers,[17] to cite just a few examples, strikes many as both surprising and inexplicable. These sorts of revelations draw ire and condemnation both from the public and from politicians.[18] More than a few have decided that the availability of enforceable noncompetition agreements necessarily results in unfairness and that we need a decisive solution—perhaps even a ban.[19] This groundswell of attention has culminated in a just-released U.S. Treasury report summarizing the state of noncompete research and crystallizing the issues of the controversy,[20] as well as a just-released White House report on noncompetes that trumpets the need

---

13.   *See, e.g.*, Associated Press, *Scrutiny on Worker Non-Compete Deals*, CHATTANOOGA TIMES FREE PRESS (Jan. 3, 2015), http://www.timesfreepress.com/ news/national/georgia/story/2015/jan/03/scrutiny-worker-non-compete-deals/280886/ [https://perma.cc/D36A-92TW] (documenting the use of noncompetes for low-wage workers and observing that noncompetes might trap such employees in their current jobs and allow employers to pay them less as a result).

14.   *See, e.g.*, Steven Greenhouse, *Noncompete Clauses Increasingly Pop Up in Array of Jobs*, N.Y. TIMES, June 8, 2014, at B1 (concluding that there has been an increase in the use of noncompetes based on a review of practitioners' litigation databases which indicate more disputes).

15.   *See, e.g.*, *id.* (mentioning surprising uses of noncompetes alongside the more well-known examples of restricting executives and technology workers).

16.   *See, e.g.*, Jamieson, *supra* note 6.

17.   *See, e.g.*, Woodman, *supra* note 6.

18.   *See, e.g.*, Evan Smith, *State House Committee Approves Stanford Bill to Ban Non-Compete Agreements*, HERALD*NET* (Feb. 20, 2015, 3:02 PM), http://www. heraldnet.com/article/20150220/BLOG5207/150229856   [https://perma.cc/TP6C-4U78] (quoting the sponsor of the bill to ban noncompetes as stating that "overused noncompete agreements reinforce the wealthy, stifle startups and inhibit the best and brightest future entrepreneurs").

19.   *See, e.g.*, Jordan Weissmann, *Ban Noncompete Agreements. Do It Now.*, SLATE (June 4, 2015, 4:39 PM), http://www.slate.com/blogs/moneybox/2015/06/04/ forcing_low_wage_workers_to_sign_noncompete_agreements_is_vicious_and_stupid. html [https://perma.cc/JBR6-RBAJ].

20.   Office of Economic Policy, U.S. Department of the Treasury, Non-compete Contracts: Economic Effects and Policy Implications 3 (Mar. 2016) [hereinafter Treasury Report], https://www.treasury.gov/resource-center/economic-policy/Documents/UST%20Non-competes%20Report.pdf [https://perma.cc/7GV5-2YHV] (examining the benefits and costs of noncompetes and concluding that "many of these benefits come at the expense of workers and the broader economy").

for additional academic research and careful policy discussion to inform this debate and potentially mend the fractured state of noncompete law across the country.[21]

   However, many observers ignore pivotal questions about these stories. To begin with, it is not clear that the employers in these examples regularly attempt to—or legally could succeed in— enforcing noncompetes against their former low-wage employees.[22] We also do not know if these agreements in fact limit employees' mobility, keeping them in their positions for longer than they would otherwise have stayed, all else equal, and perhaps precluding them from advancing in their careers. Even if mobility is lower as a result of such contracts, we have not yet determined whether affected employees are compensated for that sacrifice, whether through raises, greater job security, or internal promotion. And finally if these employees were unaware of their noncompetes and if enforcement activity is in fact minimal,[23] any impact of these contracts on employee behavior or welfare may have been negligible.

---

   21.   The White House, Non-Compete Agreements: Analysis of the Usage, Potential Issues, and State Responses (May 2016) [hereinafter White House Report], https://www.whitehouse.gov/sites/default/files/non-competes_report_final2.pdf [https://perma.cc/7GZM-JAHT].

   In the coming months, as part of the Administration's efforts to support competition in consumer product and labor markets, the White House, Treasury, and the Department of Labor will convene a group of experts in labor law, economics, government and business to facilitate discussion on non-compete agreements and their consequences. The goal will be to identify key areas where implementation and enforcement of non-competes may present issues, to examine promising practices in states, and put forward a set of best practices and call to action for state reform. By facilitating a dialogue between academic experts and those with practical expertise, we aim to identify policies that could be used to promote a fair and dynamic labor market, while remaining cognizant of real world challenges to reform. We also aim to prompt further research exploring the use and the effects of non-compete agreements.

Id. at 3.

   22.   In fact, in some cases, employers explicitly decline to enforce these agreements. *See, e.g.*, Brunner v. Liautaud, No. 14-c-5509, 2015 WL 1598106, at *1, *10 (N.D. Ill. Apr. 8, 2015). In *Brunner*, one current and one former employee sued Jimmy John's seeking, in part, a judgment voiding their noncompetes. While the court dismissed the claims for lack of standing, it found that even if there had been standing, it would not have been able to overlook the fact that the employer never enforced or intended to enforce breaches of its noncompetes. *Id.* at *10.

   23.   *See id.* (asserting that even if a worker had alleged a sufficient injury for standing, he or she "still cannot overcome the Franchisee Defendants and Jimmy John's intention [included in two affidavits] not to enforce any breach of the Confidentiality and Non-Competition Agreements").

More generally, we do not know if these agreements—especially for low-wage employees, but even for more educated, skilled, and highly paid workers—are truly commonplace. Are these anecdotes outliers or instead the proverbial "tip of the iceberg," and thus emblematic of an important policy domain in need of sustained collective concentration? Additionally, even if noncompetes are relatively common and harmful for low-wage workers, just how pervasive are they generally, and what purposes do they serve with regard to other types of employees? Do these agreements provide economic or social benefits for employers or for employees themselves, including increased investment in training? These are not merely hypothetical concerns. Many U.S. jurisdictions are in the midst of designing, proposing, and debating policy alternatives,[24] all of which are tethered to untested assumptions about the frequency, scope, consequences, and purposes of noncompetes.

For all of these reasons, in 2014, we designed and implemented a comprehensive survey of U.S. labor force participants—what we call the 2014 Noncompete Survey Project. Our principal purpose in this Article is simply to describe our data-collection undertaking in detail, in the hopes of convincing other legal and policy scholars to engage in similar efforts.

We begin in Part I with our primary motivation for the survey: furnishing basic data on the frequency of noncompetes, their substance, and what employees know about them. We review the various pockets of research related to noncompetes and highlight key data problems, developing policy reactions, and unsupported assumptions that run through the commentary and scholarship—all of which contributed to our decision to undertake such an extensive survey project. In Part II, we describe our survey and methodology at length. We discuss its content and design, its online implementation, how we cleaned and refined its data, and our treatment of sample

---

24.   *See, e.g.*, Editorial, *Noncompete Agreements: Follow Calif. Lead and Scrap Them*, Bos. Globe (Apr. 15, 2014), https://www.bostonglobe.com/opinion/editorials/2014/04/15/noncompete-agreements-follow-calif-lead-and-scrap-them/tuZk0HoOgfK5mrt9dSgl7K/story.html (urging swift legislative support for former Massachusetts Governor Deval Patrick's proposed reforms to the law governing noncompetes, which continue to stall in the state); *see also, e.g.*, Tom Erickson, *Mass. Should Get Rid of Noncompete Agreements*, Bos. Globe (May 14, 2014), https://www.bostonglobe.com/opinion/2014/05/14/mass-should-get-rid-non-compete-agreements/pYEjtd4BXqTUclnUoWkZlO/story.html (advocating for a legislative ban against noncompetes in Massachusetts in an opinion piece written by the CEO of a Massachusetts-based technology firm that voluntarily abandoned noncompetition agreements).

selection concerns. In Part III, we demonstrate the data's value by exploring the relationship between noncompete frequency and enforceability. Curiously, we find little credible evidence of any relationship between the strength of enforcement at the state level and employer noncompete use by state, raising questions about the practical significance of standard reform efforts.

We briefly conclude this Article by outlining future research opportunities arising out of the 2014 Noncompete Survey Project. We believe that our survey data will provide an important means of assessing the impact noncompete agreements have on employer and, particularly, employee choices. In the long run, we hope that our survey and the research it engenders will lead to a much better understanding not only of employment contracting activity, but of how best to improve the employment relationship through targeted, careful, and evidence-based reform.

I. PROBLEMS, ASSUMPTIONS, AND SOLUTIONS

Noncompetes have been a source of much debate in recent years.[25] Scholarly research has helped to shape and fuel this debate,[26] and new research efforts are responding to the growing dispute over restrictive covenants.[27] However, these efforts focus on the perceived widespread use and abuse of noncompetes, especially among low-wage and less empowered employees,[28] as well as the hypothesized

---

25. *See* Robert W. Gomulkiewicz, *Leaky Covenants-Not-to-Compete as the Legal Infrastructure for Innovation*, 49 U.C. DAVIS L. REV. 251, 253-56 (2015) (highlighting the flurry of academic articles and the scholarly debate surrounding noncompete reform in the last two decades).

26. *See, e.g.*, Ronald J. Gilson, *The Legal Infrastructure of High Technology Industrial Districts: Silicon Valley, Route 128, and Covenants Not to Compete*, 74 N.Y.U. L. REV. 575, 578-80 (1999) (proposing that the innovation and rapid labor mobility in California and Silicon Valley in particular are an outgrowth of California's longstanding statute banning noncompetes); Marx et al., *supra* note 9, at 875-76 (building on and partially testing Gilson's thesis about the role of noncompetes by exploiting a temporary change in Michigan's noncompete enforcement policy to study the movement of engineers with patents).

27. *See, e.g.*, Timothy P. Glynn, *Interjurisdictional Competition in Enforcing Noncompetition Agreements: Regulatory Risk Management and the Race to the Bottom*, 65 WASH. & LEE L. REV. 1381, 1418-24 (2008); *see also* Gomulkiewicz, *supra* note 25, at 258; Griffin Toronjo Pivateau, *Enforcement of Noncompetition Agreements: Protecting Public Interests Through an Entrepreneurial Approach*, 46 ST. MARY'S L.J. 483, 484 (2015).

28. *See, e.g.*, Arnow-Richman, *supra* note 1, at 1166-67 (emphasizing the ubiquity of noncompetes across a broad range of sectors and proposing an alternative theory of enforcement).

broader effects on innovation,[29] mobility,[30] work effort, training and investment,[31] economic growth,[32] and the like.[33] Yet, obtaining a sense of the *actual* prevalence and use of noncompete agreements in the United States has been glossed over at best—and at worst ignored by existing scholarship.

The absence of comprehensive data about noncompete activity is entirely predictable: Noncompetes are, at least in theory, the fruit of private contracting between individual employers and employees, negotiated on a case-by-case basis.[34] These agreements are private documents concerning confidential business interests dispersed across many millions of distinct parties. As a result, they are almost impossible to track down and catalog in a systematic and representative way. Moreover, many of the parties have no incentive to make their contracting practices and existing employment relationships transparent, even to researchers. Indeed, they often have strategic reasons *not* to make their practices known. It was with the goal of offering at least a partial remedy for this information deficit that we conceived, designed, and implemented the 2014 Noncompete Survey that we describe in detail in Part II.

---

29.   *See, e.g.*, Gilson, *supra* note 26, at 578 (discussing the hampering effect of noncompetes on technological innovation).

30.   *Id.* at 579 (discussing a similar hampering-like effect of noncompetes on employee mobility).

31.   *See, e.g.*, Evan Starr, Consider This: Firm-Sponsored Training and the Enforceability of Covenants Not to Compete 16-17 (Nov. 5, 2015) (unpublished manuscript), https://sites.google.com/site/starrevan/research [https://perma.cc/7VR7-KAS7] (investigating the impact of noncompetes on firm investments in employee training).

32.   *See, e.g.*, Samila & Sorenson, *supra* note 10, at 435-37 (discussing findings that noncompetes hinder agglomeration economies); Stuart & Sorenson, *supra* note 4, at 182-84 (suggesting enforceable noncompetes may indirectly hamper new venture creation); *see also* ORLY LOBEL, TALENT WANTS TO BE FREE: WHY WE SHOULD LEARN TO LOVE LEAKS, RAIDS, AND FREE RIDING 49-75 (2013) (viewing noncompetes as a hindrance to innovation and economic growth).

33.   Other consequences include the effect on contracting behavior and the structure of contracts in general. *See, e.g.*, Kurt Lavetti et al., Buying Loyalty: Theory and Evidence from Physicians 20-29 (Feb. 20, 2016) (unpublished manuscript), http://www.kurtlavetti.com/UIPNC_vf.pdf [https://perma.cc/5S2V-5H4T] (studying the role that noncompetes play in the contracting structure and contracting behavior of physicians).

34.   Eric A. Posner & George G. Triantis, *Covenants Not to Compete from an Incomplete Contracts Perspective* 7-10 (John M. Olin Program in Law & Econ., Working Paper No. 137, 2001), http://www.law.uchicago.edu/files/files/137.EAP_.covenants.pdf [https://perma.cc/G45C-39FB] (developing a model involving a series of potential opportunities for negotiation and renegotiation of noncompetes among employees, their employers, and potential employers).

A. Existing Research

Early American scholarly treatments of noncompetes and other restrictive covenants were based primarily on common law sources.[35] From their first use in England as early as the fifteenth century, covenants not to compete have been viewed with suspicion because of their inherently anticompetitive nature.[36] Nevertheless, by the mid-twentieth century, there was near-consensus among U.S. courts that noncompetes should be enforceable within reasonable limits.[37] The classic balancing test to determine whether the scope of a covenant not to compete puts it within reasonable bounds—sometimes referred to as the "rule of reason" test—has evolved little over time.[38] This three-part inquiry requires courts to examine the legitimate interests of the employer while also considering the impact that enforcement is likely to have on both the employee's interests and, notably, the public welfare.[39]

---

35.   Well-known early overviews of the use of employment-related noncompetes in the United States include Charles E. Carpenter, *Validity of Contracts Not to Compete*, 76 U. PA. L. REV. 244, 244-71 (1928), and Winslow Drummond, Note, *Severability of Covenants in Partial Restraint of Trade: A New Rule*, 5 DUKE B.J. 115, 115-22 (1956).

36.   For a description of the early history of noncompetes and their anticompetitive character, see Blake, *supra* note 7, at 627 (tracking the development and use of covenants not to compete in the wake of the breakdown of the guild system in Britain). For a more recent decision invoking anticompetitive concerns, see, for example, Thiesing v. Dentsply Int'l, Inc., 748 F. Supp. 2d 932, 947 (E.D. Wis. 2010) ("Restrictive covenants limit one's right to work and to earn a livelihood and are therefore 'looked upon with disfavor, cautiously considered, and carefully scrutinized.'" (citations omitted)).

37.   *See, e.g.*, Arthur Murray Dance Studios v. Witter, 105 N.E.2d 685, 691-93 (Ohio Ct. Com. Pl. 1952) (citing numerous examples of early law review articles and case law disquisitions on the "reasonableness" standard for noncompetes).

38.   *See* T. Leigh Anenson, *The Role of Equity in Employment Noncompetition Cases*, 42 AM. BUS. L.J. 1, 17 (2005) ("Notwithstanding . . . changing attitudes and conditions in the labor market, the English rule of reason remains the doctrinal scheme in a majority of states in this country."); Maureen B. Callahan, *Post-Employment Restraint Agreements: A Reassessment*, 52 U. CHI. L. REV. 703, 709 (1985) (noting that the English reasonableness rule stemming from *Mitchel v. Reynolds* "has survived virtually unchanged to the present day").

39.   *See, e.g.*, Reliable Fire Equip. Co. v. Arredondo, 965 N.E.2d 393, 396-97 (Ill. 2011). In this case, the court explains the reasonableness test as follows:

> A restrictive covenant, assuming it is ancillary to a valid employment relationship, is reasonable only if the covenant: (1) is no greater than is required for the protection of a legitimate business interest of the employer-promisee; (2) does not impose undue hardship on the employee-promisor, and (3) is not injurious to the public. . . . Further, the extent of

As the U.S. labor market has shifted away from manufacturing and toward what some refer to as the knowledge economy, restrictive covenants have come under redoubled scrutiny. Many commentators claim that noncompetes are unfair to employees,[40] perhaps because the changing nature of employment has made such provisions more common.[41] The most spirited criticism targets employers' tendencies to maximize the restrictive scope of any noncompete and to exploit their superior bargaining power in take-it-or-leave-it negotiations.[42] For instance, employee fairness advocates are extremely suspicious of noncompete clauses, contending that while such agreements may theoretically protect employer interests, it is only at the clear and demonstrable expense of employee freedom.[43]

In any event, the ideal balance between employee freedom of choice and employer business interests is a complex question, and in practice, courts still struggle to find what they hope is equipoise.[44]

---

the employer's legitimate business interest may be limited by type of activity, geographical area, and time.

Id. (citing comments to the RESTATEMENT (SECOND) OF CONTRACTS §§ 187 cmt. (b), 188(1) cmts. (a)-(d) (AM. LAW INST. 1981)).

40.   *See, e.g.*, Katherine V.W. Stone, *The New Psychological Contract: Implications for the Changing Workplace for Labor and Employment Law*, 48 UCLA L. REV. 519, 577-92 (2001) (discussing employer abuse of superior bargaining position in the noncompete context).

41.   While this assertion is subject to debate, it seems clear that the changing workplace—and the changing nature of knowledge itself—has affected how parties and courts view noncompetes. *See, e.g.*, Anenson, *supra* note 38, at 17-18 ("It is clear that businesses will continue to utilize contracts restraining competition in an effort to police faltering employee loyalty and to retain their competitive edge. The judicial trend also seems to be toward favoring employers in the protection of company-developed human capital and legitimate proprietary interests."); Griffin Toronjo Pivateau, *Preserving Human Capital: Using the Noncompete Agreement to Achieve Competitive Advantage*, 4 J. BUS. ENTREPRENEURSHIP & L. 319, 320-21 (discussing the greater need for noncompetition agreements in the face of drastic changes in information technology and the economic environment).

42.   Arnow-Richman, *supra* note 1, at 1214-15 (2001) (reviewing the substantive and procedural concerns intrinsic to drafting, negotiating, and enforcing employment noncompetes).

43.   *Id.*; *see also* Stone, *supra* note 40, at 581-82.

44.   *See, e.g.*, Michael J. Garrison & John T. Wendt, *The Evolving Law of Employee Noncompete Agreements: Recent Trends and an Alternative Policy Approach*, 45 AM. BUS. L.J. 107, 112 (2008) (finding that courts moved to be more "protective of the employee's interest in mobility" and that a trend of "heightened scrutiny of employee noncompete agreements reflects some of the fundamental changes taking place in the economy and in the workplace"); *see also* Stone, *supra* note 40, at 580 (discussing the inconsistencies with what is considered "reasonable," which "varies from state to state and from case to case").

Scholars, for their part, never cease complicating the picture by proposing refinements of the factors courts should consider and the processes courts should use in light of business realities.[45]

Early empirical research on noncompetes began in the 1990s with several articles assessing the litigation and enforcement of these contracts.[46] The last 15 years have seen many more empirical studies, almost all of which followed in the wake of Ronald Gilson's well-known article contrasting the use of noncompetes in the high-tech economies of Massachusetts's Route 128 area and California's Silicon Valley.[47] Gilson argued that the differences in noncompete enforcement regimes across the two states played a significant role in the more pronounced rise of Silicon Valley.[48] His explanation—which relied on the perception that high-tech workers in Silicon Valley were more mobile—was theoretical but intuitively attractive as a testable claim.[49] Accordingly, variation in enforcement intensity

---

45. *See generally* Norman D. Bishara, *Covenants Not to Compete in a Knowledge Economy: Balancing Innovation from Employee Mobility Against Legal Protection for Human Capital Investment*, 27 BERKELEY J. EMP. & LAB. L. 287 (2006) (advocating for learning from the law and economics, labor mobility, and employee rights approaches—including balancing benefits and costs of enforcement with protecting specific classes of employees—which would maximize positive knowledge spillovers associated with employee mobility). This tension is more visible when considering the broader business consequences of noncompetes and how they impact both business strategy and employee interests. *See generally* Bishara & Orozco, *supra* note 12 (discussing disputes between employers and employees over knowledge ownership). Ethical considerations further tweak this challenge. *See generally* Norman D. Bishara & Michelle Westermann-Behaylo, *The Law and Ethics of Restrictions on an Employee's Post-Employment Mobility*, 49 AM. BUS. L.J. 1 (2012).

46. *See, e.g.*, Helen LaVan, *A Logit Model to Predict the Enforceability of Noncompete Agreements*, 12 EMP. RESP. & RTS. J. 219, 226-34 (2000) (analyzing a random sample of litigated cases to discern whether certain factors can predict court decisions); Peter J. Whitmore, *A Statistical Analysis of Noncompetition Clauses in Employment Contracts*, 15 J. CORP. L. 483, 493-94 (1990) ("The study presented in this [a]rticle relies on the assumption that judicial decisions in noncompetition clause cases are a function of the differing combinations of facts . . . a statistical analysis of the relationships between the facts and outcomes of each case should enable lawyers to better understand and predict judicial enforcement of noncompetition clause cases.").

47. Gilson, *supra* note 26, at 578-80.

48. *Id.* at 578.

49. *See id.* This is evidenced by the fact that others have used Gilson's work as a foundation for further empirical investigation. For example, one 2006 study addressed Gilson's thesis by comparing mobility rates in high-tech industries in Silicon Valley to mobility rates in non-high-tech industries elsewhere in California. This study found that employee mobility elsewhere in the state was consistent with

across the states started drawing the attention of legal academics,[50] and empirically minded researchers began using more rigorous methods to evaluate the role of noncompetes.

   These scholars have examined a broad array of outcomes: employee mobility,[51] human capital investment and training,[52] the formation and growth of agglomeration economies,[53] innovation,[54] and new venture creation,[55] including spinouts.[56] These studies bring to bear a range of empirical strategies both to identify the various effects of noncompete enforceability and to measure noncompete enforceability itself. Yet each of these studies also relies on a set of shared underlying assumptions about the nature and importance of state enforcement measures and about the absolute and relative prevalence of noncompete usage across states, types of employees and positions, and categories of industry.[57]

---

rates of mobility in other states. Bruce Fallick et al., *Job-Hopping in Silicon Valley: Some Evidence Concerning the Microfoundations of a High-Technology Cluster*, 88 REV. ECON. & STAT. 472, 475-81 (2006).

   50.   *See, e.g.*, Glynn, *supra* note 27, at 1418-24 (discussing the effect of noncompetes in the context of intrastate choice-of-law and choice-of-forum clauses); *see also* Gillian Lester & Elizabeth Ryan, *Choice of Law and Employee Restrictive Covenants: An American Perspective*, 31 COMP. LAB. L. & POL'Y J. 389 (2010) (focusing on the choice-of-law issues surrounding noncompetes that cause courts to take different approaches).

   51.   *See, e.g.*, Marx et al., *supra* note 9, at 884 (finding that the passage of the Michigan Antitrust Reform Act in the 1980s, which permits reasonable noncompetes, negatively impacted the mobility of patent holders in Michigan).

   52.   *See* Paul H. Rubin & Peter Shedd, *Human Capital and Covenants Not to Compete*, 10 J. LEGAL STUD. 93 (1981). Although not empirical, this important piece of legal scholarship was the first to apply economist Gary Becker's general and specific human capital theory to noncompete agreements.

   53.   *See* Fallick et al., *supra* note 49, at 481 ("Our finding of a California effect on mobility lends support to Gilson's hypothesis that the unenforceability of noncompete agreements under California state law enhances mobility and agglomeration economies in IT clusters.").

   54.   *See* Garmaise, *supra* note 8 (finding lower research and development expenditures by publically traded firms in noncompete-enforcing states); *see also* Raffaele Conti, *Do Non-Competition Agreements Lead Firms to Pursue Risky R&D Projects?*, 35 STRATEGIC MGMT. J. 1230, 1232 (2014).

   55.   *See* Stuart & Sorenson, *supra* note 4, at 183 ("Enforceable non-compete and non-solicitation covenants also may indirectly hamper new venture creation by depressing the life chances of early-stage companies.").

   56.   *See, e.g.*, Evan P. Starr et al., *Screening Spinouts? How Noncompete Enforceability Affects the Creation, Growth, and Survival of New Firms* (U.S. Census Bureau Ctr. Econ. Studies, Working Paper No. 14-27, 2014) (analyzing the impact of noncompetes on the process of spinout formation).

   57.   *See infra* Part III (discussing the complexities of the measurement of state enforcement intensity).

This rise in the empirical analysis of noncompetes and their possible consequences, the underlying assumptions of this body of work, and the remaining gaps in our understanding have already been carefully assessed by noncompete scholars.[58] Nevertheless, a brief reprise of the substance of the critiques of this literature is important to understand this Article's premise: We simply have no reliable way of knowing when, how often, or even why noncompetes and other restrictive covenants are actually used.

To date, we have identified 24 significant empirical studies that consider the use of noncompetes in the United States.[59] Broadly speaking, only six of the studies contain data on whether an employee has signed a noncompete.[60] Four of those examine top U.S. public company executives (as these contracts are accessible),[61] one focuses on physicians,[62] and one concerns engineers.[63] Three other studies catalog noncompete usage across employers,[64] and two use

---

58. *See, e.g.*, Norman D. Bishara & Evan Starr, *The Incomplete Noncompete Picture*, 20 LEWIS & CLARK L. REV. 497, 514-30 (2016) (tallying methodological concerns in the empirical noncompete literature at length).

59. *Id.* at 500-01.

60. *Id.* at 514-17.

61. *See, e.g.*, Stewart J. Schwab & Randall S. Thomas, *An Empirical Analysis of CEO Employment Contracts: What Do Top Executives Bargain For?*, 63 WASH. & LEE L. REV. 231, 254-57 (2006) (examining employment contracts with CEOs of U.S. public companies); *see also* Bishara et al., *supra* note 1, at 9-10 (finding that top executives' employment contracts often contain noncompete agreements alongside other restrictive covenants); Garmaise, *supra* note 8 (analyzing the impact of noncompetition clauses on the mobility of U.S. public company executives).

62. *See* Lavetti et al., *supra* note 33 (examining how noncompetes can alleviate inefficiencies in skilled services firms, particularly focusing on physicians).

63. *See* Matt Marx, *The Firm Strikes Back: Non-Compete Agreements and the Mobility of Technical Professionals*, 76 AM. SOC. REV. 695, 697 (2011) (studying the effect of noncompetes on technical professionals based on personal interviews with a relatively modest sample of engineers).

64. *See* Steven N. Kaplan & Per Strömberg, *Financial Contracting Theory Meets the Real World: An Empirical Analysis of Venture Capital Contracts*, 70 REV. ECON. STUD. 281, 289 (2003) (studying the incidence of founder noncompete clauses across a range of startup companies); *see also* William P. Galle, Jr. & Clifford M. Koen, *Reducing Post-Termination Disputes: A National Survey of Contract Clauses Used in Employment Contracts*, 9 J. INDIVIDUAL EMP. RTS. 227, 239 (2000) (summarizing the usage of noncompete clauses in employment contracts); Soc'y for Hum. Resource Mgmt., Survey on Non-Compete Agreements (Apr. 23, 2007), *available at* https://www.shrm.org/research/surveyfindings/documents/does_20your_20organization_20require_20employees_20to_20sign_20non-compete_20agreements_.ppt (providing data on the use of noncompetes for less than 400 human resource professionals).

experimental methods to examine noncompetes.[65] The remaining studies involve the effects of noncompete enforceability, but notably, none generates or utilizes information on who is actually *bound* by noncompete agreements.[66] In our review below, we do not address the specific findings of these articles. Rather, we highlight concerns about how much we can truly learn from these studies—regardless of the specific outcome of interest—given their limitations and the centrality of their perhaps invalid assumptions.

