Decisional Shortcuts and Selection Effects: An Empirical Study of Ten Years of U.S. District Courts' Employee Misclassification Decisions

Charlotte S. Alexander[*] *and Mohammad Javad Feizollahi*[**]

This report presents the results of an empirical study of ten years of employee misclassification summary judgment decisions by U.S. district courts, in which judges were asked to determine whether a worker was an employee or an independent contractor. Using text mining, machine learning classifiers, and regression analysis, the research reveals among 747 opinions that the judge ruled that the plaintiff was an independent contractor in thirty-eight percent of cases, and that the plaintiffs' occupation was a strong predictor of outcomes. The findings also suggest that the law's failure to define "employee" effectively may lead courts to adopt decisional shortcuts, using a feature like the plaintiff's occupation as a proxy for employee status. Courts' decisions may then influence plaintiffs' lawyers' own decisions about claim viability, producing a feedback loop in which certain occupations are heavily selected for litigation.

TABLE OF CONTENTS

---

[*] Principal Investigator. Associate Professor of Legal Studies, Department of Risk Management and Insurance, J. Mack Robinson College of Business, Georgia State University; secondary appointments, Georgia State University College of Law and Institute for Insight. Director, Legal Analytics Lab. This project was funded by the U.S. Department of Labor, Chief Evaluation Office, Labor Research and Evaluation Grant EO-30275-17-60-5-13. We thank our team of research assistants who worked on this project: Harrison Alex, Ayushri Bhargava, Hardik Dave, Hayden Hillyer, Siddharth Kapoor, Siddhant Maharana, Girraj Maheshwari, Ojasvi Maleyvar, and Omkar Raut. We also thank Jessica Lohmann and Chayun Yi of the Chief Evaluation Office at the U.S. Department of Labor.

[**] Co-Principal Investigator. Assistant Professor of Business Analytics, Institute for Insight, J. Mack Robinson College of Business, Georgia State University. Faculty Affiliate, Legal Analytics Lab.

INTRODUCTION

This report presents the results of an empirical study of ten years (2008-2017) of employee misclassification summary judgment decisions by U.S. district courts, in which judges were asked to determine whether a worker was an employee or an independent contractor.

"Employee misclassification" refers to employers' practice of improperly classifying employees as independent contractors. Despite their label, these workers function like employees: they are controlled by and economically dependent on a single employer, and lack the flexibility, entrepreneurial opportunity, and autonomy of true independent contractors. By virtue of their contractor status, however, misclassified workers are granted few workplace rights, as nearly all federal labor and employment statutes apply only to employees. This means that misclassified workers may receive sub-minimum wages and no overtime, suffer discrimination with no recourse, receive no benefits from the workers' compensation system, lack protection when organizing into a union or bargaining collectively, and receive no family and medical leave rights under federal law. Misclassification also has significant tax implications, as employers evade their payroll tax obligations when classifying workers improperly.

Courts are meant to check employers' power to classify. Misclassified workers may file suit alleging violations of their workplace rights and seek reclassification as employees for purposes of the litigation. High profile class actions against FedEx, Uber, and Amazon, for example, have asserted overtime and other pay claims in connection with those companies' use of independent contractors.

As the structure of work continues to fissure in today's "gig" or "on-demand" economy, the distinction between employee and independent contractor will come under increasing scrutiny. However, public understanding of the ways in which courts actually decide misclassification disputes is inadequate. This is because there is no data set that records courts' resolution of misclassification disputes, the reasoning that judges employ, or the factors associated with different outcomes.

This report begins to fill that gap by investigating the past decade of federal district court misclassification decisions, using custom-built algorithms to extract key information from the opinions' text and to build models to explore and predict the decisions' outcomes. Specifically, this research engages with the following questions:

- What is the win/loss rate for plaintiffs who claim misclassification? In other words, how high of a barrier does misclassification construct between plaintiffs and their access to the protections of labor and employment law?

- What legal test(s) do courts use to examine employee status? As the caselaw has developed around classification, so have multiple multi-factor tests: the common law test, the economic realities test, and a hybrid of the two. To what extent do different courts adopt these different tests? Does the adoption of any given test appear to be correlated with a particular outcome or other court-specific factors?

- Within each test, which factor or factors appear to be the most strongly associated with different outcomes? In addition, how frequently do courts refer to evidence, such as the existence of a written contract, that is outside those factors?

- What other worker, employer, lawyer, and litigation variables – beyond the factors enumerated by the test – are associated with plaintiff wins and losses? How do outcomes differ across jurisdictions, or for different plaintiff occupations or defendant industries, for example?

To answer these questions in brief, and to forecast the analyses that follow, the study found that, of the 747 summary judgment decisions studied, judges ruled that the plaintiff was an independent contractor in thirty-eight percent. The remaining sixty-two percent represented worker-friendly decisions, evenly divided between judges' affirmative findings on summary judgment that the plaintiff was an employee, and denials of summary judgment due to unresolved issues of material fact, allowing the plaintiffs' claims to proceed. Cases filed under the Fair Labor Standards Act (FLSA) claiming unpaid minimum wages or overtime dominated the data set, representing over two-thirds of all cases, as did the legal test used in FLSA cases, known as the economic reality test.

Regarding the factors and other variables that may be associated with or predict outcomes, the study found that the factors concerning the plaintiff's skill, permanency of the work relationship, tax treatment, written employment contract, and the plaintiff's own supervision of assistants appeared to figure more prominently than others in judges' opinions.[1] In addition, pro se plaintiffs fared worse than their represented counterparts, and FLSA cases fared better than other claim types.[2]

Further, plaintiff occupation was a strong predictor of outcomes, and exotic dancers were the least likely to be found to be independent contractors, with a probability of only 9.5 percent, less than half of the probability of the next-lowest occupation.[3] As discussed further at length below, the occupation findings suggest that judges may employ decisional shortcuts, regarding certain occupations as a proxy for the merits of the plaintiffs' claims. Occupation then functions as an unenumerated super-factor in judges' analyses.

A more complex dynamic may also be at work, in which plaintiffs' lawyers select misclassification cases to bring to court on the basis of the plaintiffs' occupation, repeatedly targeting certain classes of plaintiffs for representation. This selection is likely driven by lawyers' perception of the viability of those claims, which judges then reinforce in issuing plaintiff-friendly classification rulings, relying heavily on precedent drawn from the same class of occupation. This analysis holds even in single-plaintiff employee misclassification cases, brought on behalf of individuals rather than classes of workers, where the occupational structure, rather than the details of the plaintiff's particular circumstance, appears to do the work in driving the courts' analysis.

---

[1] *See* Table 4, *infra*, and accompanying discussion for detail on the direction and magnitude of the effects.
[2] *Id.* Holding all else constant, the probability of an independent contractor ruling increased by thirteen to sixteen percentage points when a plaintiff was pro se, as opposed to represented by an attorney. With respect to the FLSA result, the independent contractor probability decreased by about twenty-two percentage points in cases with FLSA claims as compared to their non-FLSA counterparts.
[3] *See* Tables 4 and 5, *infra*, and accompanying discussion.

The remainder of this report explores these and other findings and their implications, drawing on a unique dataset of misclassification summary judgment decisions and deploying analytics tools to parse the text of those decisions. The report proceeds as follows. Part I describes the law of employee misclassification and presents a brief literature review of similar work. Part II describes the data assembly process and methods used in analysis. Part III presents the study's findings and discusses their implications. Part IV concludes.

I. The Law of Employee Misclassification

The issue of employee misclassification is a threshold dispute in employment law cases. In this type of lawsuit, a worker labeled as an independent contractor sues a company, arguing for the rights, benefits, or protections that he or she would have received if properly classified as an employee. Resolving the classification issue is therefore a necessary step, dictating whether the worker may proceed to argue for the unpaid wages, antidiscrimination protection, family or medical leave, or other benefits of employee status.

In resolving misclassification disputes on summary judgment,[4] judges employ one of several multi-factor tests, depending on the underlying statutory claim at issue in the lawsuit. Among the decisions examined here, six main statutory claim types are represented: FLSA, Americans with Disabilities Act (ADA), Age Discrimination in Employment Act (ADEA), Employee Retirement Income Security Act (ERISA), Family and Medical Leave Act (FMLA), and Title VII of the Civil Rights Act of 1964 (Title VII). These six statutes, like the rest of employment law, do a notoriously poor job of distinguishing cleanly between employees and independent contractors. Title VII, the main federal antidiscrimination law, for example, defines an "employee" as "an individual employed by an employer."[5] The definitions offered by other statutes are similarly circular.

In the absence of clear statutory guidance, courts have developed various legal tests composed of varying numbers of factors, which courts use to identify a worker's proper classification. These tests may be grouped into two rough categories: the common law agency test, which is used in Title VII, ADA, and ERISA cases, and the economic realities test, which is used in FLSA, ADEA, and FMLA cases.

The most common formulation of the common law agency test derives from a U.S. Supreme Court case, *Community for Creative Non Violence v. Reid*:

> In determining whether a hired party is an employee under the general common law of agency, we consider the hiring party's right to control the manner and means by which the product is accomplished. Among the other factors relevant to this inquiry are the skill required; the source of the instrumentalities and tools; the location of the work; the duration of the relationship between the parties; whether the hiring party has the right to assign additional projects to the hired party; the

[4] Summary judgment is the stage of civil litigation at which a court may dismiss a case if the evidence so favors one side that a trial would be unnecessary. Summary judgment opinions are typically fully reasoned, citing the law and the evidence that the parties have offered. As a result, they provide a good window into judges' deployment and interpretation of misclassification law. Studying these opinions is superior to studying trial outcomes in misclassification cases, because so few cases go to trial, or settlement outcomes, because settlements are generally unavailable in public court records.

[5] 42 U.S.C. § 2000e.

> extent of the hired party's discretion over when and how long to work; the method of payment; the hired party's role in hiring and paying assistants; whether the work is part of the regular business of the hiring party; whether the hiring party is in business; the provision of employee benefits; and the tax treatment of the hired party. . . .[6]

After listing these eleven factors, the *Reid* Court went on to remark, "No one of these factors is determinative."[7] Though the *Reid* factors are frequently cited by lower courts in deciding employee misclassification cases, different circuits have developed their own, bespoke sets of factors, variously adding to and subtracting from the *Reid* test in their own articulations of the common law agency test.

The economic realities test, by contrast, prompts courts to consider six factors, yet courts emphasize, as in *Reid*, that "[n]o single factor is dispositive."[8]

> The factors are (1) the degree of control that the putative employer has over the manner in which the work is performed; (2) the worker's opportunities for profit or loss dependent on his managerial skill; (3) the worker's investment in equipment or material, or his employment of other workers; (4) the degree of skill required for the work; (5) the permanence of the working relationship; and (6) the degree to which the services rendered are an integral part of the putative employer's business.[9]

As with the common law test, different courts have adopted different versions of the economic realities test, elaborating or trimming the array of relevant factors to be considered.[10]

In some Title VII cases, courts have also applied a "hybrid" test, which combines aspects of the common law agency and economic realities tests. However, most courts today use one of the two the tests identified above, dictated by the underlying statute at issue in the case.[11]

This multiplicity of tests across different claim types, and multiplicity of factors used by different courts, has caused one judge to observe: "Just because [the plaintiff] may not be Defendants' employee for purposes of one state or federal statute does not mean that he cannot be considered Defendants' employee for another. Defendants assert that these differing

---

[6] 490 U.S. 730,751-752 (1989).

[7] *Id*.

[8] Schultz v. Capital Int'l Sec., Inc., 460 F.3d 595, 602 (4th Cir. 2006).

[9] *Id*.; *see also* Wilson v. Guardian Angel Nursing, Inc., No. 3:07-0069., at *20 (M.D. Tenn. Jul. 31, 2008) (noting variation in courts' application of the economic realities test: "Other circuits have endorsed similar inquiries under the heading of a separate, seventh factor.").