The empirical articles that address individual professions are useful in that they study people who have signed noncompete agreements, including some high-impact management employees.[67] This body of research gives insight into the noncompete contracting behavior of employers and particular employee groups. However, the studies are discrete, providing information on merely a sliver of the U.S. labor force. Even if these employees are representative of their professions, they are drawn from a small sample of jurisdictions and ultimately only account for a very small proportion of the U.S. workforce—approximately 0.87%.[68] Generalizing about the effects of noncompetes from these discrete studies is therefore highly unreliable, particularly when the studies are cited as justifications for reforms that will affect all employees.[69]

---

65.   *See* Bishara & Starr, *supra* note 58, at 519-22 (discussing a behavioral study intended to monitor the motivation and performance of participants under conditions mimicking a noncompete constraint); *see also* Amir & Lobel, *supra* note 12, at 847 (describing the same behavioral study).

66.   *See* Bishara & Starr, *supra* note 58, at 533-35 (raising concerns that these studies do not account for the rate at which employees actually sign noncompetition agreements).

67.   *See, e.g.*, Bishara et al., *supra* note 1, at 3 (studying many executive employment contracts and finding that 80% of the CEOs in the sample with explicit contracts had signed a covenant not to compete).

68.   *See* Bishara & Starr, *supra* note 58, at 500-01.

69.   For example, consider one study of the impact of noncompete policy on the mobility of Michigan "knowledge workers" who file patents. Matt Marx et al., *Regional Disadvantage? Employee Non-Compete Agreements and Brain Drain*, 44 RES. POL'Y 394, 394-404 (2015). This study cleverly examines the movement of patent-filing inventors, but does not necessarily tell us anything about the mobility of the vast majority of other types of employees—for instance, innovators for whom patent filing is not particularly relevant to performance or output. Regardless, these interesting but limited results—derived exclusively from a Michigan legal change in the mid-1980s—can be overemphasized and taken out of context. Media reports, for example, may make sweeping statements about *all* types of employees when facing a study of this sort. *See* Bloomberg, *Laws on Noncompete Agreements Hurt Michigan, New Study Says*, CRAIN'S DETROIT BUS. (Mar. 18, 2015, 5:12 AM), http://www.crainsdetroit.com/article/20150318/NEWS01/150319843/laws-on-

The three articles that discuss the use of noncompetes among employers are limited by both the lack of discussion of within-firm variation in the use of noncompetes and the lack of analysis linking the use of noncompetes to firm-level outcomes.[70]

The experimental studies are helpful contributions to our understanding because the use of noncompete agreements is likely to be nonrandom; with observational data, the nonrepresentative set of individuals who are asked to sign noncompetes can make it difficult to identify the true effects of such agreements.[71] Aside from potential criticisms of the methodologies of these studies and their use of small samples of participants in laboratory situations,[72] universalizing the findings of these articles to all types of employees across a wide range of demographics such as age, gender, experience, industry, education, annual earnings, and geographic region is questionable, to say the least.[73] Furthermore, these laboratory studies randomly assign noncompete status—essentially forcing participants to sign—and hence do not allow any type of negotiation or variation in incentives or information to be built into the contracting environment.

The batch of scholarship that relies on variation in noncompete enforceability across or within jurisdictions (without data on who is bound) constitutes the bulk of the noncompete literature and suffers from numerous shortcomings related to both measuring the law and

---

noncompete-agreements-hurt-michigan-new-study-says [https://perma.cc/5YMH-NCQX] ("It's a homegrown brain drain, the researchers say, stripping some regions of local talent while enriching others.").

70. *See* Bishara & Starr, *supra* note 58, at 519.

71. *See, e.g.*, Amir & Lobel, *supra* note 12, at 847-52.

72. *See* Bishara & Starr, *supra* note 58, at 519-22.

73. For a discussion of how situational effects of small scale economics laboratory research raise external validity and generalizability issues, see Francesco Guala, *Methodological Issues in Experimental Design and Interpretation*, *in* The Oxford Handbook of Philosophy of Economics 280, 298-99 (Harold Kincaid & Don Ross eds., 2009) ("The 'external validity' of experiments is one of the most sensitive and controversial issues in experimental economics," and "[t]he common perception for a long time has been that the very possibility of the experimental approach in economics was at stake in this controversy, and, perhaps, for this reason both experimenters and their critics have been seeking universal a priori answers to the question of generalizability."); *see also* Linda J. Luecken & Rika Tanaka, *Health Psychology*, *in* Research Methods in Psychology 245, 256 (Irving B. Weiner ed., 2013) (discussing highly controlled studies and controlling for variation and causal inferences in psychological studies); *id.* ("The careful selection of participants and controlled laboratory settings allows researchers to minimize other potential sources of variation in physiological reactivity. Nevertheless, the question remains as to whether findings from controlled laboratory studies generalize to responses in 'real life' and are representative of the larger population.").

the lack of information on who signs noncompetes.[74] Unfortunately, whether a state allows for some minimal enforcement or instead prohibits all enforcement ends up having little import, and only very few states completely ban noncompetes.[75] Most states provide some level of moderate enforcement—with a few outliers on each extreme of either aggressive enforcement or a complete enforcement ban.[76] When researchers opt to rely on an outmoded and inaccurate binary legal enforcement variable, they are, in effect, incorporating into their empirical analysis demonstrably false assumptions about state legal environments.[77] Other studies rely on the summary findings drawn from practitioner treatises, but these sources may not be updated regularly and may lack a uniform approach across states.[78] Embedded in these assumptions is a reliance on the false notion that noncompete law does not evolve.[79] Legal scholars do sometimes recognize that such reliance is misplaced,[80] but researchers in other disciplines may treat "noncompete law" as unchanging.[81]

---

74.   See generally the detailed discussion in Bishara & Starr, *supra* note 58, at 522-33, regarding the limitations of studies of noncompete enforceability. See also *infra* Part III's examination of enforcement intensity and its measurement in existing scholarship.

75.   In contrast, for a comprehensive view of the relative strength of noncompete enforcement policies in all 51 U.S. jurisdictions, see generally Norman D. Bishara, *Fifty Ways to Leave Your Employer: Relative Enforcement of Covenants Not to Compete, Trends, and Implications for Employee Mobility Policy*, 13 U. PA. J. BUS. L. 751, 772-80 (2011).

76.   *Id.* at 780 ("[T]he majority of states have followed a moderate course that seems to comport with traditional noncompete aesthetics of moderation through narrowly tailored and balanced—and reasonable—protectable interests that foster business investments in workers' human capital."); *see also infra* Figure 3 (demonstrating, in one coding approach, the smooth trend in enforcement intensity across states).

77.   *See* Bishara, *supra* note 75, at 762 (critiquing the use of a "'yes' or 'no'" enforcement variable in some empirical studies, which treat the enforcement or non-enforcement of noncompetes as absolute); *see also* Jonathan Barnett & Ted M. Sichelman, *Revisiting Labor Mobility in Innovation Markets* (SSRN Working Paper), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2758854 (Apr. 4, 2016) (criticizing the underlying assumptions of the use of an enforceability variable in several studies).

78.   For a critique of empirical investigations relying solely on the Malsberger practitioner treatise for a simplified noncompete enforcement variable, see Bishara & Starr, *supra* note 58, at 522-24.

79.   *See, e.g.*, Stuart & Sorenson, *supra* note 4, at 190 (using the Malsberger treatise to categorize states as simply either noncompete enforcing or non-enforcing jurisdictions).

80.   *See, e.g.*, Grant R. Garber, *Noncompete Clauses: Employee Mobility, Innovation Ecosystems, and Multinational R&D Offshoring*, 28 BERKELEY TECH.

Leaving measurement issues aside, without data on who signs noncompetes, studies that rely on noncompete enforceability alone must necessarily be viewed with caution, especially when the authors claim that the effects of enforceability are the same as the effects of noncompete agreements themselves.[82] Differences between enforcing and non-enforcing states may be driven by some unobserved and therefore omitted state-level factor, and without being able to show that the observed differences are due to those who sign noncompetes, such studies can do very little to allay serious identification concerns. Noncompete enforceability and the actual use of noncompetes are two distinct concepts, and finding that labor market outcomes differ in states that have greater noncompete enforceability may imply little about the effects of noncompetes themselves. Noncompetes may have chilling effects on employee mobility even in non-enforcing states if individuals believe their noncompete may be enforceable. This perspective underlines the fact that when enforceability is the key variable, *awareness* of the enforceability level itself is equally crucial to understand. Most studies assume perfect knowledge of noncompete laws by both firms and employees, but this assumption is highly doubtful, especially for employees.[83]

Some commentators also place their trust in distillations of reported litigation as surrogate measures for both the prevalence of noncompetes and growth trends in their use.[84] Unfortunately, legal

---

L.J. 1079, 1095-101 (2013) (explaining the history of noncompete law and the variation in enforcement levels); Garrison & Wendt, *supra* note 44, at 111 (tracking the evolution of noncompete law and state struggles to find workable policies).

81.   *See, e.g.*, Stuart & Sorenson, *supra* note 4, at 182.

82.   *See, e.g.*, Garmaise, *supra* note 8, at 1 ("We study the effects of noncompetition agreements by analyzing time-series and cross-sectional variation in the enforceability of these contracts across US states.")

83.   *See, e.g.*, Pauline Kim, *Bargaining with Imperfect Information: A Study of Worker Perceptions of Legal Protection in an At-Will World*, 83 CORNELL L. REV. 105, 106 (1997) ("While many commentators have questioned the assumption that workers are legally informed, the debate has been characterized by a remarkable dearth of empirical information. This study, by directly testing workers' knowledge of the relevant legal rules, offers empirical evidence contradicting the assumption of full information commonly made by defenders of the at-will rule. Far from understanding 'the footing on which they have contracted,' workers appear to systematically overestimate the protections afforded by law, believing that they have far greater rights against unjust or arbitrary discharges than they in fact have under an at-will contract.").

84.   *See, e.g.*, Greenhouse, *supra* note 14; *see also* Bob Sullivan, *Are Noncompete Clauses Getting out of Control?*, CNBC (June 25, 2014, 9:35 AM), http://www.cnbc.com/2014/06/25/are-noncompete-clauses-getting-out-of-control.html [https://perma.cc/M9XP-NMMV].

database searches only reveal *reported* cases—in other words, the relative handful of legal disputes that are not settled before trial, are submitted for judicial review, and are legally interesting enough to be reported within a given jurisdiction's judicial opinion reporting system.[85] Many issues arise from the use of these case aggregations.[86] If nothing else, they inaccurately suggest the existence of relatively little noncompete activity—the very distorted tip of the iceberg. Moreover, a change in a state's legal framework may alter employee and employer behavior in unanticipated or offsetting ways that are unobservable using a reported case yardstick. For example, a strong noncompete enforcement policy might lead to more noncompete agreements but might also produce a chilling effect (or so-called *in terrorem* effect) in which fewer employees behave in ways likely to generate litigation, which might ultimately lead to *fewer* reported opinions.[87] Methodologies that rely on reported cases might conclude that noncompetes have become less important in these circumstances when exactly the opposite conclusion would be true.

Collectively, the empirical studies to date are surely important contributions to our understanding of how restrictive covenants are enforced and how they impact certain, specific groups of labor force participants. But in the end, by omitting information on who signs noncompetes and by studying only a tiny proportion of the U.S. labor force, they tell us little about how noncompetes themselves operate, and collectively they fail to provide us with any findings that can be generalized to the U.S. labor force as a whole. Obvious gaps in our understanding include worker perceptions, the dynamics of *ex ante* and *ex post* negotiation, and the prevalence and content (across and within jurisdictions and industries) of noncompetes.[88]

---

85.   For an example of research that relies on case database searches to measure the prevalence of noncompete agreements, see Whitmore, *supra* note 46, at 526. Other legal scholars have used reported noncompete cases as informative data points in explaining the relevance of noncompetition agreements in the U.S. in the last few decades. *See* Gillian Lester, *Restrictive Covenants, Employee Training, and the Limits of Transaction-Cost Analysis*, 76 IND. L.J. 49, 54-59 (2001); *see also* Stone, *supra* note 40, at 577.

86.   *See* Lee Epstein & Gary King, *The Rules of Inference*, 69 U. CHI. L. REV. 1, 52-54 (2002); *see also* George L. Priest & Benjamin Klein, *The Selection of Disputes for Litigation*, 13 J. LEGAL STUD. 1, 5 (1984).

87.   *See e.g.*, Bishara et al., *supra* note 1, at 38-40 (finding that executive contract negotiations often yield reasonable and enforceable noncompete agreements that comply with underlying noncompete enforcement principles and state law).

88.   *See* Bishara & Starr, *supra* note 58, at 530-36 (highlighting the fact that these studies' assumptions about enforceability as an option may have some impact on the *ex ante* and *ex post* behavior of the parties).

Nevertheless, scholars and policymakers alike continue to draw conclusions about the implications of this body of research, often using it—or some part of it—as a basis for developing and making policy recommendations. For example, one well-known stream of scholarship takes the tidy position that noncompetes and other competitive restrictions are fundamentally harmful to a wide range of interests, including innovation and employee mobility.[89] Orly Lobel argues that the innovation-stunting influence of noncompetes is so great that states ought not to allow them at all.[90] Viva Moffat maintains that noncompete agreements are the "wrong tool" for protecting knowledge assets and advocates for a uniform ban across the United States.[91] Others have performed targeted noncompete data collection—in specific regions and industries—to more deeply explore the arguments for and against noncompetes. For instance, Robert Gomulkiewicz has demonstrated that fears over noncompetes limiting high-tech worker mobility may be exaggerated.[92] Although other scholars have emphasized the frequency of noncompetes in high-tech industries, the effects of noncompetes they claim are perhaps more pronounced than the actual incidence of enforcement by these employers would suggest.[93]

B. Burgeoning Policy Responses

Coincident with the expanding and diverse body of academic research on noncompetes, policymakers have recently shown intense interest in the associated legal reform questions.[94] State legislatures have begun to consider taking real action in response to what many

---

89.   S*ee generally* LOBEL, *supra* note 32 (critiquing all use of restrictive covenants, including noncompetes).
90.   *Id.*
91.   Viva R. Moffat, *Making Non-Competes Unenforceable*, 54 ARIZ. L. REV. 939, 984 (2012) ("Rather than adding another proposal for incremental change, this Article advocates for wholesale reform: Employee non-competes should be unenforceable, and the states should adopt a rule to that effect."); Viva R. Moffat, *The Wrong Tool for the Job: The IP Problem with Noncompetition Agreements*, 52 WM. & MARY L. REV. 873, 874 (2010) (proposing a uniform approach—prohibition—across all states).
92.   Gomulkiewicz, *supra* note 25, at 280-86 (providing an extensive list of reasons why high-tech employers choose not to enforce noncompetes—including the financial costs of litigation, trade secret disclosure risk, and unwanted publicity—even when enforcement is available under state law).
93.   *Id.*
94.   *See, e.g.*, *Noncompete Agreements: Follow Calif. Lead and Scrap Them*, *supra* note 24; Greenhouse, *supra* note 14.

perceive to be a pervasive "noncompete problem."[95] These legislative responses are primarily linked to concerns over employee equity. They are secondarily rooted in broad assumptions about one's home state's competitive advantage over other jurisdictions in terms of innovation and business development.[96]

Existing—and longstanding—noncompete limitations include exemptions for certain categories of employees. Although these professional carve-outs may be the product of targeted lobbying from self-interested industries and professions, some of them are arguably grounded in sound public policy arguments. By supposedly allowing unfettered access to important types of professionals in these fields, these exemptions may help protect public health and welfare. For example, a few states like Idaho and Nevada prohibit noncompetes in the employment contracts of foreign doctors working under a J-1 visa.[97] This limitation is ostensibly intended to attract and protect these doctors from abusive employers and to preserve access to health care in underserved urban and rural communities.[98] Likewise, other noncompete enforcement exceptions are justified as essential to preserving access to key professionals, such as physicians,[99] other

---

95.   Noncompete law remains, to date, largely stable. The longstanding "reasonableness" evaluation of noncompetes by the courts continues to be the standard in a large majority of U.S. jurisdictions; most states therefore do allow some form of noncompete enforcement. *See* Bishara, *supra* note 75, at 754. Nonetheless, the level of enforcement, the details of how such contracts are enforced, and under what circumstances they should be enforced have long varied significantly across jurisdictions. At one extreme, California has a very strong aversion to the enforcement of noncompetes and has frequently reiterated strong policy grounds in favor of employee mobility. *See* Edwards v. Arthur Andersen LLP, 189 P.3d 285, 288, 290 (Cal. 2008) (restating California's strong public policy in favor of employee mobility and against enforcing noncompetes, which create restrictions on that mobility). At the other extreme, states such as Florida tend to prioritize employers' business interests over employees' mobility, offering strong enforcement. *See* Bishara, *supra* note 75, at 778.

96.   *See* Gomulkiewicz, *supra* note 25, at 255 ("No state governor or legislature wants to be accused of losing its innovative edge by failing to update its non-compete laws. Several states are considering legislation to ban non-competes and others are being urged to do so.").

97.   *See* IDAHO CODE ANN. § 39-6109 (West 2015); NEV. REV. STAT. § 439A.175 (2015).

98.   *See* Marshall Allen, *Indentured Doctors*, L.V. SUN (Sept. 30, 2007, 6:00 A.M.), http://lasvegassun.com/news/2007/sep/30/indentured-doctors-how-it-happens -j-1-doctors-must/ [https://perma.cc/4NT6-XMHU] ("Many states forbid noncompete clauses for J-1 doctors because they would defeat the long-term purpose of keeping J-1 doctors in underserved areas after they leave the program.").

99.   *See, e.g.*, COLO. REV. STAT. ANN. § 8-2-113(3) (2015).

health and welfare workers,[100] and attorneys.[101] In a different vein—and in an apparent effort to reduce the perceived negative impact of noncompete enforcement on lower-wage employees with limited bargaining power—Colorado has long been more permissive of noncompetes for executive and management personnel.[102] Other exceptions to the enforcement of noncompetition agreements are, arguably, less high-minded. They seem to be more a function of specific industries or discrete professional groups seeking to defend their post-termination employment options. For example, broadcast professionals in Oregon may avoid noncompete enforcement under certain circumstances,[103] and this same group of employees in New York[104] and Massachusetts[105] has been more broadly exempted from noncompete enforcement. In another example, used car salesmen in Louisiana have a specific exemption.[106]

Against this background, many states have been revisiting their general willingness to enforce restrictive covenants.[107] For example, Massachusetts's legislature and its last two governors have engaged in a vigorous debate over proposed statutory changes to curtail noncompete enforcement.[108] This debate is nourished by speculation

---

100.   Over the years, Massachusetts has exempted several types of health and social welfare workers from noncompete enforcement, including physicians, nurses, and social workers. *See* MASS. GEN. LAWS ch. 112, §§ 12X, 74D, 135C (2016).

101.   Noncompetes are already prohibited for lawyers by state bar ethics rules and the common law to ensure that clients have access to the representation of their choice. *See, e.g.*, MICH. PROF'L CONDUCT R. 5.6 (stating that a lawyer cannot participate in an employment agreement restricting the right of the lawyer to practice after leaving a firm).

102.   *See* COLO. REV. STAT. ANN. § 8-2-113(2)(d) (providing multiple exceptions to a general prohibition on noncompetes, but only blanket exceptions for executive and management employees and their professional staff).

103.   *See* OR. REV. STAT. ANN. § 653.295 (West 2016) (stating that those employed as "on-air talent" may void noncompetes); *see also* Melissa Rassas, *Explaining the Outlier: Oregon's New Non-Compete Agreement Law & the Broadcasting Industry*, 11 U. PA. J. BUS. L. 447, 448 (2009) (providing an overview of the broadcasting industry noncompete statute in Oregon and discussing other similar statutes).

104.   N.Y. LAB. LAW § 202-k (McKinney 2016) (establishing the Broadcast Employees Freedom to Work Act that is designed for the "[p]rotection of persons employed in the broadcast industry").

105.   MASS. GEN. LAWS ch. 149, § 186 (2016) (exempting those holding certain broadcasting positions from noncompete enforceability).

106.   LA. REV. STAT. ANN. § 23:921 (2015).

107.   *See, e.g.*, *Noncompete Agreements: Follow Calif. Lead and Scrap Them*, *supra* note 24.

108.   *See* Callum Borchers & Michael B. Farrell, *Patrick Looks to Eliminate Tech Noncompete Agreements*, BOS. GLOBE (Apr. 10, 2014), https://www.bostonglobe.

in media outlets—and by the academic studies of discrete groups of employees we discuss above[109]—that eliminating noncompetes will aid the state in attracting and retaining high-tech employees that otherwise may gravitate toward Silicon Valley.[110] Legislators in other states, including Michigan and Illinois,[111] have also proposed reforms to existing restrictive covenant laws or have suggested prohibition altogether. Another legislative approach has been to single out noncompete enforcement with respect to specific innovative industries perceived to be especially harmed by restraints on employee mobility. In 2015, Hawaii enacted a statute essentially eliminating noncompetes for high-tech workers.[112] Other reforms have been related to process. Oregon's 2007 amendments provided for a two-week notice requirement under which an employer must

---

com/business/2014/04/09/gov-patrick-pushes-ban-noncompete-agreements-employment-contracts/kgOq3rkbtQkhYooVIicfOO/story.html (concluding that the proposed ban on noncompete clauses was probably the "most controversial element of legislation" proposed by former Governor Patrick in 2014); *see also* Rebecca Strong, *Post-Deval, the Bid to Ban Noncompetes in Mass. Makes a Comeback*, BOSTINNO (Apr. 10, 2015, 11:43 AM), http://bostinno.streetwise.co/ 2015/04/10/boston-tech-vc-bill-filed-to-ban-noncompetes-in-mass/ [https://perma.cc/ LV39-TRYY] (describing how many previous efforts in Massachusetts to reform noncompete enforcement policy have failed, but also noting how state representatives and state senators continue to propose new legislation on the issue).

109.   *See supra* Section I.A.

110.   *See* Editorial, *Instagram Deal Highlights Need to Bolster Tech Startups in Mass.*, BOS. GLOBE (Apr. 13, 2012), https://www.bostonglobe.com/opinion/editorials/2012/04/12/instagram-deal-highlights-need-bolster-tech-startups-mass/1kx3VvY5DGdLLX0GBl6sQJ/story.html (arguing for Massachusetts reform of business laws and noncompete enforcement policy on the basis of "how much the Boston area can benefit from fostering a healthy climate for future Facebooks and Instagrams" and because "Mark Zuckerberg has argued that he moved his fledgling company from Massachusetts to California because of the greater availability of investment funding").

111.   *See* H.B. 4198, 89th Leg., Reg. Sess. (Mich. 2015) (preserving noncompetes related to the sale of goodwill of a business but making "any term in an agreement an employer obtains from an employee, contract laborer, or other individual that prohibits or limits the individual from engaging in employment" void); *see also* H.B. 0016, 97th Gen. Assemb., Reg. Sess. (Ill. 2011); Peter A. Steinmeyer, *Drafting Enforceable Noncompetition Agreements in Illinois*, CORP. LAW. (Ill. State Bar Assoc.), May 2010, at 3 (describing the proposed Illinois law, first introduced in 2010, which would ban noncompetes except for a "key employee" or "key independent contractor").

112.   *See, e.g.*, H.B. 1090, 28th Leg., Reg. Sess. (Haw. 2015); Claire Zillman, *Hawaii Ban on Noncompetes Leaves Out a Huge Chunk of Workers*, FORTUNE (July 8, 2015, 5:07 PM), http://fortune.com/2015/07/08/hawaii-noncompete-ban [https://perma.cc/RUU6-7XSA] (discussing the limitations of the Hawaii statute, which predominantly focuses on the technology industry).

request the restriction in advance of employment.[113] More recently, Oregon has reduced its per se limit on the period of enforceability from two years to 18 months.[114]

In just the last year or two, federal policymakers have also taken up the mantle of reforming noncompete enforcement policy.[115] This includes, apparently for the first time, congressional action on the matter.[116] The news of seemingly oppressive noncompetes for low-wage employees[117]—and recent litigation concerning the impact of noncompetes on those employees[118]—has brought noncompete issues to the attention of several U.S. Senators,[119] notwithstanding the fact that noncompete enforcement policy has always been considered a state-law domain.[120] This congressional limelight led to the proposed 2015 Mobility and Opportunity for Vulnerable Employees (MOVE) Act, which sought to ban covenants not to compete for workers earning less than $15 an hour.[121] In justifying the bill, sponsors contend that employers are forcing low-wage employees "to sign non-compete agreements in an effort to dissuade those workers from seeking better, higher-paying jobs within the same industry" and that "[t]his unfair use of non-compete agreements has a chilling effect on the upward economic mobility of low-wage workers and stifles their ability to climb out of poverty."[122] They

---

113.   *See* Rassas, *supra* note 103, at 460 (discussing the final provisions of the much-debated 2007 Oregon noncompete reform legislation).

114.   *See* OR. REV. STAT. ANN. § 653.295 (West 2016). Other states enforce noncompetes as long as they have per se "reasonable" limits under the law. *See* H.B. 352 §1(b)(4) (Ala. 2015) (considering employee noncompetes with durations of two years or less presumptively reasonable and placing the burden of showing undue hardship on the opposing party).

115.   *See generally* Treasury Report, *supra* note 20; White House Report, *supra* note 21.

116.   *See, e.g.*, Press Release, Sen. Chris Murphy, Murphy, Franken Introduce Bill to Ban Non-Compete Agreements for Low-Wage Workers (June 3, 2015),   http://www.murphy.senate.gov/newsroom/press-releases/murphy-franken-introduce-bill-to-ban-non-compete-agreements-for-low-wage-workers [https://perma.cc/KQ9F-Y8YY].

117.   *See, e.g.*, Greenhouse, *supra* note 14.

118.   *See, e.g.*, Brunner v. Liautaud, No. 14-c-5509, 2015 WL 1598106, at *1-2 (N.D. Ill. Apr. 8, 2015) (exemplifying a high-profile case in which a current and a former employee sued employer Jimmy John's seeking a declaration of the unenforceability of their noncompete agreements).

119.   *See* Press Release, Sen. Chris Murphy, *supra* note 116.

120.   *See* Bishara, *supra* note 75, at 768-71.

121.   *See* Mobility and Opportunity for Vulnerable Employees (MOVE) Act, S. 1504, 114th Cong. (1st Sess. 2015).

122.   Press Release, Sen. Chris Murphy, *supra* note 116.

conclude that the solution to this perceived problem is "outlawing the use of non-compete agreements for low-wage workers . . . [to] allow those currently stuck in their low-wage jobs to secure a better life for themselves and their families."[123]

C. The Need for Employee-Level Data

Better research requires better data. Moreover, as observers of the state-level debate over what, if anything, to do with noncompete enforcement policy, we have often been bemused and occasionally concerned—but also not surprised—to learn that policymakers and commentators sometimes unduly emphasize particular elements of existing research while ignoring many of the literature's critical uncertainties and caveats. Accordingly, we began to work together to diagnose the gaps in the data and in our understanding and to identify the suspect (and especially the implicit) assumptions in the existing scholarship and in the most commonly deployed policy arguments. These discussions proved to be early precursors to what became a data-gathering project.

While each of the justifications for reforming noncompete enforcement policy or banning noncompetes may be correct to some degree,[124] available data and analysis do not conclusively support many of the broad-sweeping claims reform advocates have put forth.[125] The possibility that noncompete policy innovation may actually result in negative consequences for employees—for example, for job access, for knowledge sharing and training, and for job tenure—is too often ignored. Furthermore, the assumption that noncompetes are pervasive among vulnerable, low-wage employees is based only on anecdotes, despite their being embedded in high-profile media reports.[126] In truth, we simply know very little about the use and dispersion of these contracts.

In addition, many policymakers neglect to engage seriously with the complicated and potentially offsetting effects of these restrictions. For example, even if noncompetes reduce mobility, if they also significantly enhance wage growth, or simply increase job satisfaction, coming to a verdict on noncompetes requires making a

---

123.   *Id.*
124.   *See supra* Sections I.A & I.B.
125.   For instance, these policy reforms—especially the strong calls for outright bans—may implicitly assume that noncompetes have a negative impact on employee wages, job tenure, and career mobility both internally and externally.
126.   *See, e.g.*, Jamieson, *supra* note 6; Woodman, *supra* note 6.

value tradeoff.[127] Some recommendations depend on empirically contestable (and untested) assumptions about how the law works and how employees are likely to behave when bound by noncompetes.[128] If an employee does not know whether he is a party to a noncompete (or believes that the noncompete is unenforceable), any restriction may be orthogonal to his thinking and behavior. In the end, reliance on partial data creates unsupported, overly focused, and perhaps even inaccurate arguments (no matter how well-known and cited), which may result in unwise policy choices.

In late 2013 and early 2014, we crafted a survey of employees with the aim of erecting a firmer footing for the noncompete policy debate. An extensive survey about labor market behaviors, beliefs, and outcomes generally as well as about noncompete contracting experiences was necessary to understand how, when, and why (and to what effect) employers and employees enter into these contracts. The result was the 2014 Noncompete Survey, completed by more than 11,500 labor force participants from a range of industries and with varied demographics, experiences, earnings, and expectations.[129] By examining employee beliefs and behaviors regarding noncompete negotiation, employee comprehension of the legal and practical implications of noncompetes, and a wide range of worker mobility topics, we set the stage for evaluating many assumptions embedded in noncompete scholarship and policy discourse.

---

127.   *See* Samila & Sorenson, *supra* note 10, at 425 ("But noncompete covenants also have a positive side, helping companies to protect the human capital, intellectual property, and relationships they have developed. Companies can increase their productivity by training workers, developing new products and processes, and building relationships with customers and suppliers.").

128.   Many assume that there is a noncompete chilling effect on employee behavior and that unscrupulous employers use that effect to their advantage—even when the underlying state law makes any overbroad contract unenforceable. *See* Press Release, Sen. Chris Murphy, *supra* note 116. Some even suggest that having employees sign noncompetition agreements creates a "market for lemons" that ultimately backfires for employers, keeping the least talented and least valuable workers at a firm instead of retaining the best. *See* Jeff Haden, *The Case Against Non-Compete Agreements*, INC. (Nov. 19, 2013), http://www.inc.com/jeff-haden/non-competes-could-cause-the-death-of-your-business.html (discussing the various possible negative effects of noncompetes with Orly Lobel). Both of these arguments rest on the belief that employees know and appreciate the restrictions they sign. If an employee does not recall whether he or she signed a noncompete, or does not understand its meaning (legitimate or not), then a chilling effect or a lemons problem is impossible.

129.   See *infra* Section II.A for our discussion of the survey questions and *infra* Part II generally for a discussion of the coverage and diversity of our sample of respondents.

II. SURVEY, DATA, AND METHODS

To collect the 2014 Noncompete Survey data, we employed a large-scale, online survey instrument designed primarily to gather information on employee experiences with and understanding of noncompetition agreements. In this Part of the Article, we offer a fairly detailed picture of the mechanics and methodology behind our survey data in hopes of convincing other scholars—particularly legal scholars—that the collection of individual-level data on very specific legal topics is not only possible, but a rewarding, constructive, and economical investment.[130] We also hope that our thorough discussion will provide a roadmap of key issues to consider, methods with which to experiment, and pitfalls to avoid for others weighing the relative merits of pursuing a similar survey project.

We begin below by outlining the substance of the survey. We characterize the sample population and the primary categories of questions, which include inquiries about the respondent's future expectations and plans, and the respondent's beliefs about employer behavior and the noncompete status of other employees.[131] We describe our online implementation of the survey. We also include a discussion of the costs and benefits of an online surveying approach (as opposed to using a paper or direct-dial telephone survey), and why we firmly believe that, at least in this particular context, online

---

130.   As evidence of the efficacy of modern online surveys, note that, for the 2016 Presidential Election, Reuters has partnered with an online survey company, Ipsos, to conduct *all* of its pre-election polling. Reuters explains that:

> Online surveys allow us to collect far more data and to be more flexible and fast-moving than phone research, and online is also where the future of polling lies.
>
>   This methodology may be different from the 'traditional' (telephone) approach used by others, but it is highly accurate: It was the most accurate national poll of U.S. residents published immediately before the November 2012 general election.
>
>   Our data is primarily drawn from online surveys using sampling methods developed in consultation with several outside experts. These involve recruiting respondents from the entire population of U.S.-based Internet users in addition to hundreds of thousands of individuals pre-screened by Ipsos. The responses are then weighted based on demographic information.
>
>   Because of the volume of demographic information collected, the poll provides unprecedented insight into the myriad of communities that constitute the United States in the 21st century.

About the Polling Explorer, REUTERS, http://polling.reuters.com/ (follow "About" hyperlink) (last visited Apr. 14, 2016).

131.   *See infra* Section II.A.

surveying is sensible.[132] We summarize our approach to cleaning the data and the limitations of both the survey and the data that we identified during the cleaning process.[133] We sketch our methods for refining the "clean" data to render it more reliable.[134] We address some of the sample selection concerns relevant to our surveying strategy.[135] We argue that, for many purposes, these sample selection concerns are (relatively) unimportant, suggest ways to assess the robustness of inferences to particular selection issues, and describe our reweighting techniques. Finally, we quickly relate our multiple imputation efforts,[136] and we wrap up by reporting a second round of reweighting that we employed to generate our final dataset.[137]

A. Survey Design

We conducted the survey over approximately three months—between April 22, 2014, and July 25, 2014. We sampled labor force participants aged 18 to 75,[138] who reported being in the private for-profit or nonprofit sector or being an employee of a public healthcare system.[139] We therefore excluded from the sample all self-employed individuals, government employees, and those who indicated that they were both unemployed and not looking for work.

The substance of the 2014 Noncompete Survey concentrated on three specific sets of issues related to noncompetition agreements: the respondent's personal lifetime experiences with noncompetes, the respondent's beliefs regarding the laws governing noncompetes and his or her perceptions of enforcement costs and probabilities of

---

132. *See infra* Section II.B.
133. *See infra* Section II.C.
134. *See infra* Section II.D.
135. *See infra* Section II.E.
136. *See infra* Section II.F.
137. *See infra* Section II.G.
138. We chose to survey individuals who were past the standard retirement age (ages 65–75) because these individuals are more likely to be asked to sign noncompetes as they tend to have more senior positions, all else equal. *See* David B. Ritter & Sonya Rosenberg, *The Ins and Outs of Non-Competes*, 39 COMPENSATION & BENEFITS REV. 40, 42 (2007).
139. As a point of comparison, the Current Population Survey (CPS) similarly seeks labor force participation data and studies the non-institutionalized U.S. population. Our sampling frame is similar, except that we exclude individuals who work in industries for which a noncompete appears unnecessary, such as government. The American Life Panel, run by RAND, is a similar online panel that collects data from individuals who responded to previous Census, Bureau of Labor Statistics, or Health and Retirement Surveys and wish to respond to more surveys.

enforcement, and the respondent's experience with and beliefs about noncompetes in the individual's current job.[140]

It is worth highlighting at this point three general types of questions we employed in the survey that ultimately proved useful in evaluating our respondents' answers for logical consistency and that should continue to bear quite useful and informative fruit in future analyses. First, we asked respondents to make predictions about their future, to describe "what it would take" for them to make certain changes (for example, to leave their current employer), and to record their expectations about the years ahead with respect to training, compensation, and so on. Second, we asked respondents to report their beliefs about how their employer would respond in hypothetical scenarios and their beliefs about the noncompete status, experiences, and likely behavior of third parties, such as co-workers, employees with similar positions at other employers, and other employees in the same industry. We exploit these "projection" and "belief" questions to better understand employees' perceptions of noncompete law and the use and effects of noncompetes beyond our respondents' personal experiences. Third, we asked respondents to answer retrospective questions about how noncompetes may have affected their prior employment relationships, including their current position. Such questions allow us to examine overall experiences with noncompetes and, in particular, how the existing distribution of occupations has been influenced by prior generations of these contracts.

In the rest of this Section, we briefly elaborate on our question categories primarily to introduce the survey and to provide relevant background for our methodological discussions below.[141]

---

140.   A hardcopy version of the survey instrument is available from the authors. We set down here the substance of many of the questions we asked, but for sake of brevity, and because we do not provide question language in the same order or context, we do not necessarily report the precise wording of questions.

141.   Although we do not discuss these questions below, we also collected a great deal of demographic information from respondents, some of it automatically. For example (and when applicable): In what state do you live? What is your age, your gender, and your marital status? Are you employed? What type of employer? What is your highest educational degree? Are you still in school? What degree are you working toward? How long ago did you finish school? Is your spouse or partner currently employed? How much does he or she earn? How many children or dependents live with you? Do you anticipate extraordinary expenses with respect to supporting your dependents in the next five years? As we explain in more detail below, we also asked respondents questions about their online survey experience and behavior. We asked questions such as: How many years have you taken online surveys? Why did you sign up to take online surveys? How often do you take online surveys? Why and where did you take this survey?

1. *Prior Employment and Experience with Noncompetes*

The survey focuses predominantly on the respondent's current employment relationship, but also asks questions that explore the respondent's familiarity and history with noncompetes, which may predate the respondent's current employment relationship.[142] We also sought facts that might help us determine whether certain aspects of prior employment relationships (e.g., previous noncompetes) relate to a respondent's current characteristics or status.[143]

2. *Current Employment Relationship*

In order to assess the potential causes and consequences of noncompetes, the survey requests extensive information about many dimensions of the respondent's current employment relationship.[144] Obviously, we collected basic terms of employment.[145] The survey also asks respondents about their employer—its industry, size, structure, and so on.[146] In addition, we inquired about respondents' history with their employer (e.g., other positions, training),[147] and we

---

142. E.g.: Have you heard of noncompetition agreements? Where or from whom did you hear of them? At what age? Have you ever signed one? Are you sure? Have you ever unknowingly signed one and discovered later that you had signed one? At what age did that happen? Have you ever been sued on the basis of a noncompete? Has a noncompete ever been a factor in your decision to stay with an employer? If so, how? If you left, did the noncompete affect how you departed and where you sought employment next? Did you negotiate the terms of your departure?

143. E.g.: In the last five years, how many employers have you had? How many of these relationships involved noncompetes?

144. The sample also includes respondents who are currently unemployed. The survey asks unemployed respondents questions about their unemployment. E.g.: Do you have a job? If not, are you unemployed as the result of a noncompete with a previous employer? If you are unemployed, where do you live?

145. E.g.: What is your occupation? What is your job title? What are your duties? Do you work with trade secrets, confidential information, or have access to clients? What is your monthly and annual compensation? We also asked questions such as the following: How are you compensated? What benefits do you receive? Are you represented by a union? Are you a full-time, part-time, or seasonal employee? How many hours do you work in a week, month, or year?

146. E.g.: What is your employer's industry? Does your employer operate in a different state? Where is your employer headquartered? How frequently does your employer hire someone away from a competitor? How often does someone leave your employer to join a competitor?

147. E.g.: Did a prior noncompete influence your choice of this job? How many years of experience do you have in this type of position? What is your tenure with your employer? When was your last raise or promotion? How much training have you had in the past year?

invited respondents to assess the level of their work performance and their contributions and loyalty to their employer.[148]

3. *Future Compensation, Promotions, and Opportunities*

If a noncompete influences the initial terms and conditions and the subsequent evolution of a particular employment relationship, we might observe such an effect not solely in the position's attributes today (e.g., current compensation), but in how those dimensions have evolved over time or are *expected* to change in the future (e.g., future compensation). For this reason, the survey asks respondents to speculate about or "project" their likely opportunities, decisions, and possible outcomes going forward.[149]

4. *Enforcement and Enforceability of Noncompetes*

Many commentators, as we note above, assume that employees are poorly informed about the content and likely practical and legal consequences of a noncompete clause. To explore this possibility, we ask respondents to answer a few basic questions about the law that governs noncompetes and about the likelihood of noncompetes being enforced under the conditions of various scenarios.[150]

5. *Scope and Substance of Noncompetes*

Our survey briefly inquires about prior noncompetes, but for those respondents *currently* bound by a noncompete, we asked many questions about the existing contract's content.[151] We focused on the

---

148.   E.g.: How would you rate your daily effort level, creativity, and overall job performance in the last month? How would you rate yourself relative to your co-workers in the last month? What motivates you to work hard?

149.   E.g.: If you stay with your current employer, what will your wages be in the future? How likely are a number of possible changes—good and bad—to occur in the future? In the next year, are you likely to get an offer from another employer? Will you accept such an offer and leave your employer? How much will you demand to leave? Will it depend on whether the offer comes from one of your employer's competitors or from an employer in another industry?

150.   E.g.: Are noncompetes enforced at the national, state, or local level of government? How likely is an employer to try to enforce a noncompete against an employee? Does it depend on the circumstances? Which ones? Are noncompetes enforceable by courts in your state? Are there any states that refuse to enforce noncompetes to your knowledge? Which states?

151.   E.g.: Do you understand the terms of your noncompete? Are the terms of the noncompete fair? How significantly does it limit your options? For how long

substance of the agreement's constraints, such as the length of time the respondent would be prohibited from working for a competitor and whether the contract involves a geographical restriction.

6. *Noncompete Negotiation and Process*

To better understand the bargaining and contracting process—especially in view of the significant concerns about the abuse of low-wage employees—the survey asks respondents who are currently bound by a noncompete to report when and how their employer asked them to sign it;[152] to explain the negotiating process, how they made their decision, and whether there was consideration offered in exchange for the noncompete clause;[153] and to assess an employer's decision to require or propose noncompetes.[154]

7. *Performance and Mobility Consequences of Noncompetes*

The survey prompts respondents with a noncompete to assess the effect of their noncompete on their work performance; it also asks them to gauge how their noncompete affects the possibility of their leaving their employer—i.e., whether respondents believe themselves to be mobile.[155] We explored employee mobility in depth by asking respondents questions, for example, about their on-the-job search efforts and their experiences—if any—of being recruited by

---

are you restricted from joining a competitor upon termination? Are you limited geographically? Is there a buy-out clause? How do you know the content of your noncompete for this survey? Did you re-read it when responding to this survey? How confident are you in your answers about your agreement's content? What other types of restrictive covenants have you signed?

152.   E.g.: When were you asked to sign it? Before or after you accepted your position? Would it have affected your choice to accept the position had you known about it earlier?

153.   E.g.: Did you have other employment opportunities when you agreed to your noncompete? Did you read your noncompete before you signed it? Did you consult a lawyer? Did you negotiate over the terms of your noncompete? If not, why not? If you negotiated the terms, what did you request in return for signing? Was your request successful? If you chose not to negotiate, what were your reasons? Did you ask what would happen if you refused to sign? If you had declined to sign, do you think your employer would still have hired or promoted you? Were you promised benefits if you signed? Did you receive those benefits?

154.   E.g.: Is it fair to ask an employee to sign a noncompete? Why do you think your employer asked you to sign a noncompete?

155.   E.g.: Does your noncompete, in your opinion, influence your work performance, effort, or motivation? Would your noncompete affect your willingness to leave? Would it affect your willingness to start a new business?

other employers.[156] Finally, the survey also solicits respondents who have never signed a noncompete agreement to speculate on how their behavior might change if they were bound by one.

8. *Beliefs About Noncompete Frequency, Content, and Effects*

With the aim of extracting additional data from our respondents (as well as testing the consistency of their answers),[157] the survey investigates not only respondents' personal noncompete experiences and their beliefs about their own behavior, but also their views about noncompete frequency generally (at the employer, occupation, and industry level), about what other employees' noncompete contracts typically contain, and also about how these employees are likely to behave in response to the strictures of a noncompete.[158]

B. Survey Implementation

The benefits of collecting individual-level information through an online survey are clear. Relative to random-digit dialing (RDD), mail-in, or in-person surveys, an online survey's cost per respondent is dramatically lower.[159] The data-collection process is also orders of

---

156.   E.g.: How much effort do you spend or have you spent looking for other employment? How have you searched? Has a headhunter ever approached you? Do you search for jobs with employers that compete with your current employer? Have you received offers from other employers? How many? Did you turn these offers down at least in part because of your noncompete? Was your employer aware of your outside offer(s)? If so, how did your employer respond? How much would you need to be offered to be willing to switch employers? Would the necessary amount depend on whether the offer came from one of your employer's competitors?

157.   *Cf.* David Rothschild & Justin Wolfers, Forecasting Elections: Voter Intentions Versus Expectations (Jan. 23, 2013) (unpublished paper) (on file with the University of Pennsylvania Wharton School), http://users.nber.org/~jwolfers/Papers/VoterExpectations.pdf [https://perma.cc/K526-J5CJ] (offering evidence that answers to the question "Who will win an election?" are better predictors of the actual election results than answers to the question "Who will you vote for?" and explaining that result by contending that the former question produces an answer that incorporates not only the respondent's likely voting preferences, but also the preferences of individuals in the respondent's network).

158.   E.g.: What percentage of your co-workers in your type of position (also, at your workplace and at all locations) have signed a noncompete? What percentage of employees in your type of position generally? What percentage of employees in your industry? Are the terms of your noncompete identical to what others in your position have? What percentage of others who are in your position and who have a noncompete would turn down an offer because of their noncompete?

159.   *See* ROGER TOURANGEAU ET AL., SCIENCE OF WEB SURVEYS 1-2 (2013).

magnitude faster per respondent.[160] Furthermore, an online platform allows surveyors to ask respondents complicated, nested questions in simple, easy-to-use, and interactive ways. Finally, companies such as Qualtrics that conduct online surveys are at least sometimes willing to guarantee a target sample size and thus will continue to sample the population until that sample size is reached.[161]

These practical benefits, however, come at a potentially high methodological cost: sample selection and other systematic data collection issues. These concerns may raise serious questions about the reliability of inferences drawn about the population of interest from an online survey of an online sample of these individuals. Not surprisingly, given the benefits of using an online platform to collect data, researchers have sought to test the accuracy of online surveying by comparing the findings of identical survey questions given both to online respondents and by phone using RDD. This research shows that, at least in some situations, the results can differ between the two surveying approaches.[162] Some are thus quick to dismiss data from online surveys—especially "open" surveys accessible to anyone on the web[163]—because of these observed differences, but it may also be that more traditional survey methods, such as RDD, are themselves unreliable, perhaps as a result of severe unit nonresponse.[164]

---

160.   *See* MARIO CALLEGARO ET AL., WEB SURVEY METHODOLOGY 20-21 (2015) (describing speed and other advantages of online surveying).

161.   At least in our case, in order to ensure the highest level of data quality, our surveying company was willing to replace, for free, any individual respondent's survey answers that we were able to demonstrate were partially the product of intentional noncompliance. *See infra* Section II.D.

162.   *See* Matthias Schonlau et al., *A Comparison Between Responses from a Propensity-Weighted Web Survey and an Identical RDD Survey*, 22 SOC. SCI. COMPUTER REV. 128, 136 (2004); *see also* Matthias Schonlau et al., *Selection Bias in Web Surveys and the Use of Propensity Scores* 12 (RAND Lab. & Population, Working Paper No. 279, 2006); Stephanie Steinmetz et al., *Comparing Different Weighting Procedures for Volunteer Web Surveys* 37-39 (Amsterdam Inst. for Advanced Lab. Studies, Working Paper No. 09-76, 2009). Differences in the mode of communication (online versus phone), differences in the length of the survey period, different survey administrators, and differences in the sample populations are several factors that may be behind these inconsistencies.

163.   *See* Jerry Vaske et al., *Can Weighting Compensate for Sampling Issues in Internet Surveys?*, 16 HUM. DIMENSIONS WILDLIFE 200, 212-13 (2011).

164.   For example, Andrew Kohut and his colleagues show that the response rate of a typical Pew Research telephone survey has fallen from 36% in 1997 to only 9% in 2012. *See* ANDREW KOHUT ET AL., PEW RESEARCH CTR., ASSESSING THE REPRESENTATIVENESS OF PUBLIC OPINION SURVEYS 1 (2012), http://www.people-press.org/files/legacy-pdf/Assessing%20the%20Representativeness%20of%20Public%20Opinion%20Surveys.pdf [https://perma.cc/XRM2-5R8T]. Such low response rates

In what remains of this Section, we discuss the nuts and bolts of our data-collection efforts. In the following Sections, we describe our data cleaning strategy, sample selection issues, and our methods for refining the data to make the conclusions we draw in Part III and will draw in future research more transparent and reliable.

*   *   *

To collect responses to the survey, we hired a survey and data collection firm,[165] which in turn outsourced the data collection to eight different panel partners.[166] Individuals contacted by these panel partners to take the survey had previously applied to at least one of the panel partners seeking the opportunity to receive requests to take online surveys. Each prospective respondent was sent the survey via e-mail or received the survey as part of an online game. Importantly, the survey was not available to any person who might have happened upon it on the internet, but only to those people whom we (through our agents) specifically sent an individualized link.[167]

Respondents were compensated for taking our survey, but the nature and magnitude of this compensation varied across panel partners. Respondents who attempted the survey but did not finish it

---

make one important methodological question the extent to which the final sample is truly representative of the population it is intended to represent. All of this is to say that the not-uncommon assumption that baseline population parameters can be estimated using data from RDD surveys may be seriously flawed.

165.   Specifically, see QUALTRICS, http://www.qualtrics.com [https://perma.cc/XR42-VC5S] (last visited Apr. 14, 2016).

166.   The eight panel partners that Qualtrics used as subcontractors for the data collection were ClearVoice, GMI, Sample Strategies, SSI, Innovate, Toluna, Precision Sample, and Samplify. Importantly, it is possible for a single individual respondent to have accounts with more than one of these companies. It is also possible for the same individual respondent to have multiple accounts with the same panel partner, but typically not without providing that partner with different identity information for each account. We discuss the potential pitfalls associated with these facts and how we addressed them in more detail below.

167.   The survey link is individualized so the panel partner can identify which individual respondent is taking the survey, allowing the partner to link the response to other information that the company has about that respondent. Each panel partner employs its own confirmation procedures to verify the demographic, contact, and other information of its respondents. These procedures include, but are not limited to, TrueSample, Verity, SmartSample, USPS verification, and digital fingerprinting. To probe the veracity of the claims made by the panel partners about their demographic verification systems, one of the authors signed up with SSI seeking to be a part of their online survey panel. After we submitted the required information, an SSI representative called us at a number we provided and asked to clarify some of the prospective respondent information we had submitted.

were paid only 10¢. By comparison, respondents who actually completed the survey either received $1.50, were entered into a sweepstakes drawing for various online rewards, or were awarded credits to play a particular online game.[168]

Our target size for this data-collection enterprise was 10,000 completed surveys. We were able to control the characteristics of the final sample through the use of quotas, which are simply constraints on the numbers of respondents with particular characteristics or sets of characteristics. In particular, we sought a final sample in which respondents were 50% male; 60% with at least a bachelor's degree; 50% with earnings of at least $50,000 annually from their current, highest paying job; and 30% over the age of 55. We chose these numbers either to align the sample with the corresponding sample moments for labor force participants in the 2012 American Community Survey (ACS)f[169] or to oversample certain groups of the population of particular interest for further subgroup comparisons.[170] In addition, to better understand smaller states with unusual or contested noncompete laws, our survey oversampled the residents of Colorado, Oregon, Massachusetts, and Florida.[171]

---

168.   For more insight into the methods used by Qualtrics, see Chris P. Long et al., *Fairness Monitoring: Linking Managerial Controls and Fairness Judgments in Organizations*, 54 ACAD. MGMT. J. 1045, 1051 (2011). Incentives such as online credit not only increase the number of overall participants, but also the likelihood of participants completing the survey. *See* Andrew T. Fiore et al., *Incentives to Participate in Online Research: An Experimental Examination of "Surprise" Incentives*, *in* PROCEEDINGS OF THE SIGCHI CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS 3433, 3434 (Apr. 26, 2014).

169.   The American Community Survey is large, well-known, and nationally representative and is used by governments to determine how federal and state funds should be distributed to various potential recipients. *See What Is the American Community Survey?*, U.S. CENSUS BUREAU (June 22, 2015), https://www.census.gov/programs-surveys/acs/about.html [https://perma.cc/F2Z4-2K9J] ("The American Community Survey (ACS) is an ongoing survey that provides vital information on a yearly basis about our nation and its people. Information from the survey generates data that help determine how more than $400 billion in federal and state funds are distributed each year.").

170.   The power of a test when comparing across groups is larger when the groups have the same number of observations. *See* Shayna A. Rusticus & Chris Y. Lovato, *Impact of Sample Size and Variability on the Power and Type I Error Rates of Equivalence Tests: A Simulation Study*, 19 PRAC. ASSESSMENT RES. & EVALUATION 1, 7 (2014) (noting in Table 1 that "[e]qual sample sizes are more powerful than unequal sample sizes"). As states are naturally of different sizes, we oversampled smaller states with unique noncompete laws so as to increase the statistical power of hypothesis testing in future subgroup analyses.

171.   Given the size of its population, oversampling the Golden State was not necessary, notwithstanding California's importance in the noncompete arena.

The survey was relatively long, and many of the questions were cognitively intensive.[172] Because we were concerned that some respondents might not take the survey seriously as a result, we employed the use of "attention filters,"[173] which require respondents to answer trivial questions in a certain way or else the survey session terminates.[174] In accord with standard practice in this setting,[175] we incorporated three attention filters in total—one in the beginning, one in the middle, and one toward the end of the survey.

C. Preliminary Data Cleaning

We began cleaning our survey data by carefully screening them for four different sorts of problems. In these preliminary rounds of cleaning, our goal was to identify unreliable or repeat respondents whose entire set of responses to the survey's questions should be dropped—i.e., entire observations that we concluded were unreliable or duplicative and so should not be included in the final sample to be used for analysis. In the next Section, we discuss how we went about "refining" the remaining observations when they suffered from missing data, typos, or other idiosyncratic errors, but nevertheless contained, in our view, objectively useful and reliable data.

---

172.   Cognitively difficult questions are more likely to lead to respondents providing satisfying rather than accurate answers. *See* Jon A. Krosnick, *Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys*, 5 APPLIED COGNITIVE PSYCHOL. 213 (1991). Alternatively, respondents may give up only part way through. *See* Stéphane Ganassali, *The Influence of the Design of Web Survey Questionnaires on the Quality of Responses*, 2 SURV. RES. METHODS 21, 28 (2008). Longer surveys are less likely to be taken up or finished. *See* Mirta Galesic & Michael Bosnjak, *Effects of Questionnaire Length on Participation and Indicators of Response Quality in a Web Survey*, 73 PUB. OPINION Q. 349, 355, 358 (2009). Interestingly, evidence suggests that web surveys may reduce the cognitive load relative to other forms of data collection since participants are more likely on average to re-read questions and to take more time to answer, but these tendencies vary with respondent characteristics. TOURANGEAU, *supra* note 159, at 146.

173.   Daniel M. Oppenheimer et al., *Instructional Manipulation Checks: Detecting Satisficing to Increase Statistical Power*, 45 J. EXPERIMENTAL SOC. PSYCHOL. 867, 868 (2009).

174.   For example, in a series of survey questions asking a respondent how much he or she agrees or disagrees with a particular statement, an attention filter question might simply say, "Please select Strongly Disagree." If the respondent does not then select "Strongly Disagree" in response to this directive, the survey session ends. Importantly, our survey as implemented did not allow respondents to go back and change previously selected answers to questions.

175.   During our conversations and in our other communications, Qualtrics consistently advocated the use of such filters to improve data quality.

In the first round of preliminary cleaning, we began with the full set of survey "starters" and identified those who (1) did not finish the survey, (2) were filtered out as a result of failing one of the survey's three attention filters, or (3) were not part of the sample of interest. In total, we invited over 700,000 individuals to participate in our survey. Of those who supposedly received an online invitation to take the survey from a panel partner, only 105,053 acknowledged receiving their invitation, and only 79,328 individuals actually began taking the survey. The numbers of invitations and acknowledgements by panel partner are shown in Table 1.