[10] Note that this taxonomy of legal tests only covers claims made under federal law, and does not address separate sets of factors developed by state courts in applying state law.

[11] *Compare* Nationwide Mut. Ins. Co. v. Darden, 503 U.S. 318, 323–24 (1992) (discussing the common law test), *and* Cmty. for Creative Non-Violence v. Reid, 490 U.S. 730, 751–52 (1989) (discussing the same), *with* Oestman v. Nat'l Farmers Union Ins. Co., 958 F.2d 303, 305 (10th Cir. 1992) ("The hybrid test, which is most often applied to actions under Title VII, is a combination of the economic realities test and the common law right to control test."), *and* EEOC v. Zippo Mfg. Co., 713 F.2d 32, 38 (3d Cir. 1983) ("Consequently, the hybrid standard that combines the common law 'right to control' with the 'economic realities' as applied in Title VII cases is the correct standard . . . .").

definitions of 'employee' would result in nightmares for business owners, but it is the unavoidable state of the law."[12]

In addition to this proliferation of tests and factors, the tests' non-prescriptive nature creates another source of uncertainty. There is no guidance as to how courts should measure such concepts as "control" or "opportunity for profit and loss," for example. Moreover, judges and scholars, including the authors of the Restatement (Third) of Employment Law, have expressed doubt about whether there is actually any difference in practice among the tests for employee status.[13] The Ninth Circuit agrees, commenting, "We take this opportunity to clarify what the district court ultimately recognized: there is no functional difference between the . . . formulations."[14] In that case, the Ninth Circuit went on to apply the common law test, as first articulated by *Reid* and reiterated in later Supreme Court decisions.[15]

Yet even if courts are functionally operating within a single decision-making framework, that framework is remarkably lacking in structure, as every formulation of the legal distinction between employees and independent contractors essentially boils down to a totality of the circumstances analysis. This means that employee misclassification disputes are heavily fact-dependent. Noting this, one judge, frustrated with the parties' voluminous briefing on all of the disputed factors in a misclassification case, resorted to quoting the Bible: "There is a passage in Psalms that states: 'He pulled me out of the slimy pit, out of the muck and mire and placed my feet upon a rock and gave me a firm place to stand.'. . . The Court sought a firm place to stand to perform the legal analysis required here but the muck and mire was too deep and thick, and the advocacy too slick."[16]

Some scholars have waded into this morass, attempting to document the incidence of employee misclassification as an increasingly common employment practice associated with the booming on-demand gig economy.[17] Other studies have attempted to identify the rates of misclassification across states and industries, and to quantify its impact, primarily in terms of the loss of tax revenue.[18] However, a close study of *how* and *where* the courts draw the line between employees and independent contractors is missing from the literature. Only a single previous study – by PI Alexander – has focused on the courts as the site of analysis, attempting to understand how the courts resolve misclassification disputes.[19] That previous research studied

[12] Perez v. Foreclosure Connection, Inc., No. 2:15CV653DAK (D. Utah Aug. 19, 2016).
[13] As the Restatement puts it, "Decisions interpreting the meaning of employee under the federal antidiscrimination laws illustrate the lack of any sharp distinction between the common-law test, at least as formulated in *Reid* and *Darden*, and a multifactor economic-realities test." RESTATEMENT (THIRD) OF EMPLOYMENT LAW § 1.01 cmt. d–e (AM. LAW INST., Proposed Final Draft 2014) ("The antidiscrimination-law decisions thus highlight the broad common ground covered by the common-law test and the economic-realities test in determining whether or not to classify a service provider as an employee.").
[14] Murray v. Principal Fin. Group, Inc., 613 F.3d 943, 945 (9th Cir. 2010) ("[T]here is no functional difference between the . . . formulations.").
[15] *Id*. (citing Nationwide Mut. Ins. Co. v. Darden, 503 U.S. 318 (1992)).
[16] Lovett v. SJAC Fulton IND I, LLC et al, No. 1:14-cv-983-WSD, at n.2 (N.D. Ga. Aug. 22, 2016).
[17] *See, e.g*., Orly Lobel, *The Gig Economy and the Future of Labor and Employment Law*, 51 U.S.F. L. REV. 51 (2017); Miriam A. Cherry, *Cyber Commodification*, 72 MD. L. REV. 381 (2013); Miriam A. Cherry, *Beyond Misclassification: The Digital Transformation of Work*, 37 COMP. LAB. L. & POL'Y J. 577 (2016); Miriam A. Cherry, *A Taxonomy of Virtual Work*, 45 Ga. L. Rev. 951 (2011).
[18] *See, e.g.,* Françoise Carré, *(In)dependent Contractor Misclassification*, Economic Policy Institute (June 8, 2015), available at http://www.epi.org/publication/independent-contractor-misclassification/ (collecting studies).
[19] Charlotte S. Alexander, *Misclassification and Antidiscrimination: An Empirical Analysis*, 101 MINN. L. REV. 907 (2017).

only employee misclassification disputes that arose within employment discrimination lawsuits, and found that sixty-six percent of plaintiffs lost their claims due to an independent contractor finding. In that work, the presence of a written contract between the parties was associated with an independent contractor ruling. The results presented in the Parts that follow echo this written contract finding, and also offer new insights into courts' resolution of employee misclassification disputes in employment lawsuit types beyond just discrimination.

II. DATA AND METHODOLOGY

The goal of this research was to assemble a corpus of every written opinion, 2008-2017, in which a U.S. district court judge ruled at summary judgment on a plaintiff's employment status; to describe the outcomes and associated variables in those classification disputes; to attempt to build a predictive model that could identify winning and losing misclassification challenges; and to deliver a normative assessment of judges' misclassification decision-making.

As explained below, however, data access challenges proved formidable, and the research team could not meet the goal of assembling a definitive, comprehensive corpus. This is because of serious shortcomings in the way that our country's legal system tracks and labels civil lawsuits, and makes related data available to researchers and the public. The narrative that follows documents these shortcomings, and describes the team's efforts to come as close as possible to a comprehensive set of a decade's worth of employee misclassification summary judgment decisions.

A. The Legal Data Landscape

Before detailing the specific data assembly process used in this project, it is useful to map the legal data landscape more generally. The focus of this research is relatively constrained: summary judgment opinions issued by U.S. district courts on employee misclassification during the ten-year period 2008-2017. However, data availability problems make the preliminary task of identifying the universe of relevant opinions surprisingly difficult, even before moving to an analysis of those opinions' content. There is no single, canonical data set that captures the presence of employee misclassification disputes within a lawsuit, much less judges' resolution of those disputes. The Federal Judicial Center (FJC), the research arm of the U.S. court system, does provide some statistics on claim types and outcomes.[20] However, the FJC's claim classification system – the Nature of Suit (NOS) code – is neither reliable nor granular enough to be useful in this project. When a plaintiff files a case, she or her lawyer chooses a single NOS code from a list on a required document called a civil cover sheet.[21] This code ostensibly classifies lawsuits by the main set of statutory or common law violations alleged: plaintiffs may choose only one NOS code; they are instructed to select the "most applicable" if more than one could apply.[22] These codes are then compiled in the FJC statistics.

Though seemingly helpful at first glance, NOS codes in fact erase substantial detail, as plaintiffs are allowed to choose only one, and any other claims present in the case disappear from the FJC statistics. Moreover, the NOS code system is nowhere near granular enough to identify

[20] Federal Judicial Center, *Integrated Database*, available at https://www.fjc.gov/research/idb.
[21] U.S. Courts, Services and Forms, Civil Cover Sheet, http://www.uscourts.gov/sites/default/files/js_044.pdf.
[22] *Id.* at 2, Sec. IV ("Nature of Suit").

employee misclassification claims, as employment-related claims are classified much more coarsely, on the basis of the statute under which they are brought.[23]

Further, even if the FJC data did contain granular claim-type classification, the codebook itself cautions against using some of its data.[24] Finally, even where data carries no such warning, substantial gaps can exist. For example, in a separate project, the PI downloaded all available FJC data for all civil cases with nine NOS codes associated with labor and employment law. This produced approximately 630,000 records of lawsuits, with forty-seven variables for each one. Nine of those variables had missing data for ninety-five percent or more of the lawsuits, including variables meant to capture whether the case was filed as a class action, variables relevant to transferred cases, and variables relevant to arbitrated cases. Other variables were missing values for about half of the records, including those that capture the party in whose favor final judgment was entered, and whether that judgment included a monetary award, injunctive relief, and/or attorneys' fees and costs. This problem of missing data, along with the unreliability of some of the data that are present, render the FJC statistics frustratingly irrelevant for researchers attempting to understand the work of our country's courts.

Therefore, instead of relying on the FJC's structured data, researchers often turn to court documents themselves, conducting labor-intensive classification projects in order to find patterns and extract meaning from the raw, unstructured text. Here, as legal scholars Sandra Sperino and Suja Thomas summarize, researchers have two options for assembling a set of court documents for analysis: "private databases and the federal courts' public database – both of which have significant limits."[25]

> The private databases do not contain all of the orders issued by trial court judges. Whenever researchers rely on these private databases, their data set is necessarily skewed because it only contains a small portion of the many orders that trial court judges issue . . .
>
> [The public alternative is] PACER (Public Access to Court Electronic Records). This electronic system provides access to judges' orders as well as to pleadings and other documentation related to the courts' dockets. However, PACER is not easy to search. Researchers who use the federal database engage in the time-consuming task of

---

[23] In fact, employment discrimination claims are classified even more coarsely than the statute level. There is a single "civil rights- employment" NOS code, 442, which sweeps in seven different discrimination claim types, leaving ADA claims with their own code, 445. For more detail on the shortcomings of NOS codes, see Christina L. Boyd & David A. Hoffman, *The Use and Reliability of Federal Nature of Suit Codes*, 2017 MICH. ST. L. REV. 997 (2017); Charlotte S. Alexander, *Using Text Analytics to Predict Litigation Outcomes: A Preliminary Assessment*, in LAW AS DATA: COMPUTATION, TEXT, AND THE FUTURE OF LEGAL ANALYSIS, MICHAEL LIVERMORE & DANIEL ROCKMORE, EDS. (Santa Fe Institute Press, forthcoming 2018).

[24] Federal Judicial Center, *Integrated Data Base, Civil Documentation, Field Descriptions*, available at https://www.fjc.gov/sites/default/files/idb/codebooks/Civil%20Codebook%201988%20Forward_0.pdf ("This variable is not used uniformly by the 94 district courts. The Statistics Division advises against the use for this data for analysis purposes because it is not a mandatory data field. (Some courts may be using "9999" to indicate amounts over $1 million while others may be using it as a filler or for an unknown amount.).).

[25] SANDRA F. SPERINO & SUJA A. THOMAS, UNEQUAL: HOW AMERICA'S COURTS UNDERMINE DISCRIMINATION LAW, 132 (2017); *see also* Charlotte S. Alexander & Mohammad Javad Feizollahi, *On Dragons, Caves, Teeth, and Claws: Legal Analytics and the Problem of Court Data Access*, in COMPUTATIONAL LEGAL STUDIES: THE PROMISE AND CHALLEGE OF DATA-DRIVEN LEGAL RESEARCH, RYAN WHALEN, ED. (Edward Elgar, forthcoming 2019) (discussing court data access problems).

> combing through thousands of cases to isolate the relevant ones. To make empirical projects that use PACER more manageable, researchers tend to focus on [whichever] particular courts [are relevant to their particular research questions]. As a result, it is difficult to get an overall picture of what is happening in trial courts across the country.[26]

Here again, claim classification becomes a particular problem. If, as in the present project, a researcher seeks to assemble a comprehensive set of cases that include a particular claim, she might choose to download an over-inclusive set from PACER, sweeping in all possibly relevant NOS codes, and then laboriously examine the text of the court documents to discern the claims that are actually present.