Table 1
RECEIPT OF SURVEY INVITATIONS

Panel Partner	Invitations	Acknowledged	Proportion
ClearVoice	181,811	60,603	0.33
Sample Strategies	155,849	17,720	0.11
SSI	27,700	6,420	0.23
Innovate	8,221	5,990	0.73
Precision Sample	260,360	4,893	0.02
Toluna	43,568	4,200	0.10
GMI	22,091	3,594	0.16
Samplify	12,581	1,633	0.13
Total	**712,181**	**105,053**	**0.15**

Note: The Invitations column reports the number of invitations sent by the panel partner to potential respondents. The Acknowledged column contains the number of respondents confirmed to have opened or viewed their invitation. The Proportion column is the ratio of the Acknowledged column to the Invitations column.

Table 2 shows how many of these initial survey takers made it through the first round of cleaning and, for those survey starters who were eliminated from the survey pool, the specific reason for their removal. Of those who began the survey, 28,824 respondents were simply not in the sample of interest,[176] leaving 50,504 respondents.[177] Of these remaining 50,504 respondents, 28,906 (57.2%) ultimately did not finish the survey and were eliminated for this reason.[178]

With respect to the survey's three attention filters, each of which requires a particular answer in order for the respondent to be allowed to continue, the first attention filter caught and eliminated

---

176. Again, the sample of interest was all labor force participants aged 18–75 who are in the private for-profit sector or the private nonprofit sector or who are employed in a public healthcare system.

177. The breakdown of those who were deemed not in the sample of interest is as follows: 11,073 were neither working nor looking for work; 4,417 were self-employed; 3,031 were employed by the government; 3,876 reported working for a public employer that was not a public healthcare system. Two other major indicators that a potential respondent was outside of the sample of interest were not having an IP address in the U.S. (2,253) and being outside the age range of 18 to 75 (1,920).

178. In practical terms, the label "Did Not Finish" in Table 2 indicates that these individuals did not receive a "survey complete" page from Qualtrics.

3,423 respondents, the second caught and eliminated 976, and the third caught and eliminated 1,530.[179] Overall 29% of those who started the survey in our sample of interest completed the survey and made it through round one of the cleaning process.

Table 2
ROUND ONE CLEANING: FILTERING RESPONDENTS

	Frequency	
Total Started Survey	79,328	
Not in Population of Interest	28,824	
Not working and not looking	11,073	
Self employed	4,417	
Non-healthcare public non-profit	3,876	
Government	3,031	
IP address not in U.S.	2,253	
Not age 18–75	1,920	
Over quota	1,590	
Unemployed (over quota)	631	
Not U.S. resident	33	
		Percent
Total Started Survey in Population of Interest	50,504	100.0
Did Not Finish (Not Otherwise Filtered Out)	28,906	57.2
Attention Filters	5,929	11.7
Attention Filter 1	3,423	6.8
Attention Filter 2	976	1.9
Attention Filter 3	1,530	3.0
Within-Survey Inconsistency or Unreasonableness	1,001	2.0
Signed noncompete, but says 0% signed	195	0.4
Length of interview < 1/3 of median finish time	177	0.4
More than 30 employers in last 5 years	127	0.3
Signed more CNCs than had employers in last 5 years	110	0.2
Did not sign noncompete, but says 100% signed	76	0.2
Occupation text entry invalid	75	0.2
More than 30 positions within employer	72	0.1
Unaware of noncompete, but says 100% signed	67	0.1
Duplicate from pilot (49) or other (2)	51	0.1
Industry text entry invalid	51	0.1
Total Remaining After Round One	14,668	29.0

Note: This table presents the frequencies of eliminated respondents. The Attention Filters were three trivial questions (at the beginning, middle, and end of the survey) respondents had to answer correctly in order to continue the survey. Respondents were unable to change their answers. The Within-Survey Inconsistency or Unreasonableness filtering process was installed by Qualtrics after the first batch of about 8,000 completed surveys, and thus affected only survey responses collected later. (The first batch of respondents are flagged and filtered later in the process). Respondents were flagged for recording Yes, No, None, Not Sure, or N/A in the occupation and industry text entry boxes. In round three, each entry was also manually inspected for validity.

---

179.   We also sought to identify and eliminate individuals who behaved in ways that were clearly inconsistent *within* the four corners of the survey, but in the first round of cleaning, we were only able to do this for a relatively small subsample of the data. After we received the first 8,000 or so completed surveys, we began to examine the data for inconsistencies. Qualtrics agreed to replace those respondents who completed the survey but submitted observably inconsistent or unreasonable answers. Accordingly, we built a flagging program "mid-stream" that identified respondents providing inconsistent or unreasonable answers. Qualtrics incorporated some of these flagging criteria into the final rounds of data collection in order to expedite the cleaning process. Only those respondents eliminated through the use of these flags (i.e., after more than half of the survey collection was complete) appear in Table 2. For most of the sample, we apply the flagging criteria in round four of the data cleaning process. We describe the specific flags in that discussion below.

In the second round of cleaning, we investigated the fact that multiple observations may (and do) arise from the same IP address and also that completed surveys that might otherwise be included in our data were at times not the first observation from a particular IP address. To illustrate, at the end of round one of our preliminary cleaning, of the 14,668 completed surveys, 10,446 had a unique IP address, while 2,331 shared the same IP address with one other observation. At the extreme end of the spectrum, one IP address was observed in our initial survey data 66 times, although only one of these instances involved a complete response.

Because IP addresses are tied to a specific online device, the fact that the same IP address was observed more than once raises two questions that are relevant to the proper cleaning of our survey data: First, are respondents retaking the survey, potentially changing their answers to avoid being filtered out by our screens? Second, are different observations that are linked to the same IP address from the same person or potentially from two or more different people?[180]

To address the issues that arise out of both of these questions, we opted for a conservative approach and decided to retain only the earliest observation from any particular IP address and only if that observation involved a complete set of survey answers.[181] Of the 14,668 complete responses from round one, this cleaning regimen preserved 12,369 observations. In Table 3,[182] we show the number

---

180. One possible reason for so many repeat takers from the same IP address comes from a particular screening mechanism Qualtrics began to apply after the first 8,000 or so completed surveys had been received. Specifically, a filtering question at the beginning of the survey asked if respondents earn more than $40,000 from their current, highest-paying job. Later in the survey, when respondents were asked how much they made in their current, highest-paying job, sometimes they indicated a number that contradicted their previous answer. Instead of terminating the survey and revoking payment for these individuals, Qualtrics used a screening mechanism that notified them of the contradiction and forced them to fix it. Since respondents were not allowed to "go back" in the survey, the only way for these individuals to "fix" their response was to start the survey over. This process likely resulted in some people taking the survey multiple times from the same IP address.

181. To be more concrete, if a respondent does not complete the survey the first time an IP address is observed and later someone with the same IP address does complete the survey, we remove even this completed survey from our sample.

182. In Table 3, the "Number from the Same IP Address" variable was created before we removed the observations flagged in round one of our preliminary cleaning. Retaining these observations is necessary to accurately assess whether a completed survey was in fact the first observation from a particular IP address, given that round one drops observations in which individuals—who may later complete the survey—begin the survey but fail to finish it. Thus, while we call this round two of our preliminary cleaning, we performed it simultaneously with round one.

and proportion of completes that we retained and that we dropped for each number of observations from the same IP address.[183]

Table 3
ROUND TWO CLEANING: DUPLICATE IP ADDRESSES

Number from Same IP Address	Total		Retained in Round Two		Dropped	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
1	10,446	71.22	10,446	84.45	0	0.00
2	2,331	15.89	1,284	10.38	1,047	45.54
3	927	6.32	374	3.02	553	24.05
4	387	2.64	115	0.93	272	11.83
5	243	1.66	77	0.62	166	7.22
6	160	1.09	39	0.32	121	5.26
7	78	0.53	18	0.15	60	2.61
8	24	0.16	5	0.04	19	0.83
9	12	0.08	1	0.01	11	0.48
10	14	0.10	3	0.02	11	0.48
11	16	0.11	3	0.02	13	0.57
12	4	0.03	1	0.01	3	0.13
13	3	0.02	0	0.00	3	0.13
14	2	0.01	0	0.00	2	0.09
24	1	0.01	0	0.00	1	0.04
30	4	0.03	1	0.01	3	0.13
36	5	0.03	1	0.01	4	0.17
37	10	0.07	1	0.01	9	0.39
66	1	0.01	0	0.00	1	0.04
Total	**14,668**	**100.00**	**12,369**	**100.00**	**2,299**	**100.00**

Note: This table presents the frequency of post-round one respondents (total, retained by round two, and dropped by round two) by the number of observations with the same IP address. Round two cleaning retained the chronologically first observation from an IP address when it resulted in a completed survey. Note that the number in the Number from Same IP Address column comes from the full sample to account for the fact that the earliest observation otherwise may have been removed in round one. The numbers in the percent columns are column percentages.

In the third round of cleaning, we sought to identify and remove repeat survey takers who used different devices (presumably intentionally).[184] To rid our data of these additional observations, we

183. Even after explicitly addressing the fact that the same IP address can appear in our data multiple times, casual inspection of the data revealed additional repeat takers, who must have taken the survey multiple times using different devices. Qualtrics informed us during phone conversations in late 2014 that panel partners drop duplicate responses, although sometimes duplicate responses may still occur as a result of survey re-routing—i.e., when individuals submit contradictory information, they are sometimes not allowed to "go back" in the survey when they are alerted to the contradiction, but may succeed at starting the survey over again if they attempt to do so. Leaving this concern aside, even if panel partners do drop duplicate responses, then repeat takers could nevertheless exist in our data for either of two reasons: First, respondents may have accounts with multiple panel partners, at least two of which were involved in data collection for our survey. Second, respondents may have a second account with the same panel partner. Panel partners verify demographics and attempt to mitigate this problem on the front end. However, if individuals submit different demographic information during the sign-up process, then panel partners may not be able to identify and eliminate them.

184. Reg Baker et al., *Validating Respondents' Identity in Online Samples*, *in* ONLINE PANEL RESEARCH: A DATA QUALITY PERSPECTIVE 441, 442 (Mario

closely examined the sets of individuals who have the same gender, age, and race, and who live in the same state, work in the same county, and live in the same zip code.[185] Unfortunately, if individuals are repeat survey takers and use completely different identities and different devices, then there is simply no easy way to identify and remove them as repeat takers in our data.

Table 4 shows the number of observations by the number of potential duplicates in our survey data. According to our method, of the 12,369 valid respondents after the second round of preliminary cleaning, 11,704 of them have zero potential duplicates—i.e., there does not exist another respondent with the same gender, same age, and same race, living in the same state and zip code, and working in the same county. Of the remaining 665 observations, 620 of them share the same characteristics with one other observation, 33 share the same characteristics with two other observations, and 12 share the same characteristics with three other observations.

Table 4
POTENTIAL REPEAT RESPONDENTS WITH DIFFERENT IP ADDRESSES

Number of Potential Duplicates	Frequency	Percent
0	11,704	94.62
1	620	5.01
2	33	0.27
3	12	0.10
Total	**12,369**	**100.00**

Note: This table presents a tabulation of the number of potential duplicates from different IP addresses and their frequency within the data.

Matching respondent demographic characteristics does not necessarily indicate that survey responses originated from the same individual: Panel partners may have reached distinct individuals with the same gender, age, and race, living in the same state and zip code, and working in the same county. To examine this issue further, we compared the answers from suspected duplicate surveys to assess whether and to what extent these answers also aligned.[186] We found

Callegaro et al. eds., 2014) ("[T]he incentivized nature of panels may encourage some people to go to elaborate lengths to maximize participation to their financial advantage.").

185.   This list of identifying demographic traits emerged from our inspection of the data. Whenever we found multiple surveys that we believed originated with the same individual, we adjusted our flagging algorithm so that it would identify the case in question and other similar cases.

186.   We characterized responses as duplicates when free-form text entries for occupation, industry, salary, and other important characteristics such as tenure, experience, education, and the year in which respondents finished their schooling were identical. It may be worth emphasizing that this analysis relied on affirmatively typed answers, not just answers produced by ticking a box.

that of the 665 potential repeat survey takers, 541 came from 262 individuals who took the survey at least twice.

   To better understand the manner by which these individuals were able to take the survey several times, we investigated whether the observations we classify as duplicates involved the same panel partner or different panel partners. We display the results in Table 5. Of the 249 individuals whom we conclude took the survey twice (498 observations), 206 of them (412 observations) used different panel partners. Of the nine individuals who took the survey three times (27 observations) and the four individuals who took it four times (16 observations), every respondent made use of more than one panel partner. These results imply that because panel partners verify their respondents' demographic information and disallow individuals from signing up for multiple accounts, individuals registered with multiple panel partners may more often be the source of duplication when conducting a large-scale online survey.

Table 5
REPEAT TAKERS AND PANEL PARTNERS

| Number of Duplicates | Same Partner? | | Total |
	No	Yes	
1	412	86	498
2	27	0	27
3	16	0	16
Total	**455**	**86**	**541**

Note: This table presents a cross tabulation of the number of identified duplicates and whether the duplicates came from the same panel partner.

   In the end, we resolved the repeat-taker problem by retaining only data from the first survey response submitted by an individual; we can obviously only use one set of survey answers, and later sets of answers are contaminated as a result of the individual having already taken the survey. Accordingly, we identified 541 duplicative observations submitted by 262 individuals and decided to keep only 262 of these flagged observations, dropping the remaining 279. Thus, after beginning this cleaning round with 12,369 observations, we completed it with 12,090 observations.

   In the fourth round of cleaning, we reviewed individual answer responses to identify those respondents whose survey answers were internally inconsistent or unreasonable in light of their other survey answers. This process alerted us to several issues of importance,[187] two of which we address at this stage: (1) certain individuals failed to indicate that they were self-employed, retired, or government

---

   187.   *See infra* Section II.D (detailing our approach to "fixing" answers that are inconsistent or unreasonable in otherwise valuable observations).

employees—i.e., outside of our sample population of interest—in our initial filters but later identified themselves in the survey as falling into one of these categories;[188] and (2) some people wrote gibberish, cursed, provided clearly inconsistent answers, or otherwise revealed that they did not take the survey very seriously by, for example, completing it in an implausibly small amount of time.

With respect to the second discovery, we recognized that it was important if not essential to distinguish between survey respondents who were intentionally noncompliant and survey respondents who may have made a few idiosyncratic errors.[189] More precisely, our goal was to exclude any respondent that we believed was likely to have provided many unreliable responses—e.g., obvious gibberish, containing no useful information—but to retain those observations containing real—albeit potentially mismeasured—data. Determining whether data are "real" is ultimately subjective, but we believe our application of a few bright-line, reproducible filters allowed us to further improve the quality of our final sample.

We began by considering individual respondents whose survey completion time might have been "too fast."[190] Our assumption is

---

188. During our initial review of the data, we observed that our sample of 12,090 individuals still seemed to include many respondents who reported that they were not a part of the sample of interest. These individuals claimed to be self-employed, retired, or employed in public administration, and should have been filtered out by previous questions. They were not, however—possibly because they misunderstood their status at the outset. For example, individuals often notified us about their self-employment status in questions related to occupation or in the comments section at the end of the survey. Our initial set of respondents also included individuals who indicated that they were in public administration when they were asked to click the two-digit industry category that fit them best. We were able to identify some retirees by their submissions in the end-of-survey comments section. We discovered 43 individuals who recorded that they were employed in public administration, 136 made it clear that they were business owners, and six were retired. Again, these individuals were not in our sample of interest, and so we excluded all 183 offending observations. Note that these three out-of-sample categories are not mutually exclusive, which explains why 183 is not the sum of the number of individuals in each of the three categories. We were left with 11,902 respondents, after eliminating seven additional respondents for reasons unrelated to their being out of the sample of interest. To explain, our survey randomly assigned respondents to particular information conditions—research that we do not discuss in this Article—and so we dropped the seven individuals who took the survey before Qualtrics had properly installed the randomization component.

189. We address how we refine the data to make sure it is free from idiosyncratic errors in the next Section.

190. If a respondent's survey-taking speed is correlated with substantive responses, eliminating "fast" test takers can introduce bias. But our filter eliminated only an extreme tail of the distribution—answers very unlikely, in our opinion, to be

that reliable survey data hinges on respondents reading and digesting questions before they respond and that this process requires some minimum amount of time.[191] We conservatively defined "too fast" to be one-third of the median survey completion time for the remaining 11,902 respondents.[192] Our survey's median finish time was roughly 28 minutes from start to finish, which translates to a cutoff duration for inclusion in the final sample of 9 minutes and 20 seconds. This screening process eliminated 15 observations from our data.

Next we reviewed individual text entries, including respondent-reported job duties, job title, industry of the employer, and the open-ended survey comments at the end of the survey. Individuals who do not answer the free-form text entry questions, but instead write either gibberish or curse words or use other language to indicate that they are not engaging seriously with the survey, seem highly likely to have answered many other questions inaccurately—especially when they fail to answer the employment-related questions diligently.[193] We dropped 256 more observations as a result of this process.

Finally, we used the questions on perceptions of noncompete incidence within the respondent's industry and employer as a means to screen for noncompliance.[194] We scoured the data to find answers to important questions that were clearly, demonstrably inconsistent with each other. For instance, if, at the beginning of the survey, the individual indicates that he has never heard of a noncompete and then later indicates that 100% of employees in his industry or at his firm sign a noncompete, we dropped the observation. We identified and eliminated 98 such observations. In sum, after removing those respondents who appear highly likely to have been intentionally noncompliant, 11,529 observations remained in our sample.

---

informative. We also do not see any obvious reason why time-to-completion should be correlated with our key variables of interest.

191.   Robert Greszki et al., *The Impact of Speeding on Data Quality in Nonprobability and Freshly Recruited Probability-Based Online Panels*, *in* ONLINE PANEL RESEARCH, *supra* note 184, at 238, 241 ("In effect, there is a, though not perfect, link between very quick responses and low data quality which is supported by evidence.").

192.   According to conversations with Qualtrics staff in April 2014, Qualtrics uses one-third of the median completion time as a standard cutoff.

193.   We identified these particular classes of respondents by examining survey responses by hand. While these individuals may of course answer some of the survey's questions honestly, we view the possible bias caused by leaving their potentially arbitrary responses in the data as a greater threat than the probably less significant sample selection concerns generated by ignoring these answers.

194.   *See supra* Section II.A (describing these questions about a respondent's perceptions in more detail).

D. Data Refinement

Our discussion thus far has covered the design of the survey, its online implementation, and our preliminary cleaning of the collected data. The preliminary cleaning dropped entire observations that we concluded were suspect. For example, we removed respondents who appeared to be intentionally noncompliant on the assumption that most or all of their answers were unreliable or duplicative. In this Section, we outline our approach for refining our remaining data by identifying observations with missing data fields or with entries that are inconsistent with other entries. These individuals appear to have made idiosyncratic errors—including either typographical, mental, or memory errors. We explain our approach for correcting these errors when they can be corrected, and we outline our creation of multiple samples of the survey data that rely on different assumptions.[195]

We began our refinement of the data by attempting to measure when and the extent to which each respondent made idiosyncratic mistakes in answering survey questions. We developed an algorithm to evaluate the many ways an individual could provide inconsistent or unreasonable answers, including separate analyses of responses to individual questions to assess the answer's validity as well as comparisons of the respondent's answers across different questions to detect conspicuous mistakes. The challenge of comparing answers across questions is that, if there is an inconsistency, it can be difficult to determine which question or questions (if not both or all of them) the respondent answered erroneously.[196]

---

195. Refining data to eliminate unmistakable errors is standard practice in survey research. Typically, researchers build data editing rules that flag observations that fail to meet specific characteristics. There are many approaches to resolving errors, including eliminating observations. Perhaps the most common response, however, is to replace the erroneous entries through statistical estimation and imputation. For more information on these data replacement methods, see TON DE WAAL, JEROEN PANNEKOEK & SANDER SCHOLTUS, HANDBOOK OF STATISTICAL DATA EDITING AND IMPUTATION (2011). In addition, see U.N. STATISTICAL COMM'N & ECON. COMM'N FOR EUR., CONFERENCE OF EUROPEAN STATISTICIANS METHODOLOGICAL MATERIAL, EVALUATING EFFICIENCY OF STATISTICAL DATA EDITING: GENERAL FRAMEWORK (2000), http://www.unece.org/fileadmin/DAM/ stats/publications/editingefficiency.pdf [https://perma.cc/2NDT-WJNA].

196. The following are the types of responses, categories of questions, and problematic answers we analyzed: (1) the survey is missing responses to questions; (2) the respondent has provided answers to individual questions that are impossible; (3) the respondent's educational patterns are nonsensical (e.g., has a Ph.D. and is also working toward an associate's degree); (4) the respondent reported finishing school when younger than 13 years old; (5) the respondent's tenure or experience

After flagging all questionable survey responses using our criteria, we added up the total number of flags that each respondent received. Table 6 presents the number of survey respondents by the total number of error flags.

Table 6
TOTAL NUMBER OF ERROR FLAGS

Number of Flags	Frequency	Percent	Cumulative
0	9,478	82.21	82.21
1	1,850	16.05	98.26
2	162	1.41	99.66
3	30	0.26	99.92
4	9	0.08	100.00
Total	**11,529**	**100.00**	

Note: This table presents a tabulation of the total number of flags in the final survey sample.

There are 9,478 survey respondents—or 82.21% of our final survey sample—who received no flags whatsoever, while the bulk of the remainder—1,850 respondents (16.07%)—earned exactly one error flag under our criteria. To investigate which flags are most common in the sample and how likely each kind of flag is to appear with one or more other flags, Table 7 below shows a cross tabulation of the number of flags by the flagging criteria.

The criterion that flagged the greatest number of observations is clearly the restriction that individuals cannot earn less than the minimum wage conditional on the hours and weeks they reported having worked—1,123 respondents of the 2,051 receiving any flags (54.8%). Failure to report one's income accurately may be a signal of intentional noncompliance generally, but it may also be specific to

---

indicated that the respondent has held the same position since he or she was 12 years old or younger; (6) the respondent's tenure in his or her current position with an employer is greater than his or her tenure at all positions with the firm combined; (7) the respondent's employer's establishment size (i.e., the number of employees at a particular branch or office) is larger than the overall size of the employer (i.e., all employees with employer); (8) the number of employers in the last five years is greater than 30; (9) the respondent signed a noncompete with more employers than the respondent has had in last five years; (10) the respondent reported having held more than 30 positions within the same employer; (11) the respondent's spouse's income is reported to be more than one million dollars; (12) the respondent reported having taken online surveys since before 1989 (i.e., for more than 25 years); (13) the respondent learned about noncompetes at an age older than the respondent supplied as his or her current age; (14) the respondent reported being recruited to join over 50 employers within the first 10 years of tenure at his or her current employer; (15) the respondent has had more than 50 different job offers while employed with his or her current employer; (16) the respondent's reported annual income is less than minimum wage earnings conditional on the reported hours and weeks worked; (17) respondent's reported income amounts to more than $300,000 annually.

the sensitive nature of the income question.[197] These individuals may have responded appropriately to other questions. We see evidence in favor of this proposition in the fact that for 1,007 of these flagged individuals the below-minimum-wage flag is their sole idiosyncratic survey response. The other criteria that flagged a significant number of respondents are (1) the flag for respondents who reported that their tenure in their current positions with an employer is greater than their total tenure at the firm (251 observations) and (2) the flag that identified missing responses (195 observations).[198]

Table 7
NUMBER OF FLAGS BY FLAG TYPE

Flag Type	Number of Flags				Total
	1	2	3	4	
Earned less than minimum wage	1,007	97	13	6	1,123
Tenure in current position > tenure at firm	183	55	11	2	251
Missing responses	181	9	4	1	195
Annual earnings more than $200K	157	13	3	0	173
Average # of offers/year of tenure > 10	58	30	6	3	97
Finished school before age 13	52	28	10	2	92
# noncompetes > # of employers in last 5 years	64	12	4	4	84
Recruited by over 50 employers in last year	42	23	4	2	71
Establishment larger than firm	42	15	1	0	58
More than 30 employers in last 5 years	22	6	4	2	34
Started related work before age 13	12	7	7	4	30
Values not possible	1	14	4	0	19
Started with current firm before age 13	4	3	5	4	16
Taking *online* surveys before 1990	14	2	0	0	16
Started in current position before age 13	0	2	5	4	11
More than 30 noncompetes in last 5 years	0	4	4	2	10
Learned about noncompete when older than current age	6	1	0	0	7
Spouse's annual income greater than $1M	2	2	1	0	5
Degree pursuing inconsistent with degree held	3	0	0	0	3
Finished school before 1939	0	0	3	0	3
More than 30 jobs within same employer	0	1	1	0	2
Total	**1,850**	**162**	**30**	**9**	**2,051**

Note: This table presents a tabulation of the reasons that individual respondents in the data were flagged in the final survey sample.

---

197. *See* Jeffrey C. Moore et al., *Income Measurement Error in Surveys: A Review* 5-19 (1997) (unpublished paper) (on file with the Census Bureau), https://www.census.gov/srd/papers/pdf/sm97-05.pdf [https://perma.cc/TFC3-5HRV].

198. With respect to the first flag, we acknowledge that a report that one's tenure in one's current position is longer than one's tenure at one's current firm may reflect a misunderstanding of the question: If they had occupied similar positions with their previous employers, respondents may have had in mind their tenure in a particular *type* of position beyond just their experience with their current employer. This explanation does not fully explain these inconsistencies, however, as 197 observations still reported that their tenure in their current position was longer than their total experience in their position type. Thus, these mistaken responses appear to us to be mental or typographical errors. With regard to missing responses, these occurred only in the early days of the survey from April 22, 2014, to early in the morning on April 24, 2014, when a few of the survey's questions were apparently not functioning properly.

We turn next to the question of how we ought to handle these idiosyncratic (as opposed to noncompliant) errors in our survey data. We adopted a general policy of "do no harm" to the data; rather than deciding to keep, drop, or modify these data once and for all, we instead defined four separate samples that permit us to test whether and how cleaning the idiosyncratic errors affects the results of our analyses and that allow other researchers to use whichever sample suits their research best.[199] Sample 1 ignores the idiosyncratic errors, leaving them as they are in the preliminarily cleaned data. Sample 2 excludes all individuals from the data who have received one or more flags of any type. Sample 3 retains all of the observations, but replaces offending variables with missing values. Sample 4 preserves the respondents who have flagged survey answers, but also corrects for these idiosyncratic errors by imputing the missing and offending values, conditional on other covariates, when necessary.[200] Table 8 presents and defines these sample choices.

Table 8
SAMPLE SELECTION

Sample 1	Retain offending values
Sample 2	Drop observations with any flags
Sample 3	Replace offending values as missing
Sample 4	Impute or otherwise correct for offending values

Samples 1 and 2 are obviously simple to create, but assembling Samples 3 and 4 requires more analysis and additional assumptions. In particular, as we note above, when we determined that the values of related variables were inconsistent with each other, it was often *ex ante* unclear which of the "inconsistent" survey answers was actually erroneous—in isolation, any of the answers often made sense.[201] We proceeded by assuming that the error was more likely to occur in answering the question about which a respondent was more likely to

---

    199.   *See* Joachim R. Frick & Markus M. Grabka, *Item Nonresponse on Income Questions in Panel Surveys: Incidence, Imputation and the Impact on Inequality and Mobility*, 89 ASTA ADVANCES STAT. ANALYSIS 49, 50 (2005).
    200.   We find Sample 4 attractive for many reasons: It uses all of the data and corrects, to the best of our ability, for the idiosyncratic errors we have identified. See *infra* Part III for our analysis using Sample 4.
    201.   For example, if we observe that an employer's establishment size is bigger than the size of the entire firm size, did the respondent mistakenly report the wrong establishment size or the wrong firm size? If we believe the respondent made a mistake when selecting the firm's size, we should increase the firm size to at least the establishment size. If we instead conclude that he or she mistakenly recorded the wrong establishment size, then we should reduce the establishment size to be at most the size of the firm.

be ill-informed or, alternatively, that was more challenging for the respondent to answer.[202] For example, in multi-establishment firms, employees are more likely to be knowledgeable about how many other employees work in their own place of employment than about their employer's size firm-wide. Our approach thus implies that the firm size variable is more likely to be incorrect, and that we ought to replace its original value with the closest "consistent" value—i.e., resetting it to be as large as the establishment size value.