Alternatively, she might use keyword searching within one or more private databases containing the text of judges' decisions. Keyword searching of the opinion text allows for better targeting of the particular decisions of interest – here, employee misclassification – instead of using coarse and inexact NOS codes as an initial filter, as in the FJC data and the PACER system. However, as Sperino and Thomas observe, these results would be limited to whichever documents the private databases choose to include. These commercial providers do not reveal their criteria for inclusion or exclusion, nor the algorithms that drive their keyword search functions. Indeed, as legal scholars Elizabeth McCuskey and Pauline Kim and her co-authors have demonstrated, while these providers offer user-friendly, detailed search interfaces atop huge troves of data, they also selectively omit large numbers of judges' decisions. McCuskey's work, for example, found that thirty percent of reasoned decisions issued by judges from two U.S. district courts on a particular topic were missing from Westlaw and Lexis.[27] Kim et al. also summarize multiple studies that document gaps in court data availability, including one in which the authors "were able to access written opinions for less than twenty percent of judicial orders in the cases they examined."[28]

The research team's own experience substantiates these accounts. As Part II.D below demonstrates, the research team in this project discovered a substantial lack of overlap among the results returned by the three main commercial providers, Westlaw, LexisNexis, and Bloomberg Law, and with PACER results. A researcher using this data assembly approach is therefore left with some basic questions: Should she rely on Westlaw, on Lexis, on Bloomberg, on PACER, on some other provider, on the results that are common to all searches, on the unique set of results, or on some other set? There seems to be no clear answer to this question.

Given this legal data landscape, characterized by unreliable classifications and uneven coverage, the research team in the present project experimented with three different data assembly strategies, described below, to arrive at a set of 747 U.S. district court summary judgment opinions on employee misclassification, 2008-2017. While this set cannot be labeled definitively comprehensive, it was assembled using the best strategies available.

---

[26] SPERINO & THOMAS, *supra* note 25 at 132; *see also* Peter Siegelman & John J. Donohue III, *Studying the Iceberg from Its Tip: A Comparison of Published and Unpublished Employment Discrimination Cases*, 24 LAW & SOC'Y REV. 1133 (1990) (noting selection bias problems with studying only reported or published cases).

[27] Elizabeth Y. McCuskey, *Submerged Precedent*, 16 NEV. L.J. 515, 517 (2016).

[28] Pauline T. Kim, Margo Schlanger, Christina L. Boyd and Andrew D. Martin, *How Should We Study District Judge Decision-Making?*, 29 WASH. U. J.L. & POL'Y 83, 99 (2009) (citing David A. Hoffman, Alan J. Izenman & Jeffrey R. Lidicker, *Docketology, District Courts, and Doctrine*, 85 WASH. U. L. REV. 681, 682 (2007)).

B. Data Assembly Strategy 1

The research team originally planned a data assembly approach that combined the easy searchability of commercial legal research providers with the comprehensive document availability of PACER. Specifically, the team planned to use commercial legal research provider Bloomberg Law to search the docket sheet text of all labor and employment lawsuits, identified by NOS code, filed in the study period, to find summary judgment decisions.

Notably, this process would not involve searching Bloomberg's curated set of *opinions*, which would be subject to the same problems of selective inclusion and exclusion described in the section above. Instead, these searches would crawl the text of the lawsuits' *docket sheets*, which Bloomberg pulls directly from PACER, searching for keywords that identified summary judgment opinions.[29] The intent was to assemble a comprehensive list of all labor and employment law cases during our study period in which a judge issued a summary judgment decision, regardless of issue. Test searches in Bloomberg Law suggested that this process would identify approximately 30,000 employment and labor law summary judgment decisions.

At the next step, the research team would be to obtain the full text of those opinions. Because one goal of this project was to produce shareable data, and because Bloomberg Law, like other commercial legal research providers, prohibits sharing, the team would not actually download the 30,000 summary judgment opinions directly from Bloomberg Law.[30] Instead, the team would obtain them from two sources: RECAP, a free public archive of several million federal court documents, and PACER.[31] After first searching RECAP for any of the 30,000 target documents, the research team would purchase the remaining documents directly from PACER.[32]

By the end of these first two steps, the research team planned to have amassed a superset of all summary judgment decisions, on all topics, issued in employment and labor law cases in U.S. district courts in the years 2008 through 2017. The next steps would require text-based classification in order to identify the decisions that addressed the issue of employee misclassification specifically, and then further analysis of those decisions' text to understand the way that judges were deploying the law in the cases before them.

C. Data Assembly Strategy 2

It soon became apparent, however, that the above plan was cost-prohibitive. Bloomberg Law, like other commercial legal research providers, prohibits any automation of its search function. As a result, the process of assembling a list of the estimated 30,000 target summary judgment decisions through repeated manual searches would occupy too much researcher time

---

[29] Unlike other commercial legal research providers at the time this project was planned, Bloomberg Law allowed U.S. district court docket sheet searching by keywords, while also allowing filtering by NOS code and year.

[30] Bloomberg BNA, About Us: Terms of Service: Subscription Products, Sec. 2(b), https://www.bna.com/terms-of-service-subscription-products/.

[31] RECAP is a joint project of Princeton University's Center for Information Technology Policy and the Free Law Project (FLP), a federal 501(c)(3) nonprofit. *RECAP Project – Turning PACER Around Since 2009*, https://free.law/recap/. The RECAP archive consists of court documents in electronic form that users have already purchased from PACER, and then voluntarily contributed to the archive.

[32] PACER fee waivers, if granted, could defray the cost of downloads, but would have prohibited the transfer of any documents downloaded, frustrating the shareability goal of the DOL grant.

and too many resources, before even reaching the costs of locating, downloading, and analyzing the opinions themselves.

At the same time that the research team encountered this setback, however, a development in legal data access seemed to offer another option. Free Law Project, the nonprofit organization that runs the RECAP archive, began a project to bulk-download all free material available on PACER and add it to RECAP. This was significant, because, under the E-Government Act of 2002, courts are supposed to designate as free any document in which a judge describes the reasoning for his or her decision, exempting those documents from PACER charges, and to make those opinions available online in "text searchable format."[33]

In theory, if all judges' reasoned opinions are designated as free on PACER, and if the Free Law Project amassed all free PACER material, the RECAP archive would then contain the entire summary judgment corpus of interest to this project, at no cost, searchable by court, date, and NOS code. The research team could abandon the resource-intensive Bloomberg Law methodology described above, obtain the full set of summary judgment opinions in one fell swoop from RECAP, and then begin to analyze the text to identify the subset of interest – employee misclassification summary judgment opinions. Yet as legal scholar Peter Martin has observed, and as the research team soon discovered, judges, or their staff, are shockingly bad at following the requirements of the E-Government Act.[34]

Embarking on this revised plan, the research team first pulled 110,878 documents that were identified as free on PACER from Free Law Project's RECAP archive, with the relevant set of NOS codes, in the relevant time period, from all ninety-four U.S. district courts.[35] Notably, this total was almost four times the estimated target of 30,000 labor and employment law summary judgment opinions generated via the Bloomberg Law searches described in the previous section. This, in and of itself, was not troubling, as the free document set on PACER would include not only summary judgment opinions (the focus of this project), but also, *inter alia*, judges' decisions on motions to dismiss, class certification, protective orders, and a host of evidentiary issues. The team would next need to perform substantial filtering to identify the relevant, target subset.

Despite initial hope, however, once the research team began to engage with the RECAP documents, this plan, too, proved unworkable. The team quickly discovered that the 110,878 set was both over- and under-inclusive, in different ways. Some courts appeared to designate anything issuing from a judge as free, including purely ministerial orders that lacked any reasoning. Other courts' free document sets were under-inclusive, omitting reasoned opinions that were readily discoverable via commercial legal research services. The U.S. District Court for the District of Wyoming, for example, designated zero documents as free on PACER in any labor or employment case, in the five years between 2008 and 2012, despite at least fifty known opinions issued by that court in those years, located via keyword searches in Westlaw.

---

[33] Peter W. Martin, *Online Access to Court Records - from Documents to Data, Particulars to Patterns*, 53 VILL. L. REV. 854, 7 (2008) (describing E-Government Act of 2002, Pub. L. No. 107-347, § 205(a)(5), 116 Stat. 2899, 2915 (2002).

[34] *Martin, Peter W., District Court Opinions that Remain Hidden Despite a Longstanding Congressional Mandate of Transparency – The Result of Judicial Autonomy and Systemic Indifference*, at 4-7 (January 12, 2018). Cornell Legal Studies Research Paper No. 17-38. Available at SSRN: https://ssrn.com/abstract=3034399.artin (2018) (illustrating problems with judge non-compliance with opinion availability mandate).

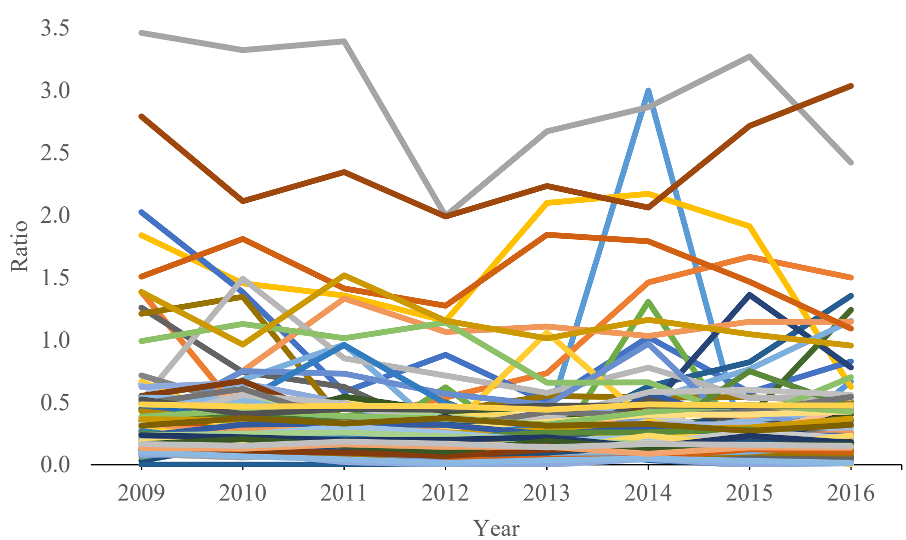
[35] Because this method relies on front-end filtering by NOS code, it is subject to the continuing caveat, explained above, about those codes' unreliability.

Yet without a canonical list of the documents that *should* appear as free on PACER, it was difficult to assess courts' patterns of inclusion or exclusion. As a next-best approach, the team created a roughly normalized measure of courts' decision designation practices, dividing each court's number of free documents on PACER each year by the number of lawsuits filed in that court in the previous year, obtained from Federal Judicial Center case-filing data.[36]

This denominator was designed to account for differences in the number of cases filed in different courts, and in the same courts in different years. The one-year lag assumed that it takes some time after a lawsuit is first filed for the case to generate reasoned judicial decisions, which judges or their staff would then designate as free on PACER. Thus, though some courts and years might be busier than others as an absolute matter, one would expect the *ratio* of free decisions to cases filed to be roughly the same across courts and years, assuming equal compliance with the E-Government Act. It is possible that, even after case-filing normalization, some courts would produce fewer reasoned opinions than others, perhaps because of different local norms among litigants, filing more or fewer motions requiring resolution. Yet even if courts differed in this respect, one would at least expect consistent ratios *within* courts, across years.

The data confounded both sets of expectations, revealing differences in ratios both *within* and *across* courts during the study period. Figures 1 and 2 illustrate both within- and across-court variation. Figure 1 includes all ninety-four U.S. district courts, with ratios on the vertical axis and years on the horizontal axis. The lines, representing each district court, are not labeled in this figure due to space constraints; Figure 2 provides more per-court detail for a selected subgroup. In Figure 1, the key observation is the difference both within and across courts over time. While most lines cluster at the bottom of the graph, indicating a relatively low, consistent ratio of free PACER documents per lawsuit filed, some courts have much higher ratios, and have ratios that vary quite dramatically between years.

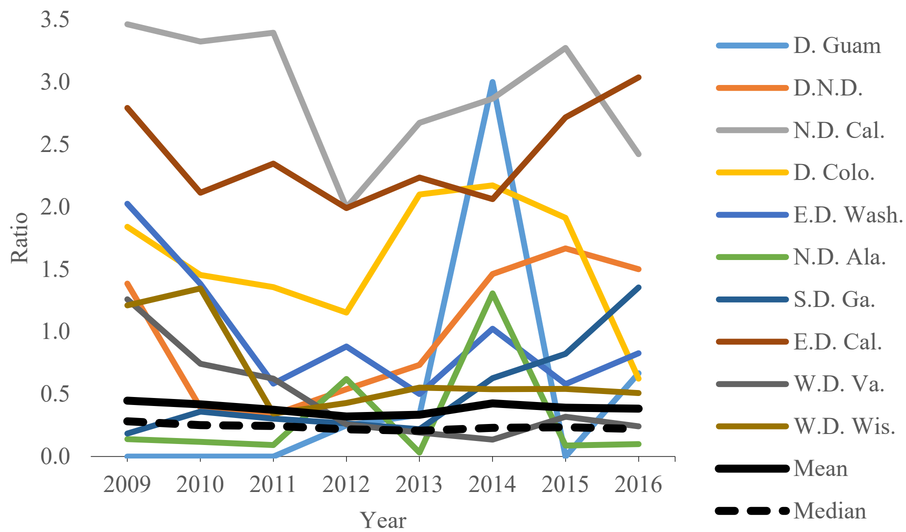
Figure 1: Ratio of Free PACER Documents Per Year to Number of Cases Filed Per Year (Lagged)



[36] Federal Judicial Center, Integrated Data Base, Civil Data, File: Civil Cases Filed, Terminated, and Pending from SY 1988 to Present, https://www.fjc.gov/research/idb. Because these data are recorded by federal fiscal year, we used data from 2009 through 2016 only for this analysis, eliminating any mismatch between fiscal and calendar years in the first and last years of the data set.

Figure 2, in turn, isolates the top ten courts by within-court standard deviation. In very rough terms, standard deviation measures the average distance from the mean of a set of data points, capturing the dispersion of the data across the set. Here, each district has eight data points: one ratio per year for eight years. The ratios for the ten courts shown in Figure 2 fluctuated the most, year to year, during that time. Figure 2 also includes the median and mean ratios for the entire data set for purposes of comparison, revealing that most of the courts that had highly *variable* ratios also had high *value* ratios, relative to other courts. In the Northern District of California, for example, shown as a gray line at the top of both graphs, the ratio of free PACER documents to lawsuits previously filed was substantially higher than in most other districts in all years except one. In addition, that district's ratio was among the most volatile, plunging from 3.39 free PACER documents per lawsuit previously filed in 2011 to 1.99 one year later.[37]

Figure 2: Ratio of Free PACER Documents Per Year to Number of Cases Filed Per Year (Lagged): Top Ten Courts by Standard Deviation



As suggested above, one plausible explanation for the dispersion in ratios both *across* courts and *within* the same courts over the years of the study period is over-inclusivity: that judges, or their staff, are sweeping in more documents than are required by the E-Government Act, and are doing so inconsistently across years. Another explanation is under-inclusivity: that judges are failing to include documents that should be free, again inconsistently. The remainder of this section discusses the research team's solution to over-inclusivity; the next section describes the team's attempts to identify and fill the gaps that we suspected were caused by under-inclusion.

With respect to over-inclusion, the team identified two categories of document that were designated as free on PACER, but were not, in fact, reasoned opinions – the proper target of the E-Government Act. These were party-filed documents such as motions and briefs, and documents that described actions by judges, but contained no reasoning, such as those setting a hearing date or summarily allowing a party to file a brief with extra pages. To remedy this

[37] The district of Guam demonstrated comparable volatility around 2014, but that phenomenon can likely be explained by the relatively small absolute number of lawsuits filed there.

problem, the team constructed a series of filters, applied to the lawsuit-level metadata derived from PACER and to the docket sheet text. We first dropped all documents with two or fewer pages, reasoning that this would exclude purely ministerial judges' orders. We next dropped documents that appeared first, second, or third on the docket sheet. Here, we reasoned that the absolute earliest that a judge's reasoned summary judgment opinion might appear on a docket would be in position four, after the complaint, a motion for summary judgment (assuming no need for discovery), and an opposition. Therefore, the first, second, and third items were not likely to be judges' reasoned opinions.

Next, the team wrote code to crawl the first three words of the docket entry description text to identify words associated with judges' summary judgment opinions, as opposed to party-filed motions or briefs. This produced a set of 49,560 possibly relevant judges' summary judgment decisions from the initial, much larger set of 110,878. Comparing this number to the initial, Bloomberg Law-derived estimate of 30,000 summary judgment opinions in labor and employment law cases over the ten-year period reveals a mismatch, perhaps because our filtering swept in extra documents, or because the initial Bloomberg Law search strategy produced an undercount.

Taking the 49,560 opinions we had identified, the team next turned to the document text, to begin to isolate the decisions that engaged with the question of employee misclassification, the issue central to our research questions. Here, we wrote code to perform a very broad, very rough keyword search, identifying all documents in which the word "employee" appeared within the same paragraph (reading three sentences forward and backward) as "independent contractor." This filter produced a working set of 2,360 decisions. We then shifted to yet another data assembly strategy, to fill the holes that we suspected were present in the data due to courts' failure to comply with the requirements of the E-Government Act of 2002.

D. Data Assembly Strategy 3

The team identified two possible gap-filling strategies by which to locate additional, potentially relevant summary judgment decisions that were not fee-exempt on PACER. First, we went back to the RECAP archive and pulled 38,627 additional court documents that had labor or employment law NOS codes from the target years, but were not tagged as free on PACER. If some judges had failed to designate their reasoned opinions as free, the team thought, perhaps another PACER user had paid to download those opinions and then voluntarily contributed them to RECAP. Using the same set of filters employed above to identify likely summary judgment decisions, this strategy yielded an additional 5,394 judges' opinions. We further filtered based on the presence of "employee" and "independent contractor" in the document text, producing a set of 955 opinions.

Second, we returned to the commercial databases. Here, the team searched the U.S. district court judicial opinion databases of Westlaw, Lexis, and Bloomberg Law in the relevant time period for the same keywords used above: "employee" within the same paragraph as "independent contractor." Table 1 below reports the results, and the overlap among the three, and with the RECAP results.[38]

[38] Determining whether a decision was present in more than one set of search results required substantial work. This is because case names are formatted differently across the result sets, and one system's citation does not necessarily apply to the others. RECAP documents, for example, appear with their court-provided civil action number; that number may or may not be listed on the version of the opinion that is retrieved from a commercial provider.

Table 1: Search Result Comparisons

	Westlaw (WL)	Lexis (LX)	Bloomberg Law (BL)	RECAP
Totals	8,318	6,390	5002	2360
Also in WL	-	5,617 (88%)	4461 (89%)	392 (17%)
Also in LX	5,617 (68%)	-	4431 (89%)	340 (14%)
Also in BL	4,461 (54%)	4,431 (69%)	-	275 (12%)
Also in RECAP	392 (5%)	340 (5%)	275 (5%)	-
Also in any other set	5,864 (70%)	5,786 (91%)	4,638 (93%)	409 (17%)

Within the combined set of results returned from all searches, there were 11,417 unique opinions. Of those, 5,374 were present in only one set of results, whereas 6,043 opinions were present in at least two results sets. There were 4,104 opinions that were present in all three of the commercial legal research providers; only 253 were present in both RECAP and all three commercial providers.

Table 1 shows the extent of the overlap among the four sources. Bloomberg Law, as a whole, overlapped the most with other sources, as ninety-three percent of Bloomberg's results appeared in at least one other result set. Eighty-eight percent of Lexis results also appeared in Westlaw's results set, but only seventy percent of Westlaw's own results, in total, appeared in any other results set.

Turning to the RECAP results, only seventeen percent appeared in any other set. This could be explained by the vast under-inclusion of reasoned opinions in the set of documents that are free on PACER, as suggested by the analyses above, and the zero results for some districts. This could also be further evidence of the flawed system of NOS code assignment. The RECAP documents were initially filtered to include only the NOS codes for labor or employment law cases, whereas the Westlaw, Lexis, and Bloomberg Law results were keyword-based, across all cases, regardless of the NOS code attached. We suspect that many of the results that appear in the commercial providers' result sets but are absent from RECAP carry a non-labor or employment NOS code, and are therefore erased from the lawsuit-level NOS classifications in PACER's metadata.

Thus, we return to the questions from above: Should a researcher rely on Westlaw, on Lexis, on Bloomberg Law, on PACER, on some other provider, on the results that are common to all searches, on the unique set of results, or on some other set? Given the lack of overlap among the four results sets, there does not seem to be a clear answer to this question.

In the end, the research team adopted a conservative approach, to try to amass the most complete corpus possible of employee misclassification summary judgment decisions. The team assembled all unique documents that were not already included in the RECAP set, regardless of which commercial set contained them. However, the team was still bound by this project's goal of shareability, and so could not merely download the opinions from their respective commercial

Moreover, opinions that are published in the Federal Reporter carry a common citation format that is traceable across the commercial providers, but not RECAP. Unpublished decisions may carry a Westlaw, Lexis, or Bloomberg Law-specific citation, but that citation is not present in the search results of other commercial providers. Matching by case name can be accomplished by using various "fuzzy matching" toolkits, but the results inevitably contain some error.

sources. Instead, the team searched Google Scholar's free database of federal court opinions (which requires a case name or case number as a search term), and paid the PACER fees to download the set that remained. In the end, this strategy yielded an additional 7,031 unique opinions, for a total set of 10,923.[39]

The team performed two additional text-based filtering steps before finally turning to the project's central research question of how judges differentiate between employees and independent contractors. First, in order to identify the opinions in which judges actually engaged in an employee misclassification analysis, rather than merely mentioning "employee" and "independent contractor" in passing, the team constructed an additional set of filters based on the presence and frequency of keywords that are used in the employee misclassification analysis at summary judgment. Deploying these filters further reduced the working set of opinions to 4,326.

However, a manual review revealed several categories of irrelevant opinions that had been swept into the 4,326. These included opinions in which courts examined the issue of joint employment, an analysis that uses one of the very same legal tests – economic realities – and many of the same keywords as in the employee misclassification analysis.[40] The team then developed a second filtering process, using machine learning classifiers to distinguish between relevant and irrelevant opinions.

A machine learning classifier is an algorithm, or a set of computer commands, that first "learns" from a training set of pre-classified documents in order to build a model that can accurately sort future documents into the correct categories. In the case of complex, unstructured text such as judges' opinions, such classifiers have the potential to perform faster and more efficiently than people reading and manually categorizing each of the 4,326 opinions as relevant or irrelevant. Such classifiers can also outperform code that relies on a series of pre-programmed rules to perform the sorting function, as it would be very difficult to write, *ex ante*, a set of rules that would be sufficiently comprehensive and flexible to account for all variations in legal text, and to generate accurate classifications across the entire set. The advantage of a machine learning approach is that the computer code generates and deploys its own, highly complex rules for classification, derived from the patterns that the algorithm detects, autonomously, within the training set.

In this project, we first manually classified a set of 1,100 opinions, drawn at random from the larger opinion set of 4,326. Of these, 518 were relevant employee misclassification summary judgment decisions and 582 were irrelevant. In addition to its relevant-irrelevant label, each of the 1,100 opinions was also associated with a frequency count for each keyword that the research team had previously identified and extracted from the text. Further, the team wrote code that identified the statute under which the underlying lawsuit was brought – Title VII, FLSA, or ADA, for example – and extracted all citations to caselaw contained in each opinion.