We apply 17 flags to our data to identify what seem to be idiosyncratic errors. Ten of these flags rely on inconsistencies *between* individual answers. In these cases, we must either drop or adjust at least one of the inconsistent variable values to achieve (perhaps minimal) consistency. Table 9 reports the flags that use the existence of inconsistent variable values to identify questionable observations and lists the offending variable that we believe ought to be dropped or replaced as the least reliable among the inconsistent set. We construct Sample 3 by replacing the least reliable variable values in Sample 1 as missing when an inconsistency arises.

Table 9
SAMPLE 3 CORRECTIONS FOR MULTIPLE VARIABLE FLAGS

Flag Type	Offending Variable
Earned less than minimum wage	Annual income
Tenure in current position > tenure at firm	Tenure in position at firm
Finished school before age 13	Year finished school
# noncompetes > # of employers in last 5 years	# noncompetes
Establishment larger than firm	Firm size
Started related work before age 13	Experience
Started with current firm before age 13	Tenure
Started in current position before age 13	Tenure in position
Learned about noncompete when older than current age	Age learned
Degree pursuing inconsistent with degree held	Ultimate degree

Note: When there is an inconsistency between, for example, two entered values, we correct the entry for the variable about which the respondent is more likely to be uncertain or which the respondent is less sensitive to reporting accurately.

In Sample 4, by contrast, we aimed to maximize the survey data's informational content by replacing inconsistent, unreasonable, or missing values—whenever possible—with new values that are both internally consistent and more likely to be accurate.[203] To accomplish this, we deploy three data replacement strategies.

---

202. We considered numerous options for how to "fix" such errors, including imputing the values of both variables. We chose to fix only one variable in order to preserve as much of the respondent's originally submitted data as possible, with the constraint that the data not be internally inconsistent. Ultimately, the choice of which inconsistent variable value to fix is a somewhat subjective decision.

203. *See* Melvin Stephens, Jr. & Takashi Unayama, *Estimating the Impacts of Program Benefits: Using Instrumental Variables with Underreported and*

First, as a general matter, we "repair" entries that are marred by idiosyncratic inconsistency; typically, our approach is to replace the less reliable offending value with the value that is closest to the originally submitted value that would not be inconsistent with the respondent's other answers.[204] Second, in a few instances in which we observe values that are extreme outliers, such as having more than 30 employers during the last five years, we do nothing—i.e., we retain the original suspect value—assuming that if researchers view the respondent's answer as unreasonable, they will choose whatever imputation method best suits their needs.[205] Third, when an answer is clearly unreasonable or missing, and there is no workable single imputation procedure, we make use of multiple imputation methods to calculate a substitute value for the original missing or unreasonable survey entry.[206]

We also reviewed by hand the values of particular critically important variables in hopes of ferreting out unreliable entries that are more difficult to detect. In effect, we used more elaborate and time-consuming approaches to expose unreasonable or unreliable entries and then returned to single or multiple imputation methods to replace the suspect values.

For instance, we examine all income values over $200,000 and any entries with potential typos individually.[207] For incomes over $200,000, we cross-checked the employee's reported occupation, job title, industry, experience, and education to determine whether the numbers reported by the respondent are reasonable. With respect to

---

Imputed Data 22 (Nat'l Bureau of Econ. Research, Working Paper No. 21248, June 2015), http://www.nber.org/papers/w21248.pdf [https://perma.cc/Z9H2-9TL6] ("Survey non-response has continued to rise in recent years which has generated an increase in the share of observations with imputed values on a number of commonly used datasets.").

204.   This form of imputation is known as single imputation. Starting with the nonmissing values of other variables, single imputation essentially deduces the value of a missing or erroneous observation. Single imputation methods include using the mean of the variable from observations with reliable values, estimating missing values based on an expected ratio between observations, replacing the offending value with the nearest existing observation, and using a regression to determine missing values, among other logical deduction methods. *See* JELKE BETHLEHEM ET AL., HANDBOOK OF NONRESPONSE IN HOUSEHOLD SURVEYS 418-42 (2011).

205.   *See* Stephens & Unayama, *supra* note 203, at 3, 10, 22-23.

206.   We describe our multiple imputation work in Section II.F, deferring the discussion until after we explain our weighting strategy in Section II.E.2.

207.   In particular, we first sought out income values that seemed to us suspicious, before checking by hand whether they were objectively within a reasonable range or whether we ought to impute them.

potential typographical errors, we built a "test" ratio by comparing the individual's current income to next year's expected income (a variable that is non-missing for all responses).[208] For yearly income values in our data, this ratio is less than 0.2 and greater than 9 in 609 cases. We scrutinized these observations as well as high-income observations and applied a situation-specific test to decide whether to replace these suspect values.[209] When we were unable to produce reliable alternative income values or when these alternative values dropped below minimum-wage levels (based on the reported number of hours and weeks worked), we swapped out the problematic values as missing in order to replace them later—along with other variables' missing values—with multiple imputation estimates.[210]

In addition to hand cleaning unreasonable income values, we also inspected the entries for the occupation and industry variables by hand. We began by having two sets of research assistants (RAs) code each survey respondent's free-form written answer to "what your current employer produces or does" as well as the respondent's job title and a written job-duties description using the 2010 Standard Occupation Classification System and the 2012 North American Industry Classification System. We then had a third RA compare the occupation and industry coding entries from the first two iterations, and revisit those entries that did not match.

This coding protocol produced three distinct sets of hand-coded occupation and industry codes; we exploited the two-digit codes that the respondents had selected themselves to choose between coding

---

208.   A typo in an entry in one year but not in the other (and assuming wage growth is relatively low) should produce a ratio close to either 0.1 or 10, which are thus indicative of a potential problem.

209.   A response-by-response analysis of these 609 wage records (taking into account each individual's wage history, occupation, hours worked, experience, tenure, and education) revealed a nontrivial number of errors in the wages income variable. The most common errors were an additional or missing zero or two (117 observations). Twenty-one individuals clearly lied about their income, conditional on their occupation and experience (e.g., grocery store bagger earning $2 million a year). Others indicated they had zero income and we were unable to use their wage history to impute their income (242 observations). When possible, we imputed incomes in the following way: If, relative to prior wage history and the expected future wage trajectory, there was simply a difference in the number of zeros, the number of zeros was adjusted to align with the other numbers. If a respondent's reported current income was wildly different from the previous year's number and the number from the following year, we replaced the current income value with the average of the surrounding years (assuming the surrounding years were reasonable).

210.   Please see *infra* Section II.F for details about the multiple imputation approaches we employ in our work.

outcomes when there was disagreement among the RAs.[211] The logic of this replacement method is as follows:

1. If two RA codes match the respondent's self-selected code, use the self-selected code.
2. If no RA codes match each other, but one RA code matches the self-selected code, use the self-selected code.
3. If two RA codes match each other and no RA codes match the self-selected code, use the matching RA codes.
4. If no RA codes match each other and none of them match the self-selected code, use the self-selected code.
5. If two RA codes match each other and one RA code matches the self-selected code, use a fourth RA to break the tie.

To conclude this review of our data refinement efforts, Table 10 displays our replacement methods for inconsistent, unreasonable, or missing values in otherwise informative observations.

Table 10
SAMPLE 4 CORRECTIONS FOR FLAGGED VARIABLES

Panel A: Flags Using Multiple Variables	
Flag Type	**Method**
Earned less than minimum wage	Impute
Tenure in current position > tenure at firm	Tenure in position = Tenure at firm
Average # of offers/year of tenure > 10	Impute
Finished school before age 13	Year finish school = YOB + 13
# noncompetes > # of employers in last 5 years	# noncompetes = # employers
Establishment larger than firm	Firm size = Establishment size
Started related work before age 13	Experience = 2014 – (YOB + 13)
Started with current firm before age 13	Tenure = 2014 – (YOB + 13)
Started in current position before age 13	Tenure in position = 2014 – (YOB + 13)
Learned about noncompete when older than current age	Age learned about noncompete = Age
Degree pursuing inconsistent with degree held	Do nothing
Panel B: Flags Using One Variable	
Flag Type	**Method**
Missing responses	Impute
Annual earnings more than $200K	Inspect individually, impute if necessary
Spouse's annual income greater than $1M	Inspect individually, do nothing or impute
Finished school before 1939	Impute as 2014 – age + estimated years of school
Values not possible	Impute or do nothing
Recruited by over 50 employers in last year	Impute
More than 30 employers in last 5 years	Impute
Taking *online* surveys before 1990	Do nothing
More than 30 noncompetes in last 5 years	Do nothing
More than 30 jobs within same employer	Impute

Note: This table displays the methods we used to correct inconsistent or implausible entries by flag type when multiple variables are involved (Panel A) and when only one variable is involved (Panel B). YOB stands for year of birth.

---

211.   In the process of examining and cleaning the occupation and industry codes, we discovered that a few of the survey respondents were actually business owners and thus did not belong in the sample. We dropped all 24 such observations, which translated to a final dataset tally of 11,505 observations.

Before we lay out our multiple imputation methods in Section II.F, we must first confront the fact that multiple imputation requires weights constructed so that the weighted sample as a whole reflects the population of interest.[212] Weighting is necessary when using multiple imputation to replace missing values because, without recalibrating the likelihood of drawing each value from the sample to match the likelihood of drawing that value from the relevant population, we would necessarily select replacement values from a potentially skewed, nonrepresentative distribution.

Therefore, to complete the assembly of Sample 4, we first weighted the data to reflect the population of interest. Second, we used multiple imputation to impute missing values. Third, we re-weighted the now completed dataset to reflect population moments.

There is no shortage of weighting methods—we describe some of them below—but all weighting schemes seek to enhance sample data's representativeness. Methods exist for reweighting observed data in a sample to match observed data in the population, but it can be tough to ascertain whether the sample reflects the population on those variables that the researcher can observe in the sample but *not* in the population. Indeed, weighting is conceptually linked to sample selection bias, which refers to the possibility that the *process* of sampling itself—by whatever method—may render certain types of individuals particularly likely or unlikely (relative to their true numbers) to find their way into a sample. When this occurs, the sample is nonrepresentative on the characteristics causing affected groups to be disproportionately likely or unlikely to appear in the sample. Surveys—and in particular online surveys—are especially dubious because individuals choose to take them. For this reason, we next discuss the possible sources, extent, and implications of any selection bias in our data, the role reweighting the data can play in mitigating this bias, and, finally, multiple imputation.

E. Sample Selection

The 2014 Noncompete Survey necessarily produced an online convenience sample, not a typically more representative probability

---

212. *See* JAMES R. CARPENTER, USING SURVEY WEIGHTS WITH MULTIPLE IMPUTATION—A MULTILEVEL APPROACH 18 (Nov. 27, 2012), http://www.lse.ac.uk/statistics/events/SpecialEventsandConferences/CarpenterJR.pdf [https://perma.cc/B8GP-J828] (noting that "[i]n the light of this, for accurate inference within domains for survey data, the imputer needs to include the weights, the domain indicator, and their interaction for valid MI inference in general").

sample.[213] Consequently, one far-reaching concern is that nonrandom selection of respondents into our sample may generate significant bias in the inferences we draw from the data about the population of interest.[214] In this Section, we describe sample selection threats to our survey project. We evaluate the severity of these threats, concluding that any nonrepresentativeness is likely minor in its consequences. Along the way, we discuss potential ways to make inferences from our survey data more robust to sample selection bias. We briefly describe a few of the benefits of reweighting convenience samples, and we then summarize our reweighting scheme in detail.

1. *Sources of Potential Selection Bias*

In an online survey of this length and complexity, a discussion of selection bias issues cannot be avoided.[215] We can divide these potential problems into four broad selection concerns:

1. Not all of the U.S. labor force is online.
2. Not all of those online register to take online surveys.
3. Not all of those who register to take online surveys are invited to take the survey.
4. Not all of those who are invited to take the survey will finish (or even start) it.

We examine each of these four selection issues in the context of our survey project. We highlight the *relative* importance of each and their likely consequences for our data and inferences. In the next Subsection, we describe a set of reweighting strategies.[216] We believe that one these approaches—depending on the specific goals of the research—can significantly mitigate these selection concerns.

---

213. Samuel J. Best & Brian S. Krueger, *Drawing Samples on the Internet*, *in* INTERNET DATA COLLECTION 16-18 (2004); Mary Hibberts et al., *Common Survey Sampling Techniques*, *in* HANDBOOK OF SURVEY METHODOLOGY FOR THE SOCIAL SCIENCES 61-62, 66-67 (Lior Gideon ed., 2012); Samuel J. Best et al., *An Assessment of the Generalizability of Internet Surveys*, 19 SOC. SCI. COMPUTER REV. 131, 131-34 (2001).

214. Kimmo Grönlund & Kim Strandberg, *Online Panels and Validity: Representativeness and Attrition in the Finnish eOpinion Panel*, *in* ONLINE PANEL RESEARCH, *supra* note 184, at 88; Hibberts et al., *supra* note 213, at 66; TOURANGEAU ET AL., *supra* note 159, at 11-13.

215. *See generally* Jelke Bethlehem, *Selection Bias in Web Surveys*, 78 INT'L STAT. REV. 161, 161-62 (2010).

216. *Id.* at 267-71; JELKE BETHLEHEM & SILVIA BIFFIGNANDI, HANDBOOK OF WEB SURVEYS 329-79 (2011).

Of the four selection issues, the first and fourth on the list are "standard" challenges. Other survey approaches must overcome their own versions of these problems.[217] With respect to the first selection concern, for example, random-digit dialing (RDD) methods can only reach the population with a telephone and in-person surveys rely on being able to locate, travel to, and interview the person.[218] Moreover, even though probability samples *begin* with a randomly selected target sample, systematic unit nonresponse can cause the final sample to be a nonrandom draw of the population.[219]

Overcoming this first issue in probability samples has been the subject of significant debate among scholars. As we have already noted, work by Kohut and his colleagues finds that RDD response rates have decreased from 36% in 1997 to just 9% in 2012.[220] Against the backdrop of falling response rates, Wang et al. (2014) shows that by putting more or less weight on particular observations—in this case, using multilevel regression methods and post-stratification[221]— a convenience sample of Xbox users can better predict the outcome of the 2012 presidential election than traditional Gallup polls.[222] Reweighting, however, is no panacea.[223] It is difficult to do well, and the proper execution of any weighting strategy requires great care. Nevertheless, the design and application of valid weighting methods is an active area of research.[224]

Recall that the target population of interest for our survey is U.S. labor force participants aged 18–75 who are not self-employed and not working for government unless they are employed by a public healthcare system. What do we know about reaching the labor

---

217.   TOURANGEAU ET AL., *supra* note 159, at 13, 39; Best & Krueger, *supra* note 213, at 14.

218.   DON A. DILLMAN, MAIL AND TELEPHONE SURVEYS: THE TOTAL DESIGN METHOD 43, 46 (1978).

219.   ROBERT M. GROVES ET AL., SURVEY NONRESPONSE 276 (2002).

220.   KOHUT ET AL., *supra* note 164, at 1.

221.   See R.J.A. Little, *Post-Stratification: A Modeler's Perspective*, 88 J. AM. STAT. ASS'N 1001 (1993), for details on post-stratification, and David K. Park et al., *Bayesian Multilevel Estimation with Poststratification: State-Level Estimates from National Polls*, 12 POL. ANALYSIS 375 (2004), for details on multilevel estimation with post-stratification.

222.   Wei Wang et al., *Forecasting Elections with Non-Representative Polls*, 31 INT'L J. FORECASTING 980, 989-90 (2014).

223.   Matthias Schonlau et al., *A Comparison Between Responses from a Propensity-Weighted Web Survey and an Identical RDD Survey*, 22 SOC. SCI. COMPUTER REV. 128, 128-38 (2004).

224.   Jelke Bethlehem & Mario Callegaro, *Introduction to Part IV*, *in* ONLINE PANEL RESEARCH, *supra* note 184, at 264, 271.

force via online methods (selection concern #1)? A recent study of internet users indicates that the overall internet penetration rate is about 87% for U.S. adults.[225] We reproduce additional findings from the Pew Research Center study in Table 11.

Table 11
INTERNET USERS IN 2014

Among adults, the percent who use the internet, email, or access the internet via a mobile device	
	Use internet
All adults	87%
Sex	
a. Men	87%
b. Women	86%
Race/ethnicity*	
a. White	85%
b. African-American	81%
c. Hispanic	83%
Age group	
a. 18–29	97%[cd]
b. 30–49	93%[d]
c. 50–64	88%[d]
d. 65 +	57%
Education level	
a. High school grad or less	76%
b. Some college	91%[a]
c. College +	97%[ab]
Household income	
a. Less than $30,000/yr	77%
b. $30,000–$49,999	85%
c. $50,000–$74,999	93%[ab]
d. $75,000 +	99%[ab]
Community type	
a. Urban	88%
b. Suburban	87%
c. Rural	83%

Source: Pew Research Center Internet Project Survey, January 9–12, 2014, n=1,006 adults.
Note: Percentages marked with a superscript letter (e.g., a) indicate a statistically significant difference between that row and the row designated by that superscript letter, among categories of each demographic characteristic (e.g., age). The results for race/ethnicity are based off a combined sample from two weekly omnibus surveys, January 9–12 and January 23–26, 2014. The combined total n for these surveys was 2,008; n=1,421 for whites, n=197 for African-Americans, and n=236 for Hispanics.

The demographic groups that are less likely to have access to the internet are the elderly, those in households earning less than $30,000 a year, and those with at most a high school education.[226] On the basis of this information and the fact that our target population is all members of the U.S. labor force, we should be most concerned about survey participation of the elderly because the percentage of

---

225.   Susannah Fox & Lee Rainie, *The Web at 25 in the U.S.*, PEW RES. CTR. (Feb. 27, 2014), http://www.pewinternet.org/2014/02/27/the-web-at-25-in-the-u-s/ [https://perma.cc/KNV8-WUSP].
226.   *See supra* Table 11. According to the Pew Research Center Project Survey in 2014, 57% of respondents at least 65 years old, 77% of respondents in households earning less than $30,000 a year, and 76% of respondents with at most a high school education or less reported using the internet.

those over 65 who use the internet is just 57%.[227] We note, however, that the population of interest in the Pew study of internet penetration is "all adults,"[228] not only *labor force participants* aged 18–75. We surmise that the internet use rate for labor force participants who are aged 65–75 is much higher than 57%.[229] Admittedly, the use rate for those aged 65–75 is probably lower than the 88% rate for those aged 50–64. Nevertheless, at least some evidence shows that the *working* elderly use the internet extensively.[230] Overall, we do not believe this source of selection to be a major concern.

However, the fact that someone must register to receive online surveys to complete our survey—the second source of selection—raises serious questions about our sample's representativeness.[231] One would expect that because online surveys take nontrivial amounts of time to complete and offer small rewards, individuals who (1) have spare time and (2) want or need the rewards offered would be most likely to sign up.[232] To get a sense of the demographic differences between those who register to participate in online surveys and the U.S. labor force, Table 12 compares the panel populations of two of our survey's most represented partners (ClearVoice Research and Sample Strategies) to U.S. *labor force* averages that we calculated using American Community Survey (ACS) data.[233]

---

227. *See supra* Table 11.

228. The Pew study was focused on generating internet penetration or use statistics for the population as a whole, not for labor force participants aged 18–75. For more details on trends in internet access over time, see Andrew Perrin & Maeve Duggan, *Americans' Internet Access: 2000-2015*, PEW RES. CTR. (June 26, 2015), http://www.pewinternet.org/2015/06/26/americans-internet-access-2000-2015/ [http://perma.cc/2K6Y-55JM].

229. For details on internet adoption and usage among the elderly, see Aaron Smith, *Usage and Adoption*, PEW RES. CTR. (Apr. 3, 2014), http://www.pewinternet. org/2014/04/03/usage-and-adoption/ [https://perma.cc/QX5Z-N8VD] (reporting that 74% of those aged 65–69 go online, while 68% of those 70–74 years old go online). *See also* JONATHAN LAZAR ET AL., UNDERSTANDING WEB CREDIBILITY: A SYNTHESIS OF THE RESEARCH LITERATURE 9-10 (2007).

230. *See* Smith, *supra* note 229. While this particular study does not examine employment status explicitly, it does offer evidence that among the elderly those with higher incomes are more likely to be using the internet. *Id.*

231. D. Sunshine Hillygus et al., *Professional Respondents in Nonprobability Online Panels*, *in* ONLINE PANEL RESEARCH, *supra* note 184, at 219.

232. *Id.* at 224; Reg Baker et al., *Research Synthesis: AAPOR Report on Online Panels*, 74 PUB. OPINION Q. 711, 720 (2010); Florian Keusch et al., *Motives for Joining Nonprobability Online Panels and Their Association with Survey Participation Behavior*, *in* ONLINE PANEL RESEARCH, *supra* note 184, at 171, 172.

233. To clarify, the ACS numbers are *not* for the entire ACS population, but only for members of the U.S. labor force, whom we are able to identify using the

Table 12
PANEL PARTNER DEMOGRAPHICS

	ClearVoice	Sample Strategies	U.S. Labor Force
Panel size	902,878	425,413	
Responses in sample	4,949	2,299	
Proportion of sample (%)	42.9	19.9	
Gender (%)			
Male	56.6	41.0	52.9
Female	43.4	59.0	47.1
Education (%)			
< High School	10.0	14.3	9.8
High School graduate	26.8	25.3	26.9
Some College	24.3	21.1	24.9
Associate's degree	7.5	5.2	9.1
Trade School	2.2	.	.
Bachelor's degree	16.3	18.5	19.9
Some postgraduate	2.5	2.0	.
Graduate degree	8.2	10.0	9.5
None of the above	2.2	3.7	.
Age (%)			
13–17	5.4	.	.
18–24	21.9	23.0	15.1
25–34	22.7	27.0	24.1
35–44	26.3	19.0	21.3
45–54	13.0	16.0	21.0
55–64	7.6	10.0	14.9
65–99	3.1	5.0	3.7
Employment (%)			
Employed full-time	35.3	51.0	82.0
Employed part-time	10.9	13.0	13.7
Unemployed	9.7	3.3	4.3
Student full-time	9.9	15.0	.
Student part-time	3.6	.	.
Homemaker	8.8	5.0	.
Self employed	7.7	5.3	.
Permanently unemployed	7.2	.	.
Retired	6.9	3.5	.

Note: This table shows the panel demographics for ClearVoice Surveys and Sample Strategies as they compare to the sample demographics of the American Community Survey. The demographic information comes from the panel books from the respective partners. The numbers in the table are column percentages.

With respect to gender balance, while the U.S. labor force is not far off of being evenly split, ClearVoice and Sample Strategies have survey populations that deviate in opposite directions from this point—with 56.6% male and 41% male, respectively. Some expert observers believe that online survey panel compositions tend to skew toward women;[234] according to Table 12, this is evidently not always true. Education-level demographics are unusually similar across the

---

same sampling frame we use to define our survey sample of interest. *See supra* note 176. Unfortunately, we do not have access to comparable labor force numbers for the panel partner populations, so we instead report statistics for each partner's *entire* panel population. We obtained demographic information on the panel populations from Qualtrics and our panel partners. Data from the ACS is available at *American Community Survey (ACS)*, U.S. CENSUS BUREAU, https://www.census.gov/programs-surveys/acs/ [https://perma.cc/68MP-HM9H] (last visited Apr. 14, 2016).

234.   We elicited this belief about skewed gender balance in survey panels from Qualtrics representatives during conversations in mid-2014.

survey panels and the U.S. labor force.[235] The panel age distributions, however, differ from the ACS labor force numbers.[236] If ClearVoice and Sample Strategies are generally representative of online panels, then survey panels underrepresent some middle-age groups (45–64) and overrepresent some younger ones (18–24).[237]

The employment characteristics of the panel partners' survey populations present the starkest contrast to the U.S. labor force. Only 35.3% of the ClearVoice panel and 51% of the Sample Strategies panel is employed full-time relative to the 82% full-employment rate for the whole of the U.S. labor force.[238] This is unsurprising: Panel partners have a survey population made up of many more part-time employees, students, homemakers, and unemployed.[239] We of course only survey individuals on these panels who are a part of the labor force, however, and thus the numbers from the full panel populations are not directly comparable to Table 12's third column. In fact, if we restrict the panel populations to include only the full-time, part-time, and unemployed rows of Table 12, the full-employment proportion for ClearVoice is 63.1% and for Sample Strategies, 75.8%—much closer to the corresponding U.S. labor force number.[240]

While these demographic comparisons offer at least a sense of who registers to take online surveys, they do little to inform us about *why* people choose to sign up. This matters in practice because when we require a reliable estimate of the incidence of noncompetition agreements, for instance, we must be wary of the possibility that individuals who are more likely to register for online surveys may also be systematically more or less likely (relative to the labor force generally) to have entered into a noncompete agreement.[241] For this

---

235.   *See supra* Table 12.
236.   *See supra* Table 12.
237.   *See supra* Table 12. It is interesting to note that the panels are not consistently undershooting or overshooting ACS population proportions. Rather, different panels appear at times to bookend the population numbers, suggesting that panel partners may distinguish themselves in the marketplace by offering distinctive survey populations.
238.   *See supra* Table 12.
239.   *See supra* Table 12. ClearVoice and Sample Strategies report 10.9% and 13.0% part-time employees, 13.5% and 15.0% students, 8.8% and 5.0% homemakers, and 9.7% and 3.3% (temporarily) unemployed, respectively.
240.   These calculations are $35.3 \div (35.3 + 10.9 + 9.7) = 63.1\%$ and $51 \div (51.0 + 13.0 + 3.3) = 75.8\%$, respectively.
241.   From a research perspective, we have initially focused on accurately estimating the incidence of noncompetition agreements in the U.S. labor force, and so selection dynamics that bias this assessment are our primary concern. However, more generally, researchers relying on survey data must keep in mind their specific

reason, we asked our survey respondents to account explicitly for their interest in completing online surveys.[242] Table 13 lists various reasons individuals reported for why they registered to receive online surveys, cutting the data by income quartile.

Table 13
WHY SIGN UP FOR ONLINE SURVEYS?

Reason	Income Quartile				Total
	0%–25%	25%–50%	50%–75%	75%–100%	
Like rewards	0.58	0.58	0.60	0.60	0.59
Share opinion	0.56	0.57	0.57	0.60	0.58
Want money	0.47	0.43	0.43	0.33	0.40
Learn	0.41	0.39	0.39	0.39	0.39
Fun	0.31	0.29	0.33	0.33	0.32
Need money	0.38	0.33	0.23	0.10	0.23
Game benefits	0.11	0.13	0.10	0.11	0.11
Other	0.02	0.01	0.01	0.01	0.01
Accident	0.00	0.00	0.00	0.01	0.00

Note: This table presents the reasons respondents selected when asked why they signed up to take online surveys in the first place. Respondents were free to provide more than one answer. The numbers in the table are column percentages.

The two most common reasons people report for registering to take online surveys are that (1) they like the rewards they receive for taking them and (2) they like to share their opinion.[243] Forty percent of respondents indicate that they registered because they "want[ed] money"; a similar fraction of respondents indicate registering to learn about new products or ongoing research.[244] Interestingly, the relative importance of the reasons given by respondents does not vary much by income level, with some exceptions: 38% of those in the 0%–25% income quartile report that they signed up because they need the money, for instance, while only 10% of the 75%–100% quartile gave the same reason.[245] If we take these explanations at face value,[246] the sample selection issue asks us to consider whether those who "like rewards" or like to "share their opinion" are differentially

research questions, and ask whether the sample population's motivation to take online surveys may be correlated with the study's outcome of interest.

242. We recognized that while many surveys can only speculate about the potential reasons somebody may or may not be in the sample produced, in designing our noncompete survey, we were at least free to ask people who decided to take the survey why they agreed to take surveys in general and our survey in particular.

243. This aligns with wider findings. *See* Baker et al., *supra* note 232, at 720.

244. *See supra* Table 13. This is surprising to us, as we assumed that most individuals would sign up to receive the monetary benefits that were available.

245. *See supra* Table 13. A smaller but similar trend is observed for those who signed up because they want money.

246. By which we mean not only that the respondents reported truthfully, but also that the list of potential reasons from among which they had to choose is reasonably comprehensive.

likely (relative to those who care less about these things) to agree to a noncompete. Unfortunately, it is not clear to us whether this is true. Overall, our best guess—based on hypotheses and anecdotes, but no rigorous evidence—is that individuals who choose to sign up for online surveys are at least somewhat less likely to be employed in jobs that require noncompetes.[247] If this supposition is true, then our incidence measures would likely be biased downward.

The third selection issue revolves around who receives the survey—i.e., whether the survey is being sent broadly to a random selection of the panel population (or the entire population) or instead to a particular (and potentially skewed) subset of that population. We learned, via conversations with ClearVoice Research and Qualtrics representatives, that panel partners typically estimate the number of invitations they must send in order to reach the target number of respondents and then randomly distribute the survey to members of the panel population.[248] However, if there are binding sample quotas in place,[249] panel partners consider only candidates who will ensure compliance with the quota criteria, sending the survey randomly to these individuals. Thus the only possible selection issue at this stage appears to be selection on the variables underlying the criteria for the quotas. Yet it is critical to recall that we intended this precise form of selection.[250] Accordingly, the data themselves allow researchers to explicitly study and account for this species of selection—through the use of weighting methods or by controlling for the quotas in a regression framework—because the variables (e.g., state, job status, earnings, age, education, etc.) with which we build the quota criteria are objective, observable, and included in our data.

The fourth selection issue is the standard threat of nonresponse bias: The individuals who receive the survey are not required to finish or even start it.[251] If the individuals who enter into noncompete agreements with their employers were systematically more or less likely to finish the survey, estimates of the incidence of noncompetes will be biased.[252] What might cause someone with a noncompete to be more or less likely to start or finish our survey?

---

247.   Our speculation is based on the observed differentials in employment in the online survey-taking population and the population overall. *See supra* Table 12.

248.   We learned these details in phone conversations in mid-2014.

249.   *See supra* Section II.B (discussing our quotas and sample criteria).

250.   See *supra* at 405, for a discussion of our quota criteria.

251.   Baker et al., *supra* note 232, at 728.

252.   TOURANGEAU ET AL., *supra* note 159, at 24. For an example of a study where a variable of interest interacted with survey uptake and completion, see Mick

One important potential source of selection bias is whether the survey was advertised or described as a "noncompete" survey *before* a prospective respondent decided whether to begin the survey. If so, those who were aware of what a noncompete agreement was or who were more or less interested in issues related to noncompetes might have been differentially more or less likely to take the survey. And if these individuals were systematically more likely to have agreed to noncompetes in the past—indeed, noncompete experiences may be at the root of any special interest), estimates from the survey data about noncompete incidence, for example, may be unreliable. Fortunately, none of the survey invitations included the word "noncompete" or something similar anywhere in the invitation,[253] and we know of no evidence that respondents had any indication of the subject matter of the survey before agreeing to begin.[254]

Unfortunately, absence of selection into who opens and begins a survey does not preclude significant selection from occurring prior to the completion of the survey. Specifically, conditional on starting the survey, if respondents who are currently bound by noncompetes were more likely to finish the survey, perhaps because they found the survey more interesting, incidence estimates would in all likelihood be upward biased.[255] To explore the scope of the danger posed by this selection concern, we simply asked respondents why they took our survey. Table 14 displays how survey *finishers* responded to this question by income quartile. Differences across income quartiles are relatively minor. Higher-income individuals—who are more likely to

---

P. Couper et al., *Noncoverage and Nonresponse in an Internet Survey*, 36 Soc. Sci. Res. 131, 146 (2007).

253.   Examples of screen shots of these online invitations are on file with the authors. To illustrate the sort of language the panel partners use in these invitations, consider ClearVoice's framing: "You are about to start a General Opinion Survey, and you can earn $3.00 for your participation."

254.   Another potential source of selection bias relates to whether potential respondents may have declined to initiate the survey because they deemed the compensation for the survey to be too small relative to the estimated 27 minutes it would probably take to complete the survey. Respondents were informed of the median finish time of the survey based on the pilot sample. We do not have a method to account for the extent of this form of selection, although we can point out that 29% of the 50,504 individuals who started the survey and were a part of the sample of interest finished it in the first round.

255.   In our discussion, we focus on the potential selection bias in estimating noncompete incidence, but we remind readers that the threat is more general. If a certain type of respondent is more likely to finish the survey, and that type of respondent is not representative on *any* measure, then it is invalid to draw inferences about how the population fares on that measure.

have signed a noncompete agreement—were more "interested" in the survey than lower-income respondents, but not by a large margin (8 percentage points). Higher-income individuals were also somewhat more likely to have decided to take the survey at least in part to share their experiences with noncompetes, but differences in this statistic across income groups were just as small.

Table 14
WHY TAKE THIS SURVEY?

Reason	Income Quartile				
	0%–25%	25%–50%	50%–75%	75%–100%	Total
Interest	0.48	0.49	0.52	0.56	0.52
Like rewards	0.44	0.43	0.43	0.42	0.43
Want money	0.38	0.36	0.35	0.26	0.33
Enjoy	0.35	0.31	0.31	0.26	0.30
CNC experience	0.18	0.19	0.21	0.26	0.21
Need money	0.30	0.27	0.17	0.08	0.18
Game benefits	0.09	0.12	0.09	0.10	0.10
Pass time	0.07	0.08	0.08	0.08	0.08
Other	0.01	0.02	0.01	0.02	0.01

Note: This table presents the reasons respondents selected when they were asked why they took this particular survey. Respondents were free to provide more than one answer. The numbers in the table are column percentages.

Nevertheless, a significant number of our respondents declared that their desire to "share my experiences with noncompetes" was among their reasons for completing the survey, and this fact raises questions about selection among the finishers. It is unclear to us whether "share my experiences with noncompetes" was interpreted to mean "share my *employment* experiences" or "share my *views* on noncompetes" or even "share *my family member's* experiences with noncompetes,"[256] but if finishers were in fact more likely to have noncompete stories to tell and if having such stories implies more experience with such agreements, care must be taken in drawing inferences from our data, at least when assessing absolute incidence levels, rather than relative differences across groups of finishers. On the other hand, there are selection pressures that cut the other way, as well: Potential respondents may have declined to finish the survey to avoid answering questions about their existing noncompete, either because the questions made them feel uncomfortable for some work-related reason or because doing so might plausibly be interpreted as violating a confidentiality clause.[257]

---

256.   In support of these alternative interpretations, we estimate that 71% of the respondents who selected this response were not currently laboring under a noncompete, and 48% of these individuals had never signed a noncompete.

257.   As a result, it is ultimately unclear how concerned we should be about respondents who select "wanted to share my experiences with noncompetes." Only 20% of respondents selected this reason in explaining why they completed the

To address this selection issue, one partial solution is to probe the robustness of any analyses by eliminating all respondents who indicated an interest in sharing their noncompete experiences in the survey. While this strategy does not address the fact that those with noncompetes may disproportionately fail to finish the survey, it does remove the set of respondents whose answers are likely to artificially inflate our incidence estimates.[258] Some fraction of these respondents would have completed the survey regardless of their experiences, however. Therefore, this tactic is necessarily a conservative strategy to resolve potentially confounding selection, allowing us to interpret the resulting incidence estimates as lower bounds.

Figure 1
RESPONDENT EXPERIENCE TAKING ONLINE SURVEYS
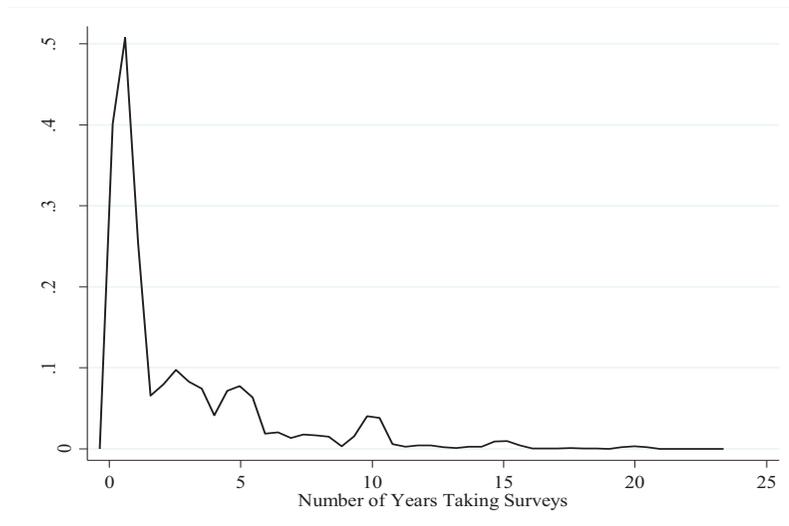
Note: Years Taking Online Surveys is truncated at 25 years. There are 338 respondents who report taking online surveys for more than 25 years. Kernel = epanechnikov, bandwidth = 0.3621.

We also investigated our survey respondents' general online survey-taking frequency and history with the objective of detecting

---

survey, and many of these respondents later indicated that they had never actually signed a noncompete. Still, users of our survey data and consumers of any results that derive from these data ought to keep this selection issue and how it might affect the reliability of any inferences in mind when engaging in interpretation.

258.   Indeed, researchers are free to calculate the incidence of noncompetes after dropping from the sample all individuals who reported that they chose to finish the survey to share their experiences about noncompetes. Importantly, because 71% of those who ticked this option were not currently bound by noncompetes, excluding these individuals only marginally affects our basic incidence results.

any other selection threats.[259] We asked our respondents to report how long (in years) they had been taking online surveys and how frequently they currently take them. Figure 1 above illustrates the distribution of the length of time (in years) respondents report having taken online surveys. The median response is 0.5 years, suggesting that most of the sample is relatively new to the online survey-taking world. Table 15 shows the frequency with which the final sample of respondents takes online surveys. Overall, higher-income individuals in our data tend to respond less frequently to survey invitations than lower-income members of the sample.

Table 15
HOW FREQUENTLY DO YOU TAKE ONLINE SURVEYS?

	Income Quartile				
	0%–25%	**25%–50%**	**50%–75%**	**75%–100%**	**Total**
First online survey	3.91	3.43	2.43	3.36	3.11
Daily	36.97	35.51	31.32	23.16	30.39
A few times per week	37.78	36.81	39.33	40.93	39.07
Once a week	3.72	4.44	5.30	6.66	5.35
A few times per month	10.26	11.56	13.19	16.05	13.38
Once a month	1.91	1.76	2.71	3.08	2.51
A few times per year	3.81	4.73	4.07	5.14	4.53
Less than a few times per year	1.63	1.76	1.65	1.62	1.66
Total	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**

Note: This table shows the frequency with which respondents take online surveys. The numbers in the table are column percentages.

Our discussion to this point has spoken primarily in terms of "overall" selection into or out of the sample, but the role that observable characteristics—such as age, occupation, education, and income—play in sample selection also merits mention. Because we use quotas to constrain the final composition of the sample with respect to many observables,[260] the real selection question is what distinguishes, for example, the college-educated managers who earn $60,000 per year that we *do* observe in our sample from those we do not observe in our sample (either because they are not online, did not sign up to take online surveys, or did not start or finish the survey). In particular, on what other dimensions, if any, do they differ, and what do these potential differences mean for the estimated incidence of noncompetes or other questions related to noncompetes?[261]

---

259. For example, if the sample were dominated by long-time survey takers, one might be worried that they use online surveys as a mode of employment, which may reflect systematic differences in the types of jobs they have had and thus whether they have signed or are currently bound by a noncompete.

260. *See supra* Section II.B. For an explanation of quota sampling, see *Quota Sampling*, LÆRD DISSERTATION, http://dissertation.laerd.com/quota-sampling.php [https://perma.cc/GH5Z-L9QQ] (last visited Apr. 14, 2016).

261. Bethlehem, *supra* note 215, at 183.

For all we know, the answer may have to do with the fact that some individuals have a network of family, friends, or colleagues who participate in online surveys regularly and that these individuals are more likely to register to take online surveys themselves as a result of their association with this network.[262] Unfortunately, we have no rigorous evidence to speak to whether this (or any other) theory is accurate. However, to the extent that contact with these networks is independent of noncompete signing status or experience *conditional* on factors such as occupation, income, and education, selection bias concerns on the whole should be minimal.[263]

So far in this Section, we have addressed four potential sources of sample selection that might affect the content of our data: the nonrepresentativeness of our respondents (1) who are online, (2) who register to take online surveys, (3) who are invited to take our online survey, and (4) who start *and complete* the survey. Although we are relatively unconcerned about the first and third forms of selection, the second and fourth forms present as more serious challenges. We consider all four and offer reasons and evidence suggesting that none is, in practice, likely to be particularly problematic. Still, because our measures are imprecise or inaccurate and our methods and arguments have limitations, we can never entirely discount selection. Moreover, although selection bias plagues all surveys, the disease is particularly virulent in a convenience survey of the sort we employed. For these reasons, we ought, when possible, to limit the extent of any potential bias.[264] One important approach to remediating selection effects in survey data is reweighting the sample.[265]

2. *Weighting Techniques*

Due to falling response rates for more traditional survey modes (random-digit dialing, mail-in surveys, etc.), scholars have developed various reweighting approaches to address selection in convenience samples. When available and appropriate, reweighting methods free

---

262.   In retrospect, we wish we would have asked respondents the question, "How did you come to sign up to receive online surveys?"

263.   BETHLEHEM & BIFFIGNANDI, *supra* note 216, at 281-94.

264.   See the discussion above with regard to methods for identifying such biases. We propose in some instances dropping certain suspect observations (e.g., those respondents who may have completed the survey precisely because they are a party to a noncompete contract) to provide a bound on any bias.

265.   *See* Paul P. Biemer & Sharon L. Christ, *Weighting Survey Data*, *in* INTERNATIONAL HANDBOOK OF SURVEY METHODOLOGY 317 (Edith D. de Leeuw et al. eds., 2008).

researchers to exploit more cost-effective data-collection strategies, including the use of online surveys.[266] In what follows, we introduce a few standard approaches; compare their properties, assumptions, and application to our survey data; and choose the one best suited for understanding noncompete activity in the U.S. labor force.

    We evaluate three weighting methods: (1) post-stratification, (2) iterative proportional fitting (raking), and (3) inverse propensity score weighting. We sketch each of these methodologies and apply them to our survey data.[267] Each method requires demographic data from the population of interest—i.e., representative data from the population that we hope the reweighted survey data will closely match.[268] We use data from the three-year American Community Survey (ACS) from 2014,[269] which contains information on gender, age, education, race, industry, occupation, annual weeks worked, state, employee class, marital status, and whether the individual is enrolled as a student. We restrict the ACS population we examine to match our population of interest: labor force participants aged 18–75 who work in the private sector or in a public healthcare system.

    Post-stratification reweights the survey data, when possible, to match the joint distribution of key variables in the population.[270] The method proceeds by placing each survey data observation into a cell defined by the values of key covariates, such as gender, age-group, income-range, and education indicators.[271] The researcher then adds up the number of observations in each cell, and calculates the proportion of the survey data that falls within each cell—s_c. Next, the researcher tabulates corresponding proportions for each cell in the population data—p_c. Post-stratification weights each cell c by the ratio of these two proportions: $w_c = p_c / s_c$.[272]

---

266.   *See generally* HANDBOOK OF SURVEY METHODOLOGY FOR THE SOCIAL SCIENCES, *supra* note 213.

267.   An interested reader may benefit from referencing Graham Kalton & Ismael Flores-Cervantes, *Weighting Methods*, 19 J. OFFICIAL STATISTICS 81 (2003), for a recent discussion.

268.   Bethlehem, *supra* note 215, at 180.

269.   To learn more about content, structure, and history of the ACS, see *supra* notes 169 & 233.

270.   The joint distribution of, say, occupation and industry reflects the proportion of individuals in each occupation across all industries (e.g., managers in manufacturing, in sales, etc.) for every occupation. BETHLEHEM & BIFFIGNANDI, *supra* note 216, at 335-36.

271.   *See* Kalton & Flores-Cervantes, *supra* note 267, at 92-94.

272.   For example, if 25% of the sample is male college graduates, while the proportion of the population that is male college graduates is 50%, then the weight applied to the male college graduates cell in the sample is 50/25 = 2.

One advantage of using post-stratification is that it reweights the sample to exactly match the joint distribution observed in the population of interest.[273] A critical downside of post-stratification, however, is the "curse of dimensionality," which can arise when one uses a large number of variables in conducting the post-stratification reweighting. Specifically, absent sufficient data, certain cells in the survey data can wind up being sparsely populated or not populated at all. The practical consequence is unstable (and therefore unreliable) estimates as a cell's weight might be laid on only a few or even a single—potentially unrepresentative—individual.[274]

Another disadvantage of post-stratification is the possibility of measurement error.[275] Post-stratification works by matching the *joint* distribution of the variables of interest, and so an erroneous entry with respect to *any* of these variables can result in the observation being placed in the wrong cell. Such measurement error may be especially likely with questions that impose a significant cognitive burden on respondents, such as requests for occupation and industry codes, which may require that each survey taker read a long list of descriptions before selecting the most appropriate match.[276]

To calculate post-stratification weights, we experimented with four sets of post-stratifying variables:

1. Post-Stratification I uses gender, age (three categories), annual compensation (three categories), industry, and occupation as post-stratifying variables, which equates to 6,248 cells. Within our sample, 16 individuals had characteristics that did not match a single individual in the ACS data, meaning they were

---

273.   *See* Kalton & Flores-Cervantes, *supra* note 267, at 84-86. This feature of post-stratification is important because, as a result, within the post-stratifying variables, sample estimates are unbiased predictors of the corresponding population values. For example, post-stratifying survey data on occupation and industry allows one to accurately predict not only the incidence of noncompetes across occupations (e.g., between managers and salespeople) and across industries (e.g., between manufacturing and sales), but also within occupations across industries (e.g., between managers who work in manufacturing and managers who work in sales).

274.   Bethlehem & Biffignandi, *supra* note 216, at 347-48.

275.   *See* Joop J. Hox, *Accommodating Measurement Errors*, *in* Int'l Handbook of Survey Methodology, *supra* note 265, at 387.

276.   Under these circumstances, asking redundant questions and engaging in painstaking data cleaning may be necessary before it is wise to use such variables in post-stratification weighting procedures. With respect to our data, we have sought to recode our occupation and industry data from text-entered answers to occupation and industry questions (comparing them to the codes entered) to match as narrowly defined as possible NAICS and SOC codes. See *supra* at 421-22, for details on how we performed this procedure with our data.

not placed in a cell. Relatedly, sample respondents did not fall into all 6,248 cells; indeed, 4,343 of the 6,248 did not contain a single sample respondent. These unfilled cells represented 13.6% of the population.

2. Post-Stratification II includes as post-stratifying variables just annual compensation (three categories), occupation, and industry, which generated 1,202 cells. There were three individuals in the sample who were not placed in a cell because of their idiosyncratic characteristics. Of the 1,202 cells, 576 had no sample respondents, although these unfilled cells represented just 3.2% of the population.

3. Post-Stratification III omits industry (on account of the fact that including industry and occupation alone generate over 400 cells), but includes other key variables: gender, age (three categories), annual compensation (three categories), education (three categories), and occupation, which created 1,188 cells. Every sample respondent fell into one of the cells. Of the 1,188 cells, 325 cells were unfilled by sample respondents. These 325 cells represented 2% of the population.

4. Post-Stratification IV includes as post-stratifying variables just occupation and industry, resulting in 413 cells. Each sample respondent fit into one of the cells. Of these 413 cells, 124 of them (representing 1.1% of the population) did not contain a single sample respondent.

Iterative proportional fitting (also known as raking) ignores the joint distribution of selected variables in the population and instead focuses on reweighting the sample so that the *marginal* probability densities of variables of interest match those of the population.[277] Because raking does not seek to match the joint distribution of the sample to the population, the curse of dimensionality does not arise—the number of cells does not increase exponentially as you add variables.[278] Raking gets its name from the way in which the iterative procedure works. Essentially, the procedure cycles through the variables matching each marginal survey density to its marginal population density, iterating until each variable is appropriately weighted.[279] The primary benefit of raking is that it is possible to

---

277. *See* Kalton & Flores-Cervantes, *supra* note 267, at 86-87.

278. *Id.*

279. To describe how raking works more intuitively, imagine taking the joint distribution of two indicator variables in the sample and arranging them in a two-by-two table. Raking works by first moving down the table, weighting each observation to match the marginal distribution of that variable in the population. Assuming the

match the marginal distribution of many variables;[280] the downside of this weighting approach that it ignores joint distribution information from the population.[281]

   Inverse probability weighting proceeds by estimating the conditional probability that a particular individual would have been a survey respondent given the nature of the population.[282] Thus, if the estimated propensity score for an individual with a particular set of covariates is 0.10,[283] we estimate that for every 10 individuals in the population with those covariate values, one of them should be a respondent in the sample.[284] Conversely, one individual in the sample

---

marginal distribution of the *other* variable in the sample does not match the marginal distribution in the population, the next step involves raking in the other direction—across the table—to match the other variable's marginal density to its population marginal density. This second step may result in the first variable's marginal density departing from the population density, so the process iterates, and raking continues until both variables match their marginal population densities.

   280.   *See* Kalton & Flores-Cervantes, *supra* note 267, at 86-87; Little, *supra* note 221, at 1009-10.

   281.   Bethlehem & Callegaro, *supra* note 224, at 269. We chose to rake our sample by gender, age (deciles), annual compensation (20 quantiles), industry, occupation, education (nine categories), an indicator for being in school, employee class (for-profit, nonprofit), an indicator for being unemployed, indicators for working more than 40 weeks per year and for working more than 40 hours per week, and state. We implemented raking with the -ipfweight- command in Stata. *See* Michael Bergmann, *IPFWEIGHT: Strata Module to Create Adjustment Weights for Surveys*, IDEAS, http://fmwww.bc.edu/repec/bocode/i/ipfweight.ado [https://perma.cc/W5Y7-TH6R] (last visited Apr. 14, 2016). We allowed for 1,000 iterations and set the maximum weight at 5. See David Izrael et al., *Extreme Survey Weight Adjustment as a Component of Sample Balancing (a.k.a. Raking)*, SAS GLOBAL F., 2009, at 2, commenting with respect to maximum weights that:

   There are no strict rules or procedures either to define extreme weights or for trimming the weights. Different surveys follow different rules and therefore in practice there are several procedures to trim extreme weights. Some common procedures for trimming large weights include: 1) identifying any weight bigger than 4 or 5 times the mean weight as an outlier weight and trimming that weight by making it equal to the limit, 2) identifying any weight bigger than the median weight plus 5 or 6 times the inter-quartile range of the weights and trimming the weight by making equal to the limit, and 3) truncating weights above a certain percentile like 95 or 99 in the distribution of weights.

For another discussion, see MATTHEW DEBELL ET AL., USER'S GUIDE TO THE ADVANCE RELEASE OF THE 2008–2009 ANES PANEL STUDY 31 (2009), http://www.electionstudies.org/studypages/2008_2009panel/anes2008_2009panel_UserGuide_AdvanceRelease.pdf [https://perma.cc/9RAC-BAJS] ("Trim weights larger than 5.").

   282.   Stephanie Steinmetz et al., *Improving Web Survey Quality*, *in* ONLINE PANEL RESEARCH, *supra* note 184, at 273, 280-81.

   283.   Bethlehem & Callegaro, *supra* note 224, at 270.

   284.   *Id.*

represents 10 people in the actual population.[285] Weights are thus attributed to individual respondents in the form $1/\hat{p}$ where $\hat{p}$ is the estimated probability of sample participation, typically predicted using a logit or probit model.[286] To implement this weighting method, we estimated a simple logit model on a set of demographically significant independent variables.[287]

Table 16
WEIGHTING SUMMARY STATISTICS AND CORRELATIONS

Panel A: Weighting Summary Statistics						
Weighting Method	**Obs.**	**Mean**	**Std. Dev.**	**Min**	**Max**	**Max/Min**
Raking	10,226	1.00	1.30	0.00	5.00	1,945
Post-Stratification I	10,210	0.87	0.95	0.00	20.25	14,087
Post-Stratification II	10,223	0.97	0.93	0.02	21.03	1,245
Post-Stratification III	10,226	0.98	1.25	0.03	65.00	2,323
Post-Stratification IV	10,226	0.99	0.78	0.02	9.20	461
IPW I	10,226	10,788	20,660	238	1,042,315	4,373
IPW II	10,226	10,787	20,660	237	1,042,314	4,391

Panel B: Correlation Between Weights							
	Raking	**Post-Strat. I**	**Post-Strat. II**	**Post-Strat. III**	**Post-Strat. IV**	**IPW I**	**IPW II**
Raking	1.00						
Post-Stratification I	0.38	1.00					
Post-Stratification II	0.43	0.65	1.00				
Post-Stratification III	0.42	0.61	0.47	1.00			
Post-Stratification IV	0.46	0.56	0.78	0.42	1.00		
IPW I	0.53	0.23	0.28	0.29	0.29	1.00	
IPW II	0.53	0.23	0.28	0.29	0.29	1.00	1.00

Note: Panel A shows summary statistics for seven weighting schemes: raking, four types of post-stratification, and two types of inverse probability weights. Raking uses gender, age (deciles), annual compensation (20 quantiles), industry, occupation, education (9 categories), an indicator for being in school, employee class (for-profit, nonprofit), an indicator for being unemployed, indicators for working more than 40 weeks per year and for working more than 40 hours per week, and state. We allowed for 1,000 iterations and set the maximum weight at 5. Post-Strat. I includes gender, age (3 categories), annual compensation (3 categories), industry, and occupation (6,248 cells). Post-Strat. II includes annual compensation (3 categories), occupation, and industry (1,202 cells). Post-Strat. III includes gender, age (3 categories), annual compensation (3 categories), education (3 categories), and occupation (1,188 cells). Post-Strat. IV includes occupation and industry (413 cells). IPW I is constructed as the inverse of the predicted probability of participating in the sample $(1/\hat{p})$ from a logit model in which an indicator for participating in the survey is regressed on gender, age (deciles), annual compensation (20 quantiles), industry, occupation, education (9 categories), an indicator for being in school, employee class (for-profit, nonprofit), an indicator for being unemployed, indicators for working more than 40 weeks per year and for working more than 40 hours per week, state, race, and an indicator for being married. IPW II is obtained by calculating $(1-\hat{p})/\hat{p}$. Panel B shows the correlation matrix of the weighting variables.

---

285. *Id.*

286. *Id.* One can improve the stability of the sample by further multiplying the weight by $1 - \hat{p}$, which increases the weight given to those respondents likely to be in the sample and decreases the weight given to those unlikely to be in the sample. *See* Peter C. Austin, *An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies*, 46 MULTIVARIATE BEHAV. RES. 399, 409 (2011) (citing Stephen L. Morgan & Jennifer J. Todd, *A Diagnostic Routine for the Detection of Consequential Heterogeneity of Causal Effects*, 38 SOC. METHODOLOGY 231 (2008)).

287. Specifically, we included gender, age (deciles), annual compensation (20 quantiles), industry, occupation, education (nine categories), an indicator for being in school, employee class (for-profit, nonprofit), an indicator for being unemployed, indicators for working more than 40 weeks per year and for working more than 40 hours per week, state, race, and an indicator for being married.

Table 16 offers summary statistics for our weighting variables and a correlation matrix.[288] In the abstract, each weighting approach has its own merits, but in practice, researchers must choose among them.[289] To select our baseline weighting approach, we analyzed the ability of each set of weights to match the distributions of important variables in our survey data to actual ACS population distributions. For each variable, we calculated the absolute value of the difference between the proportion in the cell given by the weighting method and the proportion in the cell in the population. We then summed these absolute values across all cells for a given weighting scheme. The result of this sum is the proportion of "misplaced" individuals in the weighted sample relative to the population. We report the results of this exercise across a number of variables in Table 17 below.