We then split the 1,100 into a training set, composed of about eighty percent of the 1,100, and test set, composed of the remaining twenty percent.[41] The training set would provide the machine learning classifiers with inputs from which to "learn," or to detect patterns and

[39] This figure cannot be compared directly to the initial 30,000 estimate derived from Bloomberg Law docket sheet searches, because 30,000 is the estimated number of summary judgment opinions in all labor and employment law cases, on all topics, during the ten-year period, whereas the 10,923 was the result of keyword-specific searches to identify probable employee misclassification decisions.

[40] *See, e.g.*, Salinas v. Commercial Interiors, Inc., No. 15-1915 (4th Cir. Jan. 25, 2017) (applying economic realities test in joint employment case).

[41] The training and test sets each retained the same percentages of relevant and irrelevant opinions as in the larger set of 1.100: about forty-seven percent relevant and fifty-three percent irrelevant.

configurations of keywords, citations, and statutes that best predicted the relevant-irrelevant label. Once trained, each classification algorithm would then be run on the test set, to produce a relevant or irrelevant classification label for each of the test opinions. Because the true relevant or irrelevant status of those test opinions was known by the research team, we could then choose the approach that produced the most accurate set of labels, and deploy that approach across the remainder of the full set of unlabeled opinions.

Using these procedures, the team trained and tested six different machine learning classification algorithms, listed below in Table 2, each of which deploys different approaches to deriving classification rules from the training set. The team also developed an additional citation-based classifier, referred to as "cite" in subsequent tables, which we tested on its own and also layered atop the results generated by the machine learning algorithms. This classifier generated lists of top and bottom relevant caselaw citations drawn from across the whole training set. Top citations were defined as those cases that were cited at least four times across the training set, and were cited exclusively in relevant cases. Bottom citations also appeared at least four times, and were cited in irrelevant cases ninety percent of the time. The research team experimented with the accuracy of each of the six machine learning approaches, with the citation classifier alone, and with the machine learning approaches, followed by the citation classifier as an additional filter.

Table 2: Machine Learning Classification Algorithms

Classifier	*Description*
Logistic regression (logistic)	A method for finding associations between one or more independent, or explanatory, variables and a binary or dichotomous dependent variable, which can take on only two values. Logistic regression finds the model that best fits the relationship between the two sets of variables and thus enables classification of new inputs into one of the dependent variable's two values.
Linear Discriminant Analysis (LDA)	An approach for estimating the probability that a new input will fall into a given set of classes, given the extent to which the input's features map onto the features that comprise each class. The class with the highest probability becomes the class for which the prediction is made. Can be used with smaller sample sizes than logistic regression, but assumes that the independent or predictor variables are normally distributed.
Quadratic Discriminant Analysis (QDA)	A similar approach to LDA, but allows nonlinear combinations of independent/predictor variables and relaxes some of the assumptions upon which LDA relies.
K-Nearest Neighbor (KNN)	An algorithm that plots a set of pre-labeled inputs in space, and then places each new input at its appropriate point. The new input receives the same classification as the majority of its *k* nearest

Classifier	*Description*
	neighbors, where *k* is a number assigned by the researcher.
Decision tree (DT)	A classification model that identifies natural splits within a data set and subdivides the data according to the features that best predict the splits. This produces a series of true/false pathways or sequence of conditions that best predict classification.
Random forest (RF)	A grouping of many decision trees, each of which subdivides the data differently, such that the mode or mean of all decision trees produces the best classification, and avoids the problem of overfitting that single decision trees can encounter.

We chose this approach because this set of classifiers is commonly accepted within the literature on machine learning, though each has its own advantages and limitations.[42] In this step of the process, however, the differences among techniques matters less than their ultimate performance in differentiating between relevant and irrelevant opinions. In other words, this stage of the project was highly goal-oriented – narrowing down the larger opinion set of 4,326 to a smaller subset of relevant opinions in the most efficient and accurate manner – and the pathway to that goal mattered less than achieving it.

Ultimately, we found that the random forest classifier, with the citation classifier layered on top, was the best-performing model. After multiple training iterations, this technique detected between seventy and eighty-three percent of independent contractor cases accurately within the training set, with an average of seventy-four percent.

We deployed this methodology across the larger opinion set, generating a set of algorithmically-identified relevant opinions. We then manually reviewed each opinion, in order to identify and throw out any remaining irrelevant opinions that the algorithm miscategorized, and to classify each outcome as an employee ruling, an independent contractor ruling, or a denial of summary judgment.[43] In this way, we ultimately generated the final set of 747 relevant opinions studied here. The next Part reports our findings, and then turns to our explanatory and predictive work, in which we identify the features, factors, and characteristics most associated with, and predictive of, outcomes.

III. FINDINGS AND IMPLICATIONS

Table 3 reports basic summary statistics about the outcome of the employee misclassification disputes in our data set.

[42] *See generally* AURÉLIEN GÉRON, HANDS-ON MACHINE LEARNING WITH SCIKIT-LEARN AND TENSORFLOW: CONCEPTS, TOOLS, AND TECHNIQUES TO BUILD INTELLIGENT SYSTEMS (2017).

[43] We used a manual approach here after trying and failing to write code to extract the judge's decision. We explored various approaches, from extracting all sentences that included "the court finds" or "the court concludes" language and then classifying those sentences according to outcome, to generating a list of possible conclusion sentences, e.g. "The court concludes that the plaintiffs are employees," and using word embeddings to search the full set of opinions for semantically similar sentences. None of these approaches proved successful, however, and so the research team resorted to manual classification of outcomes.

Table 3: Outcomes

Outcome	*Frequency*	*Percent*
Independent contractor ruling	281	37.6%
Employee ruling	236	31.6%
Denial of summary judgment	230	30.8%
Total	747	100%

Independent contractor is the plurality outcome, representing almost forty percent of decisions; the remaining approximately sixty percent represents employee rulings and denials of summary judgment, both worker-friendly rulings. At first glance, these statistics seem to present a much sunnier picture for workers than the findings of the PI's earlier employee misclassification study, which reported independent contractor decisions in sixty-six percent of employment discrimination cases.[44] However, a deeper dive reveals that the results are, in fact, consistent. In this study, twenty-six percent of wage and hour cases brought under the FLSA received independent contractor rulings, while the comparable figure for discrimination cases and other claim types was sixty percent. This suggests that courts may be handling employee misclassification disputes under the FLSA – with its economic realities test – differently from claims that arise under other statutory rubrics. This, as well as other relationships within the data, is explored further in the regression output discussed in the following section.

Regardless of whether an independent contractor finding represents the majority or plurality of outcomes, however, it is notable that judges entered summary judgment at all substantial numbers, given the heavily fact-based inquiry required by the multi-factor tests. Aggregating the "independent contractor" and "employee" outcomes recorded in Table 3 above, in about seventy percent of the opinions studied, the court was able to determine the worker's status, as a matter of law. Though it is impossible to calculate an ideal summary judgment percentage, this number seems high where the legal determination of employee status turns on such subjective concepts as "control" and "dependence." This may provide further support for the claim, made most prominently by Sperino and Thomas, as well as former federal judge Nancy Gertner, that U.S. district court judges are too quick to grant summary judgment as a case management tool, when they should allow the parties to reach a jury to resolve their factual disputes.[45]

Setting aside this normative discussion, the report now turns to the regression and classifier results, to explore the relationship between outcomes and a variety of possible explanatory variables.

A. Classifier and Regression Results

The research team used two types of methodologies to identify associations between outcomes and other variables, and to attempt to predict outcomes as well. Beginning with prediction, the team employed a machine learning classification approach similar to the relevant-irrelevant classification process described in Part II.D above. Here, the goal was to train different

[44] Alexander, *supra* note 19.

[45] SPERINO & THOMAS, *supra* note 25; Nancy Gertner, *Losers' Rules*, 122 Yale L.J. Forum (2012-2013), https://www.yalelawjournal.org/forum/losers-rules.

machine learning classifiers on a subset of decisions labeled with their outcome, along with each decision's array of associated features,[46] and then test the models' performance in classifying new decisions into "independent contractor" and "non-independent contractor" categories. Table 8 in the Appendix reports on the performance of the various classifiers, each of which was explained above in Table 2, in connection with the previous relevant-irrelevant classification task.

Table 8 reports three measures of performance: sensitivity, specificity, and accuracy. Here, sensitivity measures the number of correctly identified independent contractor rulings as a proportion of the total number of correctly identified independent contractor rulings (true positives) plus the number of missed independent contractor rulings (false negatives). In other words, sensitivity is a gauge of how many of the actual independent contractor rulings the algorithm was able to detect within the test set. Specificity performs a similar measure for non-independent contractor rulings, measuring the number that were correctly identified (true negatives) as a proportion of the sum of true negatives and false positives. Finally, accuracy aggregates both measures, capturing the sum of true positives and true negatives as a proportion of all classifications (true positive + true negative + false positive + false negative). The relative importance of each measure depends on the task at hand: designers of lab tests, for example, may want to maximize sensitivity at the expense of specificity in order to ensure detection of as many instances of disease as possible, even if that means some over-diagnosis. In this project, however, there is no reason to prefer one measure over the other, as the research is fundamentally exploratory, and – as explained further below – the team employed additional techniques to complement the classifiers' performance.

As Table 8 shows, the most sensitive method here – the technique that best identified independent contractor decisions and avoided false negatives – was qda.cite, or Quadratic Discriminant Analysis, followed by the top- and bottom- citation detection formula described in connection with Table 2 above. However, at sixty-eight percent of true independent contractor rulings detected, this method's performance was only mediocre. With respect to specificity, the citation method alone performed the best, correctly detecting about ninety percent of non-independent contractor rulings. Random forest, in turn, was the most accurate technique, correctly classifying a combined total of 76.3 percent of independent contractor and non-independent contractor rulings.

Importantly, regardless of their levels of sensitivity, specificity, and accuracy, the *reasons* for most of these classifiers' results are inscrutable, as most of the six classifiers do not identify the variables that were most and least important to the outcome. Moreover, predicting the outcome of a judge's decision using features derived, at least in part, from the decision's text introduces many endogeneity questions. After all, a judge might reach a conclusion on the employee misclassification question based on some pre-set ideology about the issue, and then back-fill the legal analysis in order to justify his or her decision. Any text features that could have been generated in this way, and upon which the classifiers relied, would therefore be problematic for prediction purposes. Nevertheless, due to the "black box" nature of many machine learning classifiers, the research team was not always able to determine the extent to

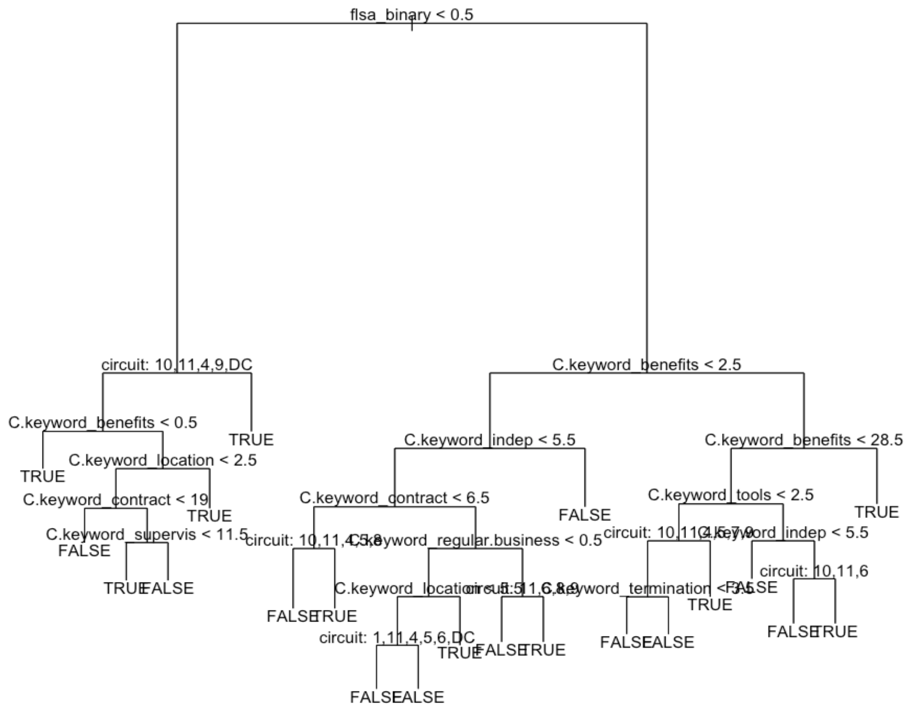
[46] The full set of features, or variables, appears in Table 6 in the Appendix, which reports summary statistics on the data set. The feature selection process is described in the logistic regression discussion below.

which the outcome prediction was influenced by features of the judge-written text versus features exogenous to the judges' own opinion construction, such as plaintiff occupation.[47]

The exceptions to this rule are logistic regression, as well as the decision tree and random forest classifiers. Decision trees produce a visual representation of the splits discovered within the data, a process described in Table 2 above. Figure 3 below is an example decision tree from this project.