Table 17
WEIGHTING METHODS AND POPULATION DISTRIBUTIONS

Variable	Raking	Post-Strat. I	Post-Strat II	Post-Strat. III	Post-Strat. IV	IPW I	IPW II
Education (9 categories)							
Percent misallocated	12.99	37.46	39.30	31.54	41.16	7.25	7.25
Per-cell misallocation average	1.44	4.16	4.37	3.50	4.57	0.81	0.81
Income (20 categories)							
Percent misallocated	3.92	26.68	27.76	26.65	29.95	8.56	8.56
Per-cell misallocation average	0.20	1.33	1.39	1.33	1.50	0.43	0.43
Age (10 categories)							
Percent misallocated	0.00	9.28	14.64	8.20	15.78	7.78	7.78
Per-cell misallocation average	0.00	0.93	1.46	0.82	1.58	0.78	0.78
Occupation (22 categories)							
Percent misallocated	0.28	10.44	2.61	2.06	1.12	9.30	9.30
Per-cell misallocation average	0.01	0.47	0.12	0.09	0.05	0.42	0.42
Industry (19 categories)							
Percent misallocated	0.05	8.29	2.15	8.53	1.04	9.88	9.89
Per-cell misallocation average	0.00	0.44	0.11	0.45	0.05	0.52	0.52
State (51 categories)							
Percent misallocated	0.00	23.53	22.43	23.38	20.75	9.66	9.66
Per-cell misallocation average	0.00	0.46	0.44	0.46	0.41	0.19	0.19
Occupation by Industry (289 categories)							
Percent misallocated	20.42	17.13	3.87	22.78	0.00	25.79	25.79
Per-cell misallocation average	0.07	0.06	0.01	0.08	0.00	0.09	0.09

Note: This table compares the ability of seven weighting schemes to match sample distributions of key variables to population distributions. For each variable of interest and for each weighting method, we aggregated across the distribution to identify the total percentage of individuals who are misallocated by the weighting method when compared to the population at large. The per-cell misallocation average is the total percentage that is misallocated across all the cells divided by the total number of cells.

---

288. Importantly, inverse probability weights are frequency weights—i.e., the number of individuals in the population represented by each respondent—while raking and post-stratification weights are cell or analytic weights that instead reflect the importance of an individual's cell relative to the population. The first set of post-stratification weights are the most unstable, which can be attributed to the sparseness of populated cells in that weighting scheme.

289. We have examined the robustness of our work to different weighting schemes, and have found no significant differences in the substance of the results.

First, we consider how well the weighting methods performed in matching the population's educational achievement distribution.[290] Post-stratification weights performed relatively poorly, misallocating between 32% and 41% of the population. Inverse probability weights scored best on this measure, misallocating 7% of individuals. Raking improperly assigned 13% of individuals, although this appears to be a consequence of the restriction we imposed that groups not receive weights exceeding 5.[291] By contrast, inverse probability weighting did better because its weights were unconstrained. Furthermore, the bulk of raking's misallocation resulted from the weight constraint limiting the size of the group that did not complete high school, placing weight instead on the high school graduate groups.[292]

With respect to other key variables—income, age, occupation, industry, and state—raking weights matched the survey and ACS distributions almost perfectly, while post-stratification fared poorly and inverse probability weighting performed only somewhat better. For instance, with our measure of income, post-stratification weights misallocated approximately 30% of the data, and inverse probability weighting misallocated 7.8%. Raking only misallocated 3.9%. Post-stratification's poor performance is no surprise given that we were able to specify relatively few post-stratification variables.

Raking works by matching marginal population distributions, and so one important test is whether raking also aligns the sample's and the population's joint occupation-industry distributions.[293] We created the proportions for each occupation-industry cell using two-digit NAICS and SOC codes.[294] We found that raking misallocated

---

290.   We are keen on matching the education distribution of the population because, as with a labor force participant's occupation and industry, we surmise that education is likely to be strongly related to the use of noncompetes.

291.   *See supra* note 281.

292.   We do not report these results, but they are available upon request.

293.   It is important that our sample closely matches the joint occupation-industry distribution because we expect noncompete use will depend not just on occupation and industry separately, but on the interaction between the two.

294.   *See Introduction to NAICS*, U.S. CENSUS BUREAU, http://www. census.gov/eos/www/naics/ [https://perma.cc/YT3A-GCJA] (last modified Mar. 4, 2016) ("The North American Industry Classification System (NAICS) is the standard used by Federal statistical agencies in classifying business establishments for the purpose of collecting, analyzing, and publishing statistical data related to the U.S. business economy."); *Standard Occupational Classification*, U.S. BUREAU OF LAB. STAT., http://www.bls.gov/soc/ [https://perma.cc/JRC8-TBEE] (last visited Apr. 14, 2016) ("The 2010 Standard Occupational Classification (SOC) system is used by Federal statistical agencies to classify workers into occupational categories for the purpose of collecting, calculating, or disseminating data.").

20.4% of the data, while inverse probability weighting misallocated even more—25.8%. Among the post-stratification choices, the ones that use occupation by industry performed noticeably better: Post-Stratification IV, for example, which uses only occupation and industry as post-stratifying variables, matched the joint distribution perfectly. Post-Stratification II performed almost as well, which is not unexpected because the only additional post-stratifying variable is income. Post-Stratification III performed poorly because it does not post-stratify on industry, and Post-Stratification I failed for the straightforward reason that it resulted in too many empty cells.

While these approaches for evaluating the performance of our weighting schemes are ad hoc,[295] we conclude for purposes of the analysis we present in Part III that reweighting using our raking approach makes the most sense because of raking's ability to match the marginal distributions while performing adequately in matching the ASC population's joint occupation by industry distribution.[296] We acknowledge that, for purposes of demonstrating how they work in practice, we assess just a few reweighting schemes in these pages.[297] We also admit that we chose to focus on variables that are relevant to our particular research aims. Given our sampling approach and our goal of understanding the role noncompetes play in the workplace, the ability of a reweighting scheme to match (as closely as possible) our survey sample to the joint occupation-industry distribution of the population seems to us particularly important,[298] but researchers with other aspirations may understandably emphasize representativeness along other dimensions.[299]

---

295.   A more sophisticated method might proceed by incorporating the size of the differences across variables of interest under each weighting method and the population of interest, and then combine those differences into overall estimates of each method's viability.

296.   We hasten to add that our incidence estimates are qualitatively similar regardless of which weighting approach we employ.

297.   More complicated weighting methods exist, including the method used by Axel Börsch-Supan, Detlev Elsner, Heino Faßbender, Rainer Kiefer, Daniel McFadden, and Joachim Winter. *See* Axel Börsch-Supan et al., How to Make Internet Surveys Representative: A Case Study of a Two-Step Weighting Procedure 3 (Mar. 30, 2004) (unpublished manuscript) (on file with the Michigan State Law Review), http://www.mea.mpisoc.mpg.de/uploads/user_mea_discussionpapers/loil50ozz 320r55b_pd1_040330%20geschuetzt.pdf [https://perma.cc/PB97-GY4F] (using "an explicit behavioral model of Internet access and survey participation decisions" to construct weights).

298.   *See supra* notes 281, 285.

299.   It is worth making explicit that the particular weighting scheme one uses may influence the extent to which the differences in our samples—i.e., the

Sample selection concerns are pervasive in empirical work.[300] Researchers must recognize them where they exist, address them when possible, and in every event, interpret their findings in light of them. In this Section, we outlined general selection issues with using our survey data to estimate the overall incidence of noncompetes, and we have suggested both ways to test the robustness of any findings to these selection issues and ways to mitigate selection's likely effects on some dimensions by reweighting.

We determined to conduct our survey using an online platform because, in our view, the benefits outweighed the costs. Selection concerns may in fact be less serious for our survey and the questions motivating it than in other important contexts.[301] Moreover, in some circumstances, selection effects can be "signed"—meaning one can at least identify the direction of any selection bias—which may permit us to establish an upper or lower bound on the answer to any particular question.[302] Finally, reweighting techniques are available to

---

different ways of cleaning and refining the data—matter in the final analysis. For example, post-stratifying on occupation and income will result in the grocery bagger who earns $2 million receiving zero weight, while raking will accord him positive though small weight. Which of these weighting approaches is more suitable turns in part on our data cleaning and refinement efforts. If the survey data for the grocery bagger are legitimate aside from his income, then post-stratification will award him too little weight, and raking would be an improvement. On the other hand, if the rest of his data are inaccurate, then giving him zero weight is equivalent to dropping him from the sample, which would appear to be the better choice. We have done our best to clean the data, so individuals who remain through the fourth round of preliminary cleaning are associated with accurate data. We also believe the imputation methods we have used are appropriate and that they will contribute to more accurate weights regardless of the weighting scheme a researcher chooses.

300. James J. Heckman, *Sample Selection Bias as a Specification Error*, 47 ECONOMETRICA 153, 153 (1979) ("This paper discusses the bias that results from using nonrandomly selected samples to estimate behavioral relationships as an ordinary specification bias that arises because of a missing data problem."). For an instructional resource that seeks to develop the ability to identify and navigate selection concerns, see Denise Dickins et al., *The Importance of Sample Selection: An Instructional Resource Using U.S. Presidential Elections*, 31 J. ACCT. EDUC. 68, 75 (2013) ("Sampling is a powerful, efficient auditing tool. However, it is susceptible to the possibility that conclusions drawn from a sample's results may not be representative of a population's actual characteristics (nature, value, outcome, etc.). To reduce this sampling risk, it is important that steps be taken to fully understand the characteristics of the population to be sampled, to use the appropriate sampling unit, and to mitigate the likelihood of biased results.").

301. Although, as we reiterate here, researchers should whenever possible explicitly ask questions to understand the extent of any such selection.

302. For example, if individuals who sign noncompetes (typically assumed to be highly educated and well-compensated) are less likely all else equal to be

reduce or eliminate the effects of at least some sources of selection bias. Nevertheless, care and caution remain important. Fortunately, for the empirical result presented in this Article, we believe there is even less reason to be concerned about selection than there may be more generally. In Part III of this Article, we compare the incidence of noncompete agreements across different noncompete enforcement regimes. Selection is likely unable to explain the geographic pattern we identify because differences in the scope or extent of noncompete enforcement seem very likely to be orthogonal to which individuals completed our survey.

F. Multiple Imputation

Multiple imputation creates multiple completed datasets (each dataset being one imputation) from a single dataset with missing data by "filling in" missing data using information from respondents for whom the relevant data are not missing. In the context of our survey, there are three types of "missing" values in our sample of completed surveys: (1) missing values that result from a survey question being added after the respondent had completed the survey; (2) "missing" values that are entries flagged as unreasonable, as we describe above in the Section on data refinement; and (3) "missing" values that arise from the inability of respondents to recall the answer to a question—e.g., whether they have agreed to a contract. The first two categories of missing observations are typical problems, but the third category is special to our context and so we describe our process for imputing these data in detail, focusing in particular on those data that relate to the use of covenants not to compete.

It is perhaps unremarkable that many labor force participants appear to be unaware of what a noncompete is or, for that matter, unsure of whether they have signed one.[303] This fact is in line with research and endless anecdotes indicating that many people do not carefully read contractual language (if they read it at all),[304] or if they

---

taking online surveys, then our survey will underestimate the incidence of noncompetes in the population.

303.   *See, e.g.*, Kim, *supra* note 83, at 106; *see also* Jesse Rudy, *What They Don't Know Won't Hurt Them: Defending Employment-at-Will in Light of Findings that Employees Believe They Possess Just Cause Protection*, 23 BERKELEY J. EMP. & LAB. L. 307, 310-11 (2002) ("Empirical evidence on the subject strongly suggests that employees do not understand the at-will employment term and its application.").

304.   *See* Kim, *supra* note 83, at 110-11; Rudy, *supra* note 303, at 340 ("[T]ry to recall the last time you read all of the disclaimers and warranty information when you purchased a kitchen appliance . . . . Chances are, there is no

do read it carefully, they fail to understand the significance of many of its potentially important provisions.[305] To be sure, whether an employment contract contains a noncompete clause will be salient for certain categories of employees with certain skills and in certain industries.[306] But in general, across almost every class of worker, at least a few individuals are simply unsure about whether they are currently subject to a noncompete (or would prefer not to say). For our purposes, this raises the central question of how best to account for—how to code and analyze—those individuals who indicate that they "don't know" or that they are "unsure" about whether they are a party to a noncompetition agreement.

In our survey, we ask our respondents a number of questions about their experience with and knowledge of noncompetes,[307] after specifically defining what a noncompetition agreement is, clarifying what it is not, and providing some helpful information about when a respondent might have been asked to sign one.[308] Our incidence analysis in Part III focuses on the current status of our respondents, which involves the answer to the following question: "Did you sign a non-competition agreement with your current employer?" Available

last time, because if you are like most consumers you have never read this information.").

305.　*See* Kim, *supra* note 83, at 151; Rudy, *supra* note 303, at 340 ("Gaining a working understanding of legal rules is costly for parties."). We explore this possibility in our survey: "Have you ever unknowingly signed a noncompete, only to realize later that you had in fact signed one?"

306.　*See* Kim, *supra* note 83, at 144 tbl. 5 (finding that 46.9% of those without a high school diploma gave correct responses to questions involving the lawfulness of various types of discharges, while 61.8% of those with an advanced degree provided correct responses).

307.　Examples of these noncompete-related questions from our 2014 survey include without limitation the following: "Have you ever heard of a noncompete [before this survey defined it for you]?"; "From where or from whom did you first learn about noncompetes?"; "At what age did you first learn about noncompetes?"; "Have you *ever* signed a non-competition agreement with *any* employer?"; "How sure are you that you have or have not signed a non-competition agreement at some point in your life?"

308.　"A non-competition agreement is a contract between an employer and an employee that *prevents the employee from joining or creating a competing company* for a period of time after the employee leaves his or her current employer." Later, the survey offers additional information to the respondent: "It might be helpful to know that non-competition agreements can be signed: (a) as stand-alone contracts related to a job; (b) as part of policy manuals or employee handbooks you acknowledged receiving; (c) as part of the 'fine print' in any employment material you have every signed at work, such as a job application form; (d) along with other contracts such as non-disclosure agreements, confidentiality agreements, non-solicitation, and non-poaching agreements."

answer choices are:[309] "Yes," "No," "Cannot remember," and "Do not want to say."[310] We interpret these answers as "yes," "no," and the last two as "maybe." Methodologically, the critical question is how should we treat the respondents whose answer is a "maybe"?

We assume that at least a fraction of respondents who answer "maybe" were actually bound by a noncompetition agreement when we surveyed them. If our goal is to understand how common these agreements are among labor force participants, we need a defensible method for determining the percentage of respondents who *would have* answered such an incidence question "yes" if they were fully informed and willing to answer truthfully. We anticipate that this is likely to be a general problem with future survey research about contracting behavior.[311] Neither missing data nor respondents lacking the knowledge to answer survey questions is a new problem. Still, we thought it might be of some value to explain how we address this problem when the variable at issue—current noncompete status—is essential to the analysis. We pursued two distinct strategies to account for the "maybe" data.[312]

First, we posed and answered a slightly different question. Rather than measure what percentage of the labor force is currently bound by a noncompetition agreement, we calculated the minimum and maximum bounds between which the true incidence percentage most likely falls. This approach allows a more nuanced answer to the incidence question: for example, "at least X% and at most Y% of employees in this category are currently bound by a noncompete." The advantages of calculating, presenting, and interpreting incidence numbers in this way is that, depending on the specific question, it is

---

309.   Other questions we ask about noncompetition agreements sometimes involve differently phrased answer options, depending on context.

310.   We decided to include the last answer option because we anticipated that some respondents might have believed their confidentiality agreements with their employer or even just their conception of employee loyalty should preclude their answering arguably sensitive questions about their private contractual arrangements with their current employer.

311.   *See* Marx, *supra* note 63, at 701 ("For six of the dyads, interviewees could not remember whether the employer had included a noncompete in the employment contract.").

312.   Another possibility is just to discard the "maybe" cases, but the listwise deletion of observations can be the medicine that kills the patient. *See* Gary King et al., *Analyzing Incomplete Political Science Data: An Alternative Algorithm for Multiple Imputation*, 95 Am. Pol. Sci. Rev. 49, 49, 51-52 ("Listwise deletion discards one-third of cases on average, which deletes both the few nonresponses and the many responses in those cases. The result is a loss of valuable information at best and severe selection bias at worst.").

intrinsically conservative.[313] We make no implicit assumptions about respondents who answered the question "maybe"; we explicitly and simultaneously consider both extreme possibilities. We incorporate both possibilities into the results we present below.

Second, to extract more information from our data, we also used multiple imputation methods to estimate the true status for respondents who answered "maybe."[314] "Multiple imputation . . . replaces each missing or deficient value with two or more acceptable values representing a distribution of possibilities."[315] As this passage hints, the challenge in imputing missing values is accounting for the uncertainty that necessarily surrounds predicted values. If we were to ignore this concern, imputation would be straightforward: We would just use what we already know to make our best guess. This might equate to calculating the mean of the variable in question from those observations in which we do observe a value.[316] A more sophisticated approach would compare observations with missing values to those with known outcomes, and would then estimate the most likely outcome using these comparisons.[317] In other words, we would replace missing values with the average value for *similar* cases—i.e., conditional on observed values of other variables.[318]

---

313. In fact, one could describe this route as the most conservative approach because it lets the data speak entirely for themselves.

314. We discuss only one example of multiple imputation here, but the basic ideas are applicable to other situations in which the researcher is unclear how the respondent should have answered in an ideal world in which respondents answer every question, information is complete, and memories are reliable.

315. DONALD B. RUBIN, MULTIPLE IMPUTATION FOR NONRESPONSE IN SURVEYS 2 (1987); *see also generally* HANDBOOK OF STATISTICAL MODELING FOR THE SOCIAL AND BEHAVIORAL SCIENCES (Gerhard Arminger et al. eds., 1995).

316. *See* JEFFREY C. WAYMAN, MULTIPLE IMPUTATION FOR MISSING DATA: WHAT IS IT AND HOW CAN I USE IT? 2 (presented at the 2003 Annual Meeting of the Am. Educ. Research Assoc., Chi., Ill.), http://www.csos.jhu.edu/contact/staff/jwayman_pub/wayman_multimp_aera2003.pdf [https://perma.cc/BET2-PS9B] ("Some of the most popular missing data methods involve ad-hoc deletion or replacement of missing data.").

317. Put differently, we would use the data we observe for all respondents to model the observed outcome, in this case, the answers "yes" and "no." Specifically, using a subset of the data—the set of respondents who answered "yes" or "no"—we might run a regression to estimate the relationships between observables (e.g., age, education, gender, industry, etc.) and the outcome in question. With these estimates in hand, we would then use them to predict the outcome (either "yes" or "no") for those who answered "maybe" using the same observables.

318. *See* Stephens & Unayama, *supra* note 203, at 10 ("A number of straightforward approaches are available when the non-reporting of values follows the selection on observables assumption.").

We are interested in testing hypotheses and making inferences, however, which means we must not only estimate the most likely value of noncompete incidence, but also how certain we are about that value—i.e., the estimate's reliability.[319] Multiple imputation methods accomplish this by building many plausible complete datasets in which we explicitly incorporate randomness (making them different, not mere copies), conducting the analysis on each of them, and then combining the results to calculate final estimates (including estimates of variance in the form of standard errors).[320] For purposes of this example, we predicted whether individuals in the "maybe" category had signed a noncompete 25 times, building 25 complete, but different, datasets, based on the characteristics of those who answered "yes" and "no."[321] This allows us to estimate noncompete incidence in a reliable way in Part III, taking into account uncertainty surrounding who in the "maybe" group is bound by a noncompete.

We imputed noncompete signing status, income, and all other missing, flagged, or "maybe" variables in one step.[322] We included as

319.   Yang C. Yuan, Multiple Imputation for Missing Data: Concepts and New Development 1 (n.d.) (unpublished manuscript) (on file with SAS Institute), http://www.ats.ucla.edu/stat/sas/library/multipleimputation.pdf [https://perma.cc/8V9X-SX42] ("Single imputation does not reflect the uncertainty about the predictions of the unknown missing values, and the resulting estimated variances of the parameter estimates will be biased toward zero.").

320.   *See* King et al., *supra* note 312, at 53 ("Multiple imputation involves imputing *m* values for each missing item and creating *m* completed data sets. Across these completed data sets, the observed values are the same, but the missing values are filled in with different imputations to reflect uncertainty levels. That is, for missing cells the model predicts well, variation across the imputations is small; for other cases, the variation may be larger, or asymmetric, to reflect whatever knowledge and level of certainty is available about the missing information. Analysts can then conveniently apply the statistical method they would have used if there were no missing values to each of the *m* data sets, and use a simple procedure . . . to combine the *m* results. As we explain below, *m* can be as small as 5 or 10.").

321.   *See* John W. Graham et al., *How Many Imputations Are Really Needed? Some Practical Clarifications of Multiple Imputation Theory*, 8 PREVENTION SCI. 206, 206-13 (2007). In Table 5, the authors show the acceptable power falloff for different levels of "missingness." *Id.* at 212. In our data, 29.67% of our unweighted sample report not knowing whether they signed a noncompete. Following the third row, the authors recommend 20 imputations. *Id.* We added an additional five imputations to increase power further.

322.   We used Stata's "mi impute chained" command. See STATA MULTIPLE-IMPUTATION REFERENCE MANUAL: RELEASE 13, STATACORP LP, (2013), https://www.stata.com/manuals13/mi.pdf, for additional details. We imputed the following survey variables: whether the respondent has a retirement plan, deferred compensation, stock options, and/or is unionized; is currently constrained by a

predictive variables in the imputation all variables that are likely to be related to noncompetes, as well as all of the variables that are also being imputed in the same step.[323] We used a logit model to impute

noncompete; has ever signed a noncompete; is aware of his or her employer suing a co-worker over a noncompete; negotiated over his or her initial pay, current pay, or training level; agrees or strongly agrees that he or she is loyal to his or her employer, that his or her job is secure, that he or she would become a "boomerang" employee if he or she ever left, that he or she is satisfied with the job, and that his or her employer shares all valuable work-related information with him or her, is committed to upgrading his or her skills, and values creativity. We also imputed whether the employer has a reputation for threatening employees over their noncompete, and whether in the last year or (separately) in the last two years, the respondent received a merit raise, a raise due to an outside offer, a promotion, an offer from a competitor, an offer from a noncompetitor, formal firm-sponsored training, informal firm-sponsored training, or self-sponsored training; whether the respondent has been recruited by a competitor or noncompetitor throughout his or her tenure, has received an offer from a competitor or noncompetitor throughout his or her tenure; feels like the noncompete limits his or her job options (both before and after an information experiment). While the prior list of variables are all binary, we also imputed the following continuous variables: the likelihood a respondent believes that his or her firm would sue him or her over a noncompete or that a state court would enforce it (both before and after the information experiment); respondent search effort directed toward competitors or noncompetitors; the proportion of the occupation that is present in respondent's industry; the number of establishments and employment in respondent's county and industry; the number of positions respondent has held with his or her employer; the probabilities that respondent leaves for a competitor or noncompetitor in the next year and that respondent receives an offer from a competitor or noncompetitor in the next year; the number of hours worked per week; weeks worked per year; income; and raking weights.

323.   These variables include an indicator for the respondent claiming he or she will never move, a third-degree polynomial in the log of the competitor-specific reservation wage, a third-degree polynomial in tenure, a third-degree polynomial in experience, a third-degree polynomial in age, indicators for job status (unemployed, employed, multiple jobs), education, how the respondent is paid, having signed a non-disclosure, non-solicitation, non-poaching, arbitration, IP pre-assignment contract, whether the respondent reports having heard of noncompetes, whether the respondent expects to work more than 10 years at his or her employer, gender, industry, occupation, multi-unit firm, multi-state firm, categories for establishment size, firm size, the type of sensitive information the respondent possesses, whether the noncompete is a factor in moving or starting a firm, whether the respondent reports receiving an offer, whether the respondent reports that he or she took the survey because of a noncompete, whether the respondent searched for another job in the last year by contacting other employers or sending out applications. We also included continuous controls for respondent self-reported motivation (including desires for internal and industry advancement, desire for money, dedication to employer, desire to master skills, and desire to have an impact on the world) and controls for self-reported measures of effort, performance, and creativity. Finally, we included a third-degree polynomial in noncompete enforceability (see *infra* Part III for a discussion of this data) from Starr (2015) fully interacted with an indicator

binary variables, and predictive mean matching to impute continuous variables. The latter procedure involves creating a predicted value for non-missing values from a regression of the missing variable on a set of covariates, and then picking randomly among a set of nearest neighbors.[324] By selecting randomly among nearest neighbors, we ensured that the predicted variables maintain consistent and plausible characteristics (e.g., income cannot be less than zero).[325]

The Stata Manual provides a simple exposition of how multiple imputation works:

> Multiple imputation (MI) is a flexible, simulation-based statistical technique for handling missing data. Multiple imputation consists of three steps:
>
> 1. *Imputation step*. *M* imputations (completed datasets) are generated under some chosen imputation model.
>
> 2. *Completed-data analysis (estimation) step*. The desired analysis is performed separately on each imputation $m = 1, \ldots, M$. This is called completed-data analysis and is the primary analysis to be performed once missing data have been imputed.
>
> 3. *Pooling step*. The results obtained from *M* completed-data analyses are combined into a single multiple-imputation result. The completed-data analysis step and the pooling step can be combined and thought of generally as the analysis step.[326]

The imputation step itself involves three steps: First, we fit a regression model with the observed data, predicting a variable we would like to impute. Second, we simulate new coefficients based on the joint posterior distribution of the coefficient estimates and standard deviations. Third, we apply these simulated coefficient estimates to the observed covariates to impute the missing values. We repeated this process 25 times to generate 25 fully imputed datasets.[327] Each variable that we predicted, along with the percentage of missing values and descriptive statistics for the non-missing values, is shown in Table 18.

---

for being in the information experiment treatment group and measures of the importance of being sued or having a noncompete enforced in determining whether the respondent would move.

324.   *See* STATA, *supra* note 322, at 241-46. The Stata manual cites a source that recommends at least 10 nearest neighbors, and potentially more with larger sample sizes. *Id.* at 243. We chose to use 15 nearest neighbors, but supplementary analysis confirmed that choosing 10 as opposed to 15 does not change the results.

325.   We allowed for the imputed value to be selected from among a set of the 15 nearest neighbors. *See* STATA, *supra* note 322, at 241-46.

326.   *Id.* at 3.

327.   *See id.* at 245-46.

Table 18
CHARACTERISTICS OF IMPUTED VARIABLES

Variable Description	Number of Missing Observations	Mean of Nonmissing Observations	Std. Dev. of Nonmissing Observations
Negotiated initial pay	7,374	0.31	0.46
Negotiated current pay	7,374	0.31	0.46
Employer has a reputation for taking actions to prevent employees from violating their noncompetes or other agreements	6,261	0.20	0.40
Currently signed noncompete	3,414	0.22	0.41
Ever signed noncompete	3,105	0.43	0.50
Aware of an instance when employer sued departing coworker over noncompete	2,851	0.08	0.28
Negotiated over training	2,764	0.07	0.25
Percent of total employment in industry made up by respondent's occupation	1,433	0.28	0.30
Total employment in respondent's occupation in respondent's industry	1,433	91,612.95	153,168.60
Deferred compensation	1,316	0.13	0.34
Recruited in last year to join competitor	1,315	0.23	0.42
Raking weights	1,279	1.00	1.30
Annual compensation	1,270	50,680.44	41,436.23
After experiment, respondent would negotiate	1,202	0.74	0.44
Received merit based raise within last year	962	0.50	0.50
Received raise as a result of outside offer in last year	962	0.09	0.29
Received promotion in last year	962	0.17	0.37
Received an offer from a noncompetitor in last year	962	0.12	0.32
Received an offer from a competitor in last year	962	0.11	0.31
Received firm-sponsored formal training in last year	962	0.33	0.47
Received firm-sponsored informal training in last year	962	0.47	0.50
Received self-sponsored training in last year	962	0.10	0.30
Received merit based raise in last two years	962	0.59	0.49
Received raise as a result of outside offer in last two years	962	0.12	0.32
Received promotion in last two years	962	0.25	0.43
Received an offer from a noncompetitor in last two years	962	0.15	0.36
Received an offer from a competitor in last two years	962	0.14	0.35
Received firm-sponsored formal training in last two years	962	0.40	0.49
Received firm-sponsored informal training in last two years	962	0.54	0.50
Received self-sponsored training in last two years	962	0.12	0.33
Probability of leaving for a competitor in next year	962	13.15	21.97
Probability of leaving for a noncompetitor in next year	962	19.72	28.41
Probability of receiving an offer from a noncompetitor in next year	962	22.17	30.36
Probability of receiving an offer from a competitor in next year	962	18.82	27.19
Recruited in last year to join noncompetitor	933	0.28	0.45
Agree or strongly agree that respondent's job is secure	782	0.78	0.42
Agree or strongly agree that respondent is loyal to employer	782	0.81	0.39
Agree or strongly agree that respondent is satisfied with job	782	0.68	0.47
Agree or strongly agree that employer shares all work-relevant information with respondent	782	0.56	0.50
Agree or strongly agree that employer is committed to upgrading respondent's skills	782	0.54	0.50
Agree or strongly agree that employer values creativity	782	0.61	0.49
Agree or strongly agree that respondent would boomerang back to employer if ever left employment	782	0.67	0.47

Note: This table shows all of the variables that are imputed, how frequently they are missing, and the mean and standard deviation of the nonmissing observations. The final sample size is 11,505.

One can observe the impact of replacing suspect values of the income variable by comparing the distribution of Sample 1's income variable to Sample 4's distribution in Figure 2. The tails of Sample 1's distribution are much longer on both ends. This occurs because

Table 18 (continued)
CHARACTERISTICS OF IMPUTED VARIABLES

Variable Description	Number of Missing Observations	Mean of Nonmissing Observations	Std. Dev. of Nonmissing Observations
Search effort (0–10) directed towards noncompetitor in last year	782	3.33	3.29
Search effort (0–10) directed towards competitor in last year	782	2.73	3.05
Stock options at current employment	551	0.20	0.40
Unionized at current employment	374	0.07	0.25
Retirement plan at current employment	265	0.61	0.49
Number of positions respondent has had within employer	218	1.65	1.22
Number of weeks worked per year	203	48.41	6.77
After experiment, believes noncompete limits options	198	0.60	0.49
Before experiment, believes noncompete limits options	198	0.57	0.49
Received job offer from a competitor while employed with current employer	78	0.22	0.41
Received job offer from a noncompetitor while employed with current employer	77	0.26	0.44
Before experiment, probability firm will sue over noncompete	55	39.51	35.52
After experiment, probability firm will sue over noncompete	55	37.93	34.67
Before experiment, probability court will enforce noncompete	55	43.35	36.69
After experiment, probability court will enforce noncompete	55	42.32	36.38
Hours respondent works per week	20	38.95	11.16
Number of establishments in respondent's county and industry in 2014	16	2,329.73	4,230.92
Number of employees in respondent's county and industry in 2014	16	39,389.48	66,184.81

Note: This table shows all of the variables that are imputed, how frequently they are missing, and the mean and standard deviation of the nonmissing observations. The final sample size is 11,505.

Figure 2
SAMPLE 1 V. SAMPLE 4 LOG INCOME DISTRIBUTIONS
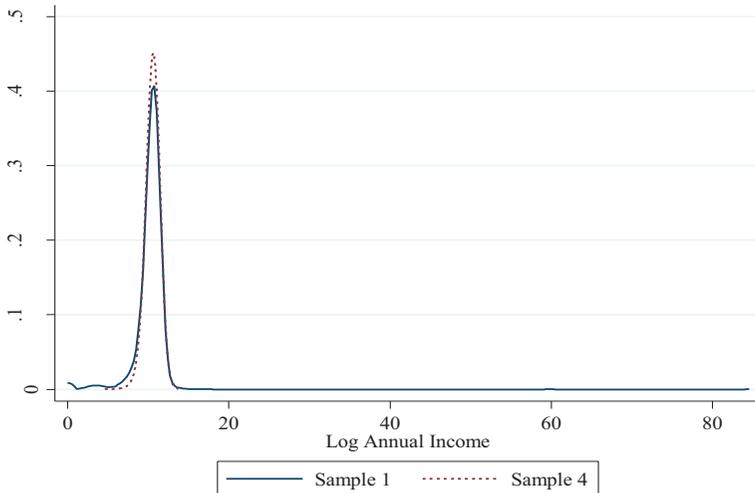
Note: kernel = epanechnikov, bandwidth = 0.1000

there are respondents who report unreliable zeros and extremely large (and highly improbable) numbers. Replacement and imputation

redistributes the mass of these observations to the central part of the distribution in Sample 4.[328] In the end, deciding how best to identify and treat missing values and idiosyncratic errors requires careful thinking about the raw quality of the data, the research question, the source of item nonresponse,[329] and the costs and benefits of various imputation methods.[330] No easy recipe exists, unfortunately.[331]

G. Re-Weighting

After we implemented the imputation procedure above, we finalized the dataset by re-weighting it using iterative proportional fitting (raking).[332] With the dataset comprising 25 separately imputed datasets, we used raking to create weights for each observation so that each dataset would be nationally representative with respect to our demographic variables of interest. After this procedure, we were left with a clean and weighted dataset ready for analysis.

III. NONCOMPETE INCIDENCE AND ENFORCEMENT

After demonstrating the need for individual employee-level data in Part I and laying out our survey approach and data methods in Part II, we have accomplished the bulk of what we set out to do in this Article. We are confident that the 2014 Noncompete Survey data

---

328. As a result of the implausibility of some of the respondent-reported income numbers and the propensity for typos in this domain, we believe that the cleaned income measure is a better choice for any analysis involving income levels. We include our constructed income measure in every sample except for Sample 1.

329. We note that in some circumstances investigation can mitigate such concerns. For example, we were able to identify the source of item nonresponse for weeks worked, hours worked, and number of positions: every respondent with missing values for these variables took our initial pilot version of the survey. Item nonresponse was therefore due only to selection on being the first to take the survey, which we believe is unlikely to be correlated with unobservables.

330. *See generally* RODERICK J.A. LITTLE & DONALD B. RUBIN, STATISTICAL ANALYSIS WITH MISSING DATA (2d ed. 2002).

331. Overall, for purposes of this Article, we take an agnostic approach in which we let the data speak for themselves, but our preferred sample for most analyses is Sample 4 because of the plausible way in which it accounts for most types of idiosyncratic error. Moreover, given the easy availability of other samples constructed using alternative assumptions, we are able to evaluate the robustness of our research to this choice in most situations. Indeed, we employed Sample 4 data in our analysis below in Part III, and yet we found the same basic conclusion when we attempted the same analysis on Samples 1, 2, and 3.

332. We must perform this step because the previous weighting step generated weights only for the observations with non-missing data.

will supply the starting block in the years ahead for many important strides in our understanding of the role that noncompetes play in the U.S. labor market.[333] Furthermore, we hope that this work—and the work of other researchers, using our survey data as well as other sources—will better inform the increasingly important policy debates over the regulation of noncompetes, keeping them rooted in evidence whenever possible and promoting an acknowledgment of ignorance whenever not.[334] In this Part, we aim to highlight the potential value of our survey data by examining a single hypothesis.[335] Specifically, we test whether noncompetition agreements are more common—as one would naturally expect—in jurisdictions where such agreements are supposedly more enforceable than in jurisdictions where they are rarely enforceable or unenforceable.

In order to assess whether noncompete incidence—essentially, the regularity of noncompete use—is correlated with a jurisdiction's (in this case, a state's) noncompete enforcement regime, we must be able to gauge the relative frequency of noncompetition agreements by state. Our survey data contains information on the residence and employment location of each respondent,[336] and we can use the bounding approach or multiple imputation to estimate whether a respondent is currently restricted by a noncompete.[337] The remaining required input is some measure of noncompete enforcement intensity (or enforceability) by state. Our survey did *not* assemble information

---

333.   *See, e.g.*, Evan Starr, The Use and Impacts of Covenants Not to Compete (Kauffman Foundation Proposal, Sept. 2015) (unpublished manuscript) (on file with authors) (seeking support for many future and ongoing research projects making use of the survey data).

334.   *See, e.g.*, Lydia DePillis, *The Rise of the Non-Compete Agreement, from Tech Workers to Sandwich Makers*, WASH. POST (Feb. 21, 2015), https://www. washingtonpost.com/news/wonk/wp/2015/02/21/the-rise-of-the-non-compete-agreement-from-tech-workers-to-sandwich-makers [https://perma.cc/RVV5-GAN3]; Patrick Thibodeau, *As Noncompete Agreement Use Expands, Backlash Grows*, COMPUTERWORLD (Feb. 5, 2016, 1:12 PM), http://www.computerworld.com/article/ 3030641/it-careers/as-noncompete-agreement-use-expands-backlash-grows.html [https://perma.cc/HX72-XEMJ]; Russell Beck, *Noncompete Studies Suggest Noncompete Enforcement Improves Training and Quality of Startups*, FAIR COMPETITION L. (Feb. 22, 2015), http://faircompetitionlaw.com/2015/02/22/noncompete-studies-suggest-noncompete-enforcement-improves-training-and-quality-of-startups [https://perma.cc/ 2VJ4-DMZY].

335.   Our presentation here summarizes and elaborates on analysis we initially reported in a working paper. *See* Evan Starr, Norman Bishara, & J.J. Prescott, *Noncompetes in the U.S. Labor Force* (SSRN Working Paper), http://ssrn.com/abstract=2625714 (May 10, 2016).

336.   *See supra* Section II.D.

337.   *See supra* Section II.D.

on this dimension of the noncompete system. Fortunately, as we note in Part I, most existing empirical research on this topic invokes legal enforcement metrics to draw conclusions about the consequences of noncompetes.[338] We therefore borrow from this literature, exploiting a recently developed enforcement measure that builds on Bishara's multi-dimensional measure of state enforcement intensity.[339]

Measuring the "strength" of a noncompete legal enforcement regime is challenging: Relevant doctrines are ancient, multifaceted, and heterogeneous.[340] Enforcement—or enforceability—is governed by both statute and precedential case law,[341] and interjurisdictional issues are complex.[342] Still, it is possible to summarize the key dimensions along which legal variation is most significant.

A few states such as California and North Dakota do not enforce noncompetes at all.[343] Most states will enforce one, but only if the agreement is "reasonable."[344] All states are in accord that there must be some "protectable interest" at issue, such as a trade secret, confidential information, or a client list, but states differ with respect to what those interests must be.[345] Florida and Kentucky, for instance, include general skills training.[346] Some states, but not others, will only enforce a noncompete against someone who voluntarily quits.[347]

---

338. *See infra* discussion surrounding notes 352-57 (discussing different measures of enforcement intensity used by various scholars).

339. *See* Starr, *supra* note 31, at 7-8 (relying on Bishara, *supra* note 75, at 772-79).

340. Blake, *supra* note 7, at 645-46 (finding that by the end of the nineteenth century, courts had "accepted the method of decision on which modern refinements were to develop," but that the "restraint-of-trade doctrine is not unitary").

341. There were approximately 1,000 reported noncompete decisions in 2014. Russell Beck, *Trade Secret and Noncompete Survey—National Case Graph 2015*, FAIR COMPETITION L. (Jan. 17, 2015), http://faircompetitionlaw.com/2015/01/17/trade-secret-and-noncompete-survey-national-case-graph-2015/ [https://perma.cc/ M5RM-LF9D].

342. *See, e.g.*, Glynn, *supra* note 27, at 1418-24 (discussing the complex problems of interjurisdictional competition in the context of determining which legal regime should govern noncompetes).

343. *See* Bishara et al., *supra* note 1, at 4; Bishara, *supra* note 75, at 757, 767, 778.

344. *See* Anenson, *supra* note 38, at 17 (stating that the "English rule of reason," which focuses on a reliability analysis, "remains the doctrinal scheme in the majority of states" in the United States).

345. Bishara, *supra* note 75, at 773.

346. *See* Lester, *supra* note 85, at 58 nn.40 & 44; Brandon S. Long, Note, *Protecting Employer Investment in Training: Noncompetes vs. Repayment Agreements*, 54 DUKE L.J. 1295, 1310, 1312 n.95 (2005).

347. *See, e.g.*, SIFCO Indus., Inc. v. Advanced Plating Techs., Inc., 867 F. Supp. 155, 157 (S.D.N.Y. 1994) ("New York courts will not enforce a non-

Certain states will rewrite an "unreasonably" broad noncompete,[348] while others will simply refuse to enforce it (like Wisconsin).[349] Colorado courts will not enforce a noncompete against anyone who is not in upper management.[350] States also differ in the procedure and consideration necessary for enforcement. Oregon requires that firms notify employees of any noncompete two weeks before they begin their employment, and if they do not, the firm must offer additional consideration for the modification.[351]

　　Malsberger and his colleagues began collecting and organizing these dimensions by state years ago, making them available in a survey volume.[352] Over the last 15 years, a number of scholars have sought to translate Malsberger's qualitative work into a quantitative measure that captures something akin to "enforcement intensity." In 2003, Stuart and Sorenson created a single indicator variable—either zero or one—to measure enforceability.[353] Garmaise aggregated 12 dimensions of enforceability (scoring each as a zero or a one) into a single numerical index in 2009.[354] In 2011, Bishara assembled two indices—one for 1991, one for 2009—using seven dimensions. These indices assign states a score between one and 10 on each dimension, and then aggregate these dimensional scores using

---

competition provision in an employment agreement where the former employee was involuntarily terminated.").

　　348.　　*See, e.g.*, FLA. STAT. ANN. § 542.335(c) (West 2015) ("If a contractually specified restraint is overbroad, overlong, or otherwise not reasonably necessary to protect the legitimate business interest or interests, a court shall modify the restraint and grant only the relief reasonably necessary to protect such interest or interests.").

　　349.　　WIS. STAT. ANN. § 103.465 (West 2015) ("Any covenant, described in this subsection, imposing an unreasonable restraint is illegal, void and unenforceable even as to any part of the covenant or performance that would be a reasonable restraint.").

　　350.　　COLO. REV. STAT. ANN. § 8-2-113 (West 2015) ("Any covenant not to compete which restricts the right of any person to receive compensation for performance of skilled or unskilled labor . . . shall be void, but this subsection . . . shall not apply to . . . executive [management] and management personnel.").

　　351.　　OR. REV. STAT. ANN. § 653.295(1) (West 2015) ("A noncompetition agreement entered into between an employer and employee is voidable and may not be enforced by a court of this state unless . . . [t]he employer informs the employee in a written employment offer received by the employee at least two weeks before the first day of the employee's employment . . . .").

　　352.　　*See* BRIAN M. MALSBERGER ET AL., COVENANTS NOT TO COMPETE: A STATE-BY-STATE SURVEY (2015).

　　353.　　Stuart & Sorenson, *supra* note 4, at 190.

　　354.　　Garmaise, *supra* note 8, at 388-90. To be precise, every of the 12 dimensions is measured as either a zero or a one, and the score for a state is the sum of all 12 of these dimensional measures. *Id.*

weights selected to capture each dimension's relative importance.[355] In this Article, we rely on Starr's 2015 enforceability indices. Starr builds on Bishara's measure by replacing its subjective dimensional weights with weights calculated using confirmatory factor analysis.[356] Figure 3 reproduces Starr's indices.[357]

Figure 3
STARR (2015) FACTOR ANALYSIS ENFORCEABILITY INDICES

We conducted our analysis by comparing incidence in states with low levels of enforceability to incidence in states with high levels of enforceability. We proceeded in the following way: First, we divided states into six groups based on Starr's "enforceability" coding for the year 2009. North Dakota and California occupy one end of the spectrum because these states largely refuse to enforce employment noncompetes. We divided the remaining states into five quantiles of enforcement intensity, with states like Florida, which freely enforces noncompetes under its statutory and case law,[358] at the other end of the enforceability spectrum. These quintiles do not,

---

355.   Bishara, *supra* note 75, at 772-79.
356.   *See* Starr, *supra* note 31, at 7-8.
357.   *See id.* at 9 fig.1. Our Figure 3 shows the index for both 1991 and 2009 to illustrate how the index has changed over time. In our analysis, we use only the 2009 enforceability index.
358.   It is understandably very difficult to know the strength of enforcement in a particular state solely using legal indices like Malsberger's compendium. These concerns are treated in other papers, and we do not revisit them here. *See* Bishara, *supra* note 75, at 762-67.

however, contain the same number of states, although our measure of enforceability is relatively smooth. In any event, we do not survey the same number of people in each state, which results in variation in the precision of our incidence estimates across these quintiles.[359] Second, for each of these quintiles, we used our multiple imputation method to estimate the percentage of employees in each quintile who are currently subject to a noncompete.[360] We also used our bounding method as a robustness check—we calculated the maximum and minimum percentages of employees with an existing noncompete, assuming both extreme cases for the "maybe" respondents. We also calculated 95% confidence intervals for all of these quantities.[361] We report our findings in Figure 4.

Figure 4
NONCOMPETE INCIDENCE AND ENFORCEABILITY
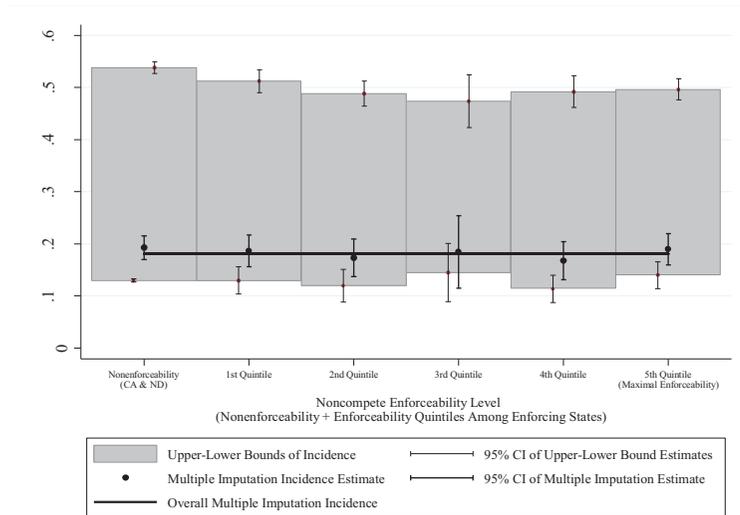
Note: The upper and lower bounds of incidence assume that those who do not know whether they have signed a noncompete did and did not sign, respectively. The noncompete enforceability measure from Starr (2015) is used to divide states into nonenforcing states and quintiles of enforcing states. CI stands for confidence interval.

    While our work is preliminary, to our eyes, the pattern that emerges is relatively stark and ratifies our reasons for conducting the

_____

    359.   *See supra* Section I.B.
    360.   In other words, we used all of the data we had at hand. For our multiple imputation estimates, we included in the tabulations those respondents who answer "yes," those who answer "no," as well as those who answer "maybe." We used the imputation methodology we detail in *supra* Section II.F to sort each "maybe" into the "yes" or "no" bins.
    361.   We clustered the standard errors at the state level.

survey in the first place. According to prior research, *considerable* variation in the enforceability of noncompetes across states exists, and yet the incidence of these contracts does not appear to vary across diverse enforcement regimes. On average, across all quintiles, about 18% of labor force participants are bound by noncompetition agreements. There are some differences across quintiles, but these differences are not statistically significant, and the estimates do not change in the order one expects. Non-enforcing states like California and North Dakota, for instance, have an estimated noncompete incidence of approximately 19.3%, which is actually higher than the corresponding level for every enforceability quintile (the highest enforcing quintile has an incidence of 19.0%).

Importantly, if we consider other observable information about respondents—such as age, occupation, and industry—in a regression framework, the substance of the resulting pattern does not change.[362] Controlling for a host of employee- and firm-level characteristics,[363] we find that noncompetes are 2.5 percentage points more common in the most enforcing quintile than in California and North Dakota (the least enforcing states). But the magnitude of this difference is small, and it cannot be statistically distinguished from zero.[364]

We note that there are implicit assumptions in our approach, as there are in all empirical exercises.[365] We present the bounds (which show a similar "flat" pattern) in case there is concern or confusion over our use of multiple imputation methods. Because the "maybes"

---

362.   These results are available from the authors upon request.

363.   In particular, we included controls for occupation by industry fixed effects, protectable interests (e.g., access to confidential information), the size and structure of the employer, the employee's education, the hours worked per week, the weeks worked per year, an interaction between hours and weeks worked, and a third degree polynomial in age. Thus, the possibility that we might, by random chance, have surveyed different types of employees in different states in a particular confounding pattern seems unlikely to explain our findings.

364.   In fact, the states where noncompetes are most enforceable do not have statistically different incidence rates when compared to non-enforcing states. When incidence is linearly regressed on the enforceability index and other covariates, we estimate a positive, but small and at best marginally significant coefficient (p-value = 0.078). Even if we ignore the evidence of a nonlinear relationship in the data, any relationship between enforceability and incidence appears to be very weak.

365.   *See generally supra* Part II. Importantly, aside from the decisions we made in assembling the data, the empirical approach itself will necessarily rely on assumptions about the data. For some assumptions that are typically seen in this type of data analysis, see Bishara, *supra* note 75, at 762-67.

make up a relatively large percentage of respondent answers,[366] if the respondents who answered "maybe" were more likely to have signed in high enforcement states relative to low enforcement states, Figure 4 would be misleading. But the story that generates this possibility, without also increasing the total number of "yes" answers is hard for us to conjure. Even if there were a satisfying explanation, the data nevertheless tell the story that any relationship between noncompete incidence and enforceability is likely to be relatively weak.

Policymakers interested in the effects of noncompetes might easily start with the assumption that employers and employees care about noncompete enforceability or enforcement intensity—i.e., that both parties would consider the enforceability of a noncompete agreement when deciding whether to require or acquiesce to one. Admittedly, it might not be all that remarkable that employee behavior does not vary much by legal regime. Employees seem much less likely to be aware of governing noncompete law, and may believe that all contractual language is enforceable.[367] Still, it is likely to come as more of a shock that employer behavior appears, at first blush, to be invariant to noncompete enforceability.

One possible interpretation—which we only suggest here—is that *actual* enforceability may be unimportant to parties; instead, *perceived* enforceability on the part of employees is what is critical, and the actual content of the law may have little relationship to what employees perceive.[368] Indeed, employers may misinform employees simply by asking them to sign a noncompete, which is something they may be more likely to do whenever there is a chance that an employee might view it as binding, or perhaps even a chance that the

---

366. We agree with those who find this pattern in the data interesting in its own right. Cognitive limits—like mental fatigue and imperfect memory—may be an important consideration in designing optimal approaches for enforcement. *See* Donald C. Langevoort, *Behavioral Theories of Judgment and Decision Making in Legal Scholarship: A Literature Review*, 51 VAND. L. REV. 1499, 1511 (1998). *See generally* Melvin Aron Eisenberg, *The Limits of Cognition and the Limits of Contract*, 47 STAN. L. REV. 211, 214-16 (1995).

367. *See* Kim, *supra* note 83, at 116-17 ("[I]nformation failures [are] likely to render employees unable to bargain meaningfully . . . . [E]mployer and employee rarely negotiate individual employment contracts in a formal sense."); Rudy, *supra* note 303, at 326 (finding "that employees in Nebraska and Virginia do not understand the legal rules that govern their employment relationships").

368. If this theory has some explanatory power, it creates a significant hurdle for existing empirical scholarship, which has relied on measures of enforceability to determine the consequences of noncompetition agreements. *See supra* Section I.A for a discussion of articles in this category; *see also* Bishara & Starr, *supra* note 58, at 520-30 (describing and critiquing these studies).

employee will subsequently view it as a promise or other expressive device that will likely improve the relationship.

In sum, our findings are at least somewhat at odds with the dominant belief that a state's noncompete enforcement policy will influence the scale of a state's noncompete contracting activity. Furthermore, because our findings suggest that the status quo legal regime in a state may not matter on the ground (at least in terms of incidence levels), policymakers should be wary of presuming that black-letter-law reform can be a useful tool to change employer practices.[369] The fact that the frequency of noncompetes in a state appears unrelated to the governing legal regime does *not* necessarily imply there is also no relationship between noncompete law and employee behavior and outcomes.[370] Even so, our finding does raise the specter that, at some basic level, reflexively curtailing or banning noncompetition agreements in something like the California mold may accomplish much less than many scholars, commentators, and policymakers currently imagine.

CONCLUSION

The impetus for engaging in this large-scale data-collection project has always been to develop a much fuller and more credible understanding of the use of restrictive covenants in the United States. Many important and vigorous discussions are underway about the propriety of noncompetes in employment contracts, yet we know remarkably little about these contractual devices as they operate in the real world. Until now, researchers have had no way of knowing how common noncompetes are, if and how they are negotiated, or whether employees understand their content or legal enforceability. We surveyed over 11,500 labor force participants to provide a representative picture of how employees understand and experience noncompetes and to construct a data resource capable of supplying reliable evidence on critical policy questions.

---

369. However, we ought to stress that certainly some noncompete reforms—for example, mandated employer practices such as Oregon's notice requirement, *see* Rassas, *supra* note 103, at 460—can influence how and when employers request noncompetes. In addition, the fact that noncompete enforceability does not seem to matter to incidence levels hints that policies aimed at educating employees about the content of noncompete law, and ensuring greater understanding of the practical implications of contracting arrangements, is worth exploring.

370. In other words, even if the law does not affect how many noncompete contracts are made, it may still be the case that behavior *under* these contracts—e.g., mobility—differs depending on the governing legal framework.

In this Article, we describe the content and implementation of our survey and the cleaning and refining of the data. We do this not only to lay the groundwork for our future research in this area, but also to provide a model for researchers whose fields might benefit from surveying individuals.[371] Assessing the potential consequences from legal change requires the ability to observe the world clearly. When it comes to the law that governs individual behavior and private interactions, data is all too often sparse or non-existent.

In addition to introducing the survey project, we made a quick foray into analysis to demonstrate its value. We investigated a simple empirical question: Is noncompete incidence positively correlated with state-level enforceability? One might reasonably hypothesize that unenforceable contracts are likely to be rare (at least if they are costly to negotiate) and less common whenever enforcement is more challenging. Yet we find essentially no relationship between formal noncompete law (which receives virtually all of the policy attention in this domain) and the level of noncompete activity. Noncompetes appear to thrive outside of law's shadow, with employers requiring them for reasons unrelated to their formal legal effect.

In the future,[372] the 2014 Noncompete Survey data and the research it engenders will furnish a far more accurate picture of noncompete contracting activity. We will know more about just how common these contractual devices are. We will gain insight into how employees comprehend noncompete law and how those beliefs drive behavior like deciding whether to accept, negotiate over, or leave a job under possible threat of noncompete enforcement—whether that constraint is real, imagined, or simply ignored. More generally, we will understand more about the consequences of noncompetes and enforcement intensity on employment relationships, innovation, and entrepreneurship. And perhaps most importantly, along the way, we will have the tools we need to critically evaluate the assumptions and recommendations of scholars, the media, and policymakers.

---

371.   Our survey instrument itself, and our experience in developing it, can also serve as a useful roadmap for conducting interesting studies concerning smaller slices of the workforce. For instance, one could further break down the labor force with variables such as geography, job description, and other demographics such as age, gender, wages, training investments, and the relevant legal regimes.

372.   We do not describe in detail our research agenda in this Article. Projects underway include a careful exploration of noncompete incidence in the U.S., a fine-grained investigation into the role noncompetes may play on many different mobility-related dimensions (e.g., job search behavior and recruitment activity), and an evaluation of an information experiment to isolate employee assumptions about noncompete enforceability and how those assumptions influence behavior.