[47] The goal of this final phase of the research project was to explore associations between features of each case and the outcome. Therefore, while the predictions generated by KNN are interesting, and relatively accurate, they are less useful when the goal is exploratory or explanatory. This is the reason that the research team added logistic regression as a complementary technique. In contrast, KNN and its other machine learning counterparts *were* useful in the previous stage of the project, described in Part II.D above, where the relevant-irrelevant classification itself was the goal, not an explanation or understanding of why the algorithm generated the classifications it did.

Figure 3: Decision Tree



Random forest, too, allows discovery of the variables that were most important to the ultimate classification. Table 9 in the Appendix lists those variables in order from most important to least. However, the mean decrease in the Gini coefficient measure reported in Table 9 is not directly interpretable as a measure of effect size, a limitation of random forest classification. Indeed, random forest variable importance scores may be better understood as providing a rough approximation of variable importance, rather than a definitive ranking. For these reasons, the team employed another, complementary technique, which does provide some window onto the direction and magnitude of the relationships within the data: logistic regression.

Table 4 reports the results of a logistic regression in which the binary dependent variable captures independent contractor outcomes versus all other outcomes. The independent variables include controls for year and circuit, as well as variables that represent the plaintiffs' pro se status and occupation, judge characteristics supplied by the Federal Judicial Center, whether the case was an FLSA lawsuit or other claim type, and a series of keyword clusters that capture the number of times the opinion used keywords associated with the various factors in the relevant legal tests.[48] The legal tests themselves do not appear as a separate independent variable, because they are highly correlated with claim type: all FLSA cases employed the economic realities test, and all discrimination cases used the common law agency test, for example. The research team chose this set of independent variables in order to capture the variation of interest to the research questions: plaintiff and judge characteristics, differing applicable precedent across years and circuits, different legal tests (captured by claim type), and different factors present or absent within the legal test employed.

[48] Table 7 in the Appendix details the keywords that comprise each cluster.

An additional note is necessary here regarding judges. Eighty-two of the 747 opinions, or about eleven percent, were written by magistrate judges, who are not appointed by the President as are district court judges, but are employees of the U.S. district courts. A magistrate writes a summary judgment opinion either in the form of a report and recommendation to the district court judge, who then reviews the magistrate's findings and issues a final decision, or as a final summary judgment opinion, if the parties have agreed that all matters in the case be resolved by a magistrate.[49] In the latter situation, the magistrate essentially stands in the shoes of the district court judge. Here, our text analysis did not distinguish between the two types of magistrate opinion. More work is needed to distinguish between magistrates' recommendations and final opinions; for purposes of this report, the team ran regressions on the full set of all judges and on the set of district judges only. In any case, as the table below reveals, the array of results does not differ dramatically between the two models.

Table 4 below reports the results of the logistic regression, run on decisions by all judges and by U.S. district court judges only.

Table 4: Logistic Regression Results, Marginal Effects

Variables	*All judges*	*District judges only*
Pro se plaintiff	0.127*	0.161*
FLSA case	-0.215***	-0.225***
Factors		
Skill	-0.0127**	-0.0103
Permanency	-0.0149**	-0.0119**
Tax treatment	-0.00977*	-0.00876*
Supervision	0.00528	0.00782**
Tools	-0.00174	0.00163
Benefits	0.00242	0.00201
Control	-0.00145	-0.00179
Duration	0.0222	0.0226
Independence	-0.00599	-0.0122
Location	0.00992	0.00898
Method of payment	0.00177	-0.0123
Regular business	-0.00672	0.00409
Plaintiffs' occupation		
Administrative/ Office support worker	0.3628***	0.3096**
Installer/ Technician	0.1279*	0.1112
Construction/ Agriculture/ Manufacturing worker	0.1517**	0.1221
Cleaning/ Maintenance worker	0.1860**	0.1754*
Doctor/ Nurse/ Teacher/ Coach	0.2921***	0.2849***
Direct sales/ Marketing worker	0.2485***	0.2253***
Security guard	0.1513	0.1454
Transportation/ Delivery driver	0.3195***	0.3096***

[49] Michelle H. Burns, *U.S. Magistrate Judges: The Breadth and Depth of Their Service*, THE FEDERAL LAWYER 63 (May/June 2014), www.fedbar.org/Resources_1/Federal-Lawyer-Magazine/2014/MayJune/Features/US-Magistrate-Judges-The-Breadth-and-Depth-of-Their-Service.aspx?FT=.pdf.

Variables	All judges	District judges only
Other service occupation	0.4014***	0.3601***
Written contract	0.00303***	0.00244***
Judge		
Type (magistrate/ district)	0.0739	
Gender	0.0283	0.0414
Birth year		-0.00308
Political party of appointing president		0.000521
Commission year		0.000774
Circuit and year controls	omitted from table	omitted from table
Observations	747	665

NOTES: *** p<0.01, ** p<0.05, * p<0.1. Exotic dancer occupation dummy variable in the base category. All predictors set at their mean values in computation of marginal effects.

Notably, Table 4 reports statistical significance, though the 747 opinion set studied here was not assembled randomly or via a process designed to ensure representativeness of a larger universe. As Part II suggested, the central problem with using the tools of inferential statistics to explore these data is that the size and characteristics of the true universe of employee misclassification cases are unknown, meaning that traditional approaches to sampling are inapplicable. Nevertheless, this report includes statistical significance for two reasons. First, if we assume that the 747 opinions are, in fact, randomly drawn from a larger unknown set, then statistical significance is important. Indeed, we have no reason to believe that our multi-pronged, multi-sourced data assembly approach systematically included or excluded certain categories of cases, as we took measures to address both over- and under-inclusivity. It may therefore be safe to assume that missing opinions from our data set are missing at random, that our set of 747 is representative of the larger whole, and that statistical significance is relevant as a measure of certainty.

Second, we could make the even stronger assumption that the research team succeeded in identifying all employee misclassification summary judgment decisions within our universe. Even in this circumstance, statistical significance would still be important, as readers would inevitably infer from these results (limited to ten years and one type of court: U.S. district courts) to a larger universe of present and future employee misclassification opinions. In this circumstance, statistical significance is also important.

Of course, readers may also choose to treat the regression results as purely descriptive, merely reporting on associations found within the set of 747 opinions, with attention to effect direction and magnitude only. The remainder of this report highlights the findings that are statistically significant; readers may examine the whole results set in Table 4.

Turning to the statistically significant coefficients, to begin, the pro se status and FLSA claim type variables were, respectively, positively and negatively associated with an independent contractor outcome, versus a plaintiff-friendly outcome. Specifically, holding all else constant, the probability of an independent contractor ruling increased by thirteen to sixteen percentage points when a plaintiff was pro se, as opposed to represented by an attorney.[50] This result is unsurprising, as past research has shown unrepresented parties to be at a disadvantage in litigation.[51]

With respect to the FLSA result, the independent contractor probability decreased by about twenty-two percentage points in cases with FLSA claims as compared to their non-FLSA counterparts. This result may stem from the underlying structure of the law, as the economic realities test used to judge employee status in FLSA cases is generally considered more worker-friendly than the alternative common law agency test.[52] However, as explored further below, this result may also reflect plaintiffs' lawyers' preferences for FLSA cases, and in particular cases brought by plaintiffs from occupations with high summary judgment success rates.

The remaining sets of statistically significant results – factors from the multi-factor tests, the written contract result, and plaintiffs' occupation – require a more in-depth treatment, as they may signal the courts' use of decisional shortcuts and the complex influence of selection effects, as plaintiffs' lawyers and judges engage in a self-reinforcing cycle of employee misclassification case selection and rulings.

B. Factor Variables[53]

With respect to the variables that represented the factors that comprise courts' various multi-factor tests, three were negatively associated with independent contractor rulings: the skill, permanency of the work relationship, and tax treatment variables.[54] One was positively associated with an independent contractor outcome: the plaintiff's own supervision of assistants.

These variables were derived from the caselaw described above in Part I, and were assembled by capturing keywords that a judge might use in analyzing each factor in the various multi-factor tests. Table 7 in the Appendix lists the keywords associated with each factor; the research team wrote code to tally each keyword, for each opinion. This method provides only a rough measure of the relative importance of these factors to outcomes, as the keyword lists may overlook some relevant words and the code may sweep in separate keyword usage that is unrelated to a judge's factor analysis. Nevertheless, these variables provide a rough indicator of the associations between factors and outcomes, as measured by keyword counts.[55]

---

[50] Because logistic regression coefficients take the form of log odds, they are difficult to interpret. The results reported above are therefore converted to marginal effects, holding all other independent variables at their means. For a simple tutorial on the computation of marginal effects using Stata, the statistical package used here, see Oscar Torres-Reyna, Princeton University Data and Statistics Services, *Predicted Probabilities and Marginal Effects After (Ordered) Logit/Probit Using* `margins` *in Stata,* http://dss.princeton.edu/training/.

[51] *See, e.g.,* Lois Bloom & Helen Hershkoff, *Federal Courts, Magistrate Judges, and the Pro Se Plaintiff*, 16 NOTRE DAME J.L. ETHICS & PUB. POL'Y 475 (2002) (collecting studies).

[52] *See, e.g.*, Nancy E. Dowd, *The Test of Employee Status: Economic Realities and Title VII*, 26 WM. & MARY L. REV. 75, 114 (1984).

[53] The term "factor variables" is not used here as it is in the Stata context, where it is a synonym for categorical variables, but instead refers to the variables that the research team created to represent the factors that courts use in their multi-factor tests for employee status.

[54] The skill variable was statistically significant only in the all-judge version of the regression.

[55] This factor analysis is also subject to the same circularity and endogeneity problem as noted above: predicting outcomes from features of the text may not be useful, as judges may have chosen those text features ex post, to justify the outcome. Nevertheless, if opinion text is dominated by one or more sets of keywords, and if the relevant factors are consistently associated with particular outcomes, this is still an interesting finding, as it suggests that judges may turn to a particular subset of factors to take the proverbial laboring oar in employee misclassification analysis.

Here, the negative coefficients on the first three variables mean that the more keywords present in an opinion that were associated with those three factors, the less likely a judge was to issue an independent contractor ruling. Specifically, for each additional appearance of the skill, permanency, and tax treatment keywords in a summary judgment opinion, the probability of an independent contractor ruling dropped by about one percentage point. The opposite was true for the supervision factor, where each additional related keyword increased the chances of an independent contractor ruling by about 0.7 percentage points.

Importantly, we cannot tell from these results how much or what types of worker skills were at issue in a misclassification dispute, for example, or how long "permanency" lasted; we can only suggest that judges' use of language related to those factors, holding all else equal, was important to the disputes' outcome. These findings suggest pathways for future research, to further refine the lists of keywords that are associated with the different sets of factors considered in the various multi-factor tests. If it emerges that only a subset of factors appears associated with outcomes, then a simplification of the legal tests may be in order, to improve clarity and reduce the "muck and mire" associated with current employee misclassification jurisprudence.

C. Written Contracts

This section now turns to the written contracts variable. As Table 7 in the Appendix shows, the research team constructed this variable by generating keyword counts per opinion of the following terms: clear and unambiguous, contract, contract language, independent contractor agreement, intention, signed, subjective belief, subjective intent, written agreement, and written contract. These were designed to identify passages of opinions that engaged with the presence of a written contract between the parties, purporting to designate the plaintiff as an independent contractor. The regression results then revealed this factor to be positively associated with independent contractor rulings, but the effect size is relatively small: for every additional relevant keyword appearing in an opinion, the probability of an independent contractor ruling increased by about 0.2 to 0.3 percentage points.

This finding squares with the results of the PI's previous study of employee misclassification decisions in discrimination cases. There, in cases in which the parties had a written contract, courts were more likely to rule that the plaintiff was properly classified as an independent contractor, and to point to the contract itself as definitive.[56]

These findings are somewhat perplexing, given caselaw that instructs courts to disregard the parties' contracts in the classification inquiry. Some courts, particularly in FLSA cases using the economic realities test, explicitly and strongly reject any reliance on contracts. As one court noted, quoting the Sixth Circuit, "the reason for looking past contractual arrangements 'is simple' [because] [t]he FLSA is designed to defeat rather than implement contractual arrangements."[57]

However, other formulations of the multi-factor tests allow courts to consider the parties' subjective intent to form an independent contractor or employee relationship.[58] The results in this and the PI's previous study suggest that courts may be relying on contracts as evidence of that

[56] Alexander, *supra* note 19.
[57] Wilson v. Guardian Angel Nursing, Inc., No. 3:07-0069., at *20 (M.D. Tenn. Jul. 31, 2008).
[58] *See, e.g.,* Jones v. Royal Admin. Servs., Inc., 887 F.3d 443, 450 (9th Cir. 2018) (considering, in a vicarious liability analysis, "the subjective intent of the parties" in structuring their work relationship).

intent. The presence of a written contract becomes a kind of decisional shortcut, where the definitional uncertainty around "true" employee status leads courts to view contracts, *ipso facto*, as evidence of true independent contractor status. Intellectual property scholar Barton Beebe has identified a similar phenomenon in trademark law, where, against the background of vague and unhelpful legal tests, courts adopt "'fast and frugal' heuristic[s] to short-circuit [the relevant] multi-factor analysis."[59]

Such rulings in the employee misclassification context are troubling, not least because of the possibility that workers might be pressured to sign sham contracts. More broadly, if courts accept the presence of a contract as proof of independent contractor status, they are essentially permitting workers to contract away the entire suite of labor and employment protections that are dependent on employee status – a body of law that is explicitly unwaiveable, even by its beneficiaries.

Alternatively, this regression result may be picking up on a correlation between the presence of written contracts and other factors that are probative of independent contractor status. In this circumstance, a judge's use of contract-related keywords would not signal anything untoward, if other factors in the analysis also revealed an independent contractor relationship. From the analysis conducted here, it is not possible to distinguish between the two scenarios, and more research is necessary to understand more clearly the role that written contracts play in the employee misclassification analysis.

D. Plaintiffs' Occupation

Turning to the final set of statistically significant results, plaintiffs' occupation, Table 4 treats occupation as a categorical variable, with one value assigned to the base, or reference, category: exotic dancers, the occupation with the lowest probability of an independent contractor ruling. Thus, the marginal effects of each occupation reported in the table on the likelihood of an independent contractor ruling should all be interpreted as relative to the likelihood of such a ruling when the plaintiffs' occupation was exotic dancer, holding all else constant. These results show that nearly all other occupations were more strongly associated with independent contractor outcomes than exotic dancers, by a margin of between thirteen and thirty-six percentage points. Table 5 below provides a more detailed look at the probability of an independent contractor ruling for each occupation, holding all else constant, among all judge rulings.

Table 5: Probability of Independent Contractor Ruling, by Occupation (All Judges)

Occupation	*Probability*
Other service occupation	0.497***
Administrative/ Office support worker	0.458***
Transportation/ Delivery driver	0.415***
Doctor/ Nurse/ Teacher/ Coach	0.387***
Direct sales/ Marketing worker	0.344***
Cleaning/ Maintenance worker	0.281***

[59] Barton Beebe, *An Empirical Study of the Multifactor Tests for Trademark Infringement*, 94 CAL. L. REV. 1581, 1581 (2006). Notably, the same phenomenon could be occurring with respect to the tax treatment variable discussed in the previous section: courts may be relying on employers' tax labeling as a stand-in for independent contractor status.

Occupation	*Probability*
Construction/ Agriculture/ Manufacturing worker	0.247***
Security guard	0.246**
Installer/ Technician	0.223***
Exotic dancer	0.0951*
Observations	747

NOTE*:* *** p<0.01, ** p<0.05, * p<0.1.

As Table 5 shows, all else equal, the probability of an independent contractor ruling ranged from forty-nine percent at the highest – other service occupations – to nine percent at the lowest – exotic dancers. Notably, Table 5 reveals a substantial drop-off between the exotic dancer probability and the next-highest: installer/technicians were more than twice as likely as exotic dancers to be held to be independent contractors; other service workers were more than five times as likely. These probabilities are displayed graphically, along with confidence intervals, in Figure 4 in the Appendix.

These results suggest that the same sorts of decisional shortcuts as explored above may be at work here, with judges looking to precedent involving workers in the same occupations and replicating those decisions. Some summary judgment decisions involving exotic dancers are actually quite explicit on this front. In a decision from the U.S. District Court for the Southern District of New York, for example, the judge first stated, "The application of the economic realities test is inherently case-specific, turning on the totality of the relevant circumstances."[60] Yet the very next sentence continues: "However, before embarking on an examination of how the factors identified in the Second Circuit's economic realities test apply to [the defendant in the case at hand], the Court notes that it is not the first court to address whether exotic dancers at a strip club such as [the defendant's] are employees under the FLSA. Nearly '[w]ithout exception, these courts have found an employment relationship and required the nightclub to pay its dancers a minimum wage.'"[61] The opinion then cites eleven different employee misclassification opinions involving exotic dancer plaintiffs from ten different courts, and concludes the paragraph by noting, "Defense counsel acknowledged at argument that the clear majority of cases have found exotic dancers to be employees under the FLSA."[62] Similar passages, and similar strings of occupation-specific precedent, can be found in other employee misclassification opinions.[63]

While the practice of finding factual similarities and distinctions within previous caselaw is central to legal reasoning, these courts' approach seems to rely more on the fact of the plaintiffs' shared occupations than mapping the details of one job onto another. Indeed, the court's note above that defense counsel had acknowledged the string of worker-friendly exotic dancer precedent suggests that the occupation label itself, rather than the particular job structures at issue, was doing the bulk of the work in the analysis. Of course, the structure of exotic dancing may be remarkably uniform across employers, and the occupation label may therefore serve as a useful decisional shortcut for courts. However, if courts rely too much on occupation as a proxy for employee or independent contractor status, their shortcuts risk undermining the purpose of the various legal tests, which is to engage with the factual minutiae of the work relationship,

[60] Hart v. Rick's Cabaret Int'l, Inc., 967 F. Supp. 2d 901, 912 (S.D.N.Y. 2013).
[61] *Id.*
[62] *Id.*
[63] *See, e.g.,* Brown v. Mustang Sally's Spirits and Grill, Inc. No. 12-CV-529S (W.D.N.Y. Sept. 12, 2013) (citing eight cases from eight different courts).

rather than rest on labels. Moreover, it is notable that many of these cases were brought by individual workers, rather than classes. In this context, heavy reliance on the collective label that is occupation requires even more caution, as the larger occupational structure may crowd out the details of the individual plaintiffs' work relationships.

In addition to decisional shortcuts, a second, related, phenomenon may help explain the significance of the occupation variables: selection effects. Here, it is important to acknowledge the powerful gatekeeping role that plaintiffs' lawyers play in choosing which lawsuits to bring to court. As the PI has explored in previous work, lawyers assess the viability of the claims that potential clients could bring, and select for the ones that appear most likely to win in court or generate a high-dollar settlement.[64] Research on litigation patterns in other areas of employment law, for example, has revealed industry-based targeting behavior on the part of plaintiffs' lawyers, as they screen for potential claims, and even actively advertise to recruit clients who work in particular industries known for violations. These patterns are evident in coordinated minimum wage and overtime litigation targeting public school employers and the oil and gas industries.[65]

The results of the present research suggest a similar phenomenon, in which plaintiffs' lawyers may prefer employee misclassification cases brought by exotic dancers over those by other types of workers. In this conception, these lawyers' targeting is then rewarded by the courts, which rule especially favorably in those case types. From the perspective of plaintiffs' lawyers, the win rate differential may send a clear message: that occupation can be a proxy for the likelihood of success on the employee misclassification question, and that lawyers should make case selection decisions accordingly.[66]

To further add to this story, the graph in Figure 4 shows the number of employee misclassification decisions from all judges in the data set for exotic dancers, as well as for the top three independent contractor operations as listed in Table 5, combined, as well as for all other occupations, combined. The trend lines are indexed to a common starting point, and cover 2009-2017, the years in which opinions were available for all occupations.

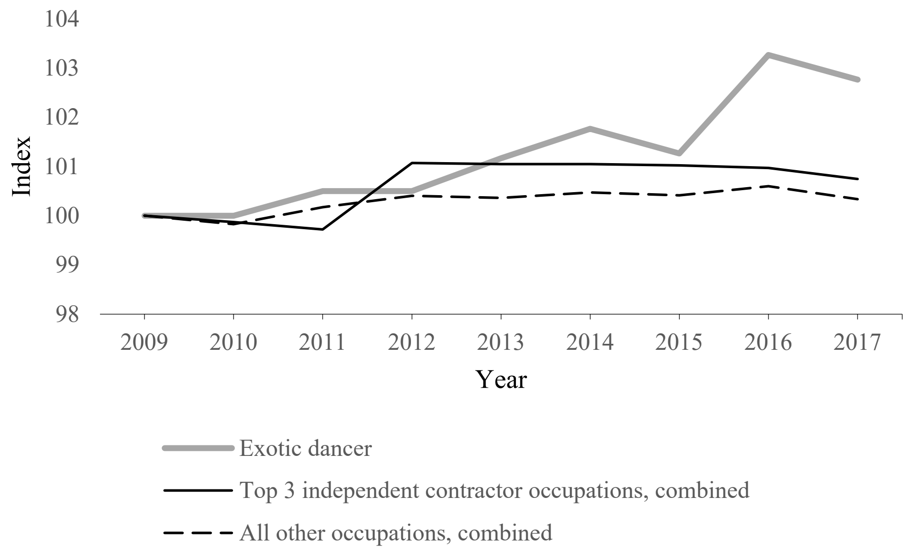
---

[64] Charlotte S. Alexander, *Litigation Migrants*, __ AM. BUS. L.J. __ (forthcoming 2019). John Coffee is the leader in studying the behavior of plaintiffs' lawyers, having published numerous articles since the 1980s and a recent book that explores and evaluates the role of entrepreneurial private attorneys in securities litigation. *See, e.g.,* JOHN C. COFFEE, JR., ENTREPRENEURIAL LITIGATION: ITS RISE, FALL, AND FUTURE (2015) (exploring private attorneys' role in securities litigation); *see also* David Freeman Engstrom, *Private Enforcement's Pathways: Lessons from Qui Tam Litigation*, 114 COLUM. L. REV. 1913 (2014) (exploring private attorneys' role in qui tam litigation); David A. Hyman & Charles Silver, *Medical Malpractice Litigation and Tort Reform: It's the Incentives, Stupid*, 59 VAND. L. REV. 1085, 1120 (2006) (exploring private attorneys' role in medical malpractice litigation); Myriam Gilles & Gary B. Friedman, *Exploding the Class Action Agency Costs Myth: The Social Utility of Entrepreneurial Lawyers*, 155 U. PA. L. REV. 103 (2006) (exploring private attorneys' role in civil rights litigation).

[65] Alexander, *supra* note 64.

[66] Though plaintiffs' law firms were not in the data the research team analyzed, there may even be discernable patterns in which certain firms specialize in certain occupations.

Figure 4: Number of Employee Misclassification Decisions by Occupation, Indexed (All Judges)



Admittedly, the chart in Figure 4 should not be the basis for causal conclusions. However, it does suggest the need for additional, future research on the hypothesis that plaintiffs' lawyers and judges are in conversation with one another about the occupation-based viability of plaintiffs' claims. The gray line, representing exotic dancers, shows a general upward trajectory throughout the time period, whereas the solid black and dotted lines, representing the top three independent-contractor heavy occupations, as well as all other occupations combined, stay relatively flat or slope slightly downward for the latter years in the data set.[67] Notably, this chart presents a very narrow slice of the whole picture, omitting misclassification disputes that do not have a summary judgment opinion, or that settle. Nevertheless, the increase in decision numbers in the exotic dancer category may suggest that plaintiffs' lawyers are receiving the viability signals sent by the courts, and then reflecting them back in the form of additional litigation. This over-selection of exotic dancer cases – all but one of which were brought under the FLSA – may also explain the FLSA's own substantial representation within the data set, representing over two-thirds of all cases. The authors are continuing to explore these possibilities in ongoing work.

Thus, recalling the discussion above regarding decisional shortcuts, judges' proxy for viability – plaintiffs' occupation – may also become lawyers' in a self-reinforcing pattern. Misclassification law, taken as a whole, then becomes a commentary on the structure of occupations, with courts approving certain work forms and censuring others. Of course, this is exactly what misclassification law is meant to do: to prompt courts to check employers' power to classify. However, if courts rely too heavily on decisional shortcuts –whether occupations or written contracts – they risk elevating employers' choices to the status of law. Lauren Edelman has observed a similar phenomenon in her extensive study of courts' treatment of employers' internal antidiscrimination policies.[68] She observes that when courts "infer fair practices" based on "the presence of antidiscrimination policies[,] personnel manuals, [and] diversity training programs," they run the same proxy-related risk identified here, in that they "wind up condoning

[67] Notably, all lines slope downward slightly in 2016. It is unclear whether this is an artefact of the data collection process used in this project, or an indication that fewer employee misclassification disputes have been resolved on summary judgment in recent years.

[68] LAUREN EDELMAN, WORKING LAW: COURTS, CORPORATIONS, AND SYMBOLIC CIVIL RIGHTS (2016).

practices that deviate considerably from the legal ideals" without interrogating the facts of those practices on the ground.

As this report suggests, courts may be more tempted to rely on decisional shortcuts where, as in the employee misclassification context, the relevant legal tests fail to provide clear and useful guidance. This is also the case in the context that Edelman studies, where the legal standards around judges' review of employer personnel policies are ambiguous. However, courts should be attuned to the risks associated with such heuristics, especially when courts' signals about viability are received by plaintiffs' lawyers, and then reflected back in the form of increased case filing of the very same type.

IV. Conclusion

The research summarized in this report was designed to fill a hole in the literature on employee misclassification by studying the text of judges' summary judgment opinions, in which courts determine whether a worker is an employee or an independent contractor, and explain their reasoning. After confronting substantial data access problems, the research team assembled a set of 747 opinions issued by all U.S. district courts in the years 2008-2017, and extracted a picture of judges' employee misclassification decision-making from the text. In doing so, this project addressed the set of research questions listed at this report's outset, exploring the win/loss rate for plaintiffs who claim misclassification; the legal tests used, which were highly correlated with claim type; the factors associated with independent contractor rulings as opposed to plaintiff-friendly outcomes; and other party and judge-related variables associated with plaintiff wins and losses.

Future work will return to the themes developed here, exploring the possibility that judges may rely on a subset of factors as decision-making heuristics or proxies to manage the ambiguity introduced by statutory definitions and the caselaw's multi-factor tests. Future work will also continue to investigate the selection effects theory suggested here, where courts' decisions influence plaintiffs' lawyers' own decisions about claim viability, producing a feedback loop in which certain occupations – or industries, claim types, or fact patterns – are heavily selected for litigation.

APPENDIX

Table 6: Summary Statistics

Variable	*Frequency*	*Percent*
Outcome		
Denial of summary judgment	230	30.8%
Employee	236	31.6%
Independent contractor	281	37.6%
Decision year		
2008	27	3.6%
2009	60	8.0%
2010	51	6.8%
2011	59	7.9%
2012	91	12.2%
2013	90	12.0%
2014	98	13.1%
2015	90	12.0%
2016	105	14.1%
2017	76	10.2%
Court		
C.D. Cal.	9	1.2%
C.D. Ill.	1	0.1%
D. Alaska	1	0.1%
D. Ariz.	9	1.2%
D. Colo.	8	1.1%
D. Conn.	10	1.3%
D. Del.	4	0.5%
D. Haw.	1	0.1%
D. Idaho	4	0.5%
D. Kan.	11	1.5%
D. Mass.	10	1.3%
D. Md.	24	3.2%
D. Me.	3	0.4%
D. Minn.	1	0.1%
D. Mont.	1	0.1%
D. Nev.	5	0.7%
D. Ore.	14	1.9%
D. Utah	5	0.7%
D.D.C.	9	1.2%
D.N.D.	2	0.3%
D.N.J.	30	4.0%

Variable	*Frequency*	*Percent*
D.N.M.	5	0.7%
D.P.R.	8	1.1%
D.S.C.	12	1.6%
D.S.D.	1	0.1%
D.V.I.	2	0.3%
E.D. Ark.	10	1.3%
E.D. Cal.	9	1.2%
E.D. Ky.	4	0.5%
E.D. La.	7	0.9%
E.D. Mich.	16	2.1%
E.D. Mo.	23	3.1%
E.D. Okla.	1	0.1%
E.D. Pa.	12	1.6%
E.D. Tenn.	2	0.3%
E.D. Tex.	5	0.7%
E.D. Va.	7	0.9%
E.D. Wis.	8	1.1%
E.D.N.C.	6	0.8%
E.D.N.Y.	29	3.9%
M.D. Ala.	3	0.4%
M.D. Fla.	29	3.9%
M.D. Ga.	3	0.4%
M.D. Pa.	6	0.8%
M.D. Tenn.	9	1.2%
N.D. Ala.	7	0.9%
N.D. Cal.	21	2.8%
N.D. Fla.	7	0.9%
N.D. Ga.	16	2.1%
N.D. Ill.	38	5.1%
N.D. Ind.	7	0.9%
N.D. Iowa	2	0.3%
N.D. Miss.	8	1.1%
N.D. Ohio	19	2.5%
N.D. Okla.	3	0.4%
N.D. Tex.	9	1.2%
N.D. W. Va.	4	0.5%
N.D.N.Y.	5	0.7%
S.D. Ala.	3	0.4%
S.D. Cal.	12	1.6%
S.D. Fla.	56	7.5%
S.D. Ga.	1	0.1%
S.D. Ill.	1	0.1%
S.D. Ind.	10	1.3%

Variable	*Frequency*	*Percent*
S.D. Iowa	4	0.5%
S.D. Miss.	9	1.2%
S.D. Ohio	6	0.8%
S.D. Tex.	30	4.0%
S.D. W. Va.	2	0.3%
S.D.N.Y.	26	3.5%
W.D. Ark.	3	0.4%
W.D. Ky.	3	0.4%
W.D. La.	3	0.4%
W.D. Mich.	6	0.8%
W.D. Mo.	6	0.8%
W.D. Okla.	2	0.3%
W.D. Pa.	3	0.4%
W.D. Tenn.	2	0.3%
W.D. Tex.	7	0.9%
W.D. Va.	5	0.7%
W.D. Wash.	8	1.1%
W.D. Wis.	3	0.4%
W.D.N.C.	7	0.9%
W.D.N.Y.	4	0.5%
Circuit		
1	21	2.8%
2	74	9.9%
3	57	7.6%
4	67	9.0%
5	78	10.4%
6	67	9.0%
7	68	9.1%
8	52	7.0%
9	94	12.6%
10	35	4.7%
11	125	16.7%
DC	9	1.2%
Case type		
Non-FLSA	245	32.8%
FLSA	502	67.2%
Pro se status		
Pro se plaintiff	54	7.2%
Represented plaintiff	693	92.8%

Variable	*Frequency*	*Percent*
Plaintiff occupation		
Administrative/Office support worker	42	5.6%
Cleaning/Maintenance worker	46	6.2%
Construction/ Agriculture/ Manufacturing worker	76	10.2%
Direct sales/ Marketing worker	111	14.9%
Doctor/ Nurse/ Teacher/ Coach	90	12.0%
Exotic dancer	47	6.3%
Installer/ Technician	71	9.5%
Other service occupation	114	15.3%
Security guard	18	2.4%
Transportation/Delivery driver	132	17.7%
Judge type		
District	665	89.0%
Magistrate	82	11.0%
Judge birth year		
1920	2	0.3%
1921	3	0.4%
1924	7	0.9%
1925	4	0.5%
1926	5	0.7%
1927	2	0.3%
1929	3	0.4%
1930	2	0.3%
1932	4	0.5%
1933	3	0.4%
1934	16	2.1%
1935	11	1.5%
1936	10	1.3%
1937	12	1.6%
1938	5	0.7%
1939	9	1.2%
1940	17	2.3%
1941	11	1.5%
1942	17	2.3%
1943	12	1.6%
1944	15	2.0%
1945	17	2.3%
1946	25	3.3%
1947	29	3.9%
1948	24	3.2%
1949	36	4.8%

Variable	*Frequency*	*Percent*
1950	31	4.1%
1951	45	6.0%
1952	42	5.6%
1953	24	3.2%
1954	26	3.5%
1955	28	3.7%
1956	18	2.4%
1957	14	1.9%
1958	8	1.1%
1959	6	0.8%
1960	18	2.4%
1961	9	1.2%
1962	23	3.1%
1963	7	0.9%
1964	6	0.8%
1965	10	1.3%
1966	10	1.3%
1967	10	1.3%
1968	6	0.8%
1969	12	1.6%
1971	4	0.5%
1972	4	0.5%
1975	3	0.4%
Unknown (magistrates)	82	11.0%
Judge gender		
Female	228	30.5%
Male	519	69.5%
Political party of judge's appointing President		
Democratic	336	45.0%
Republican	329	44.0%
N/A (magistrates)	82	11.0%
Judge commission year		
1962	1	0.1%
1967	2	0.3%
1970	1	0.1%
1971	4	0.5%
1977	3	0.4%
1978	2	0.3%
1979	17	2.3%
1980	2	0.3%

Variable	*Frequency*	*Percent*
1981	8	1.1%
1982	7	0.9%
1983	7	0.9%
1984	7	0.9%
1985	14	1.9%
1986	11	1.5%
1987	7	0.9%
1988	8	1.1%
1989	7	0.9%
1990	22	2.9%
1991	11	1.5%
1992	35	4.7%
1993	11	1.5%
1994	25	3.3%
1995	23	3.1%
1996	7	0.9%
1997	18	2.4%
1998	31	4.1%
1999	25	3.3%
2000	14	1.9%
2001	9	1.2%
2002	25	3.3%
2003	32	4.3%
2004	28	3.7%
2005	11	1.5%
2006	23	3.1%
2007	25	3.3%
2008	25	3.3%
2010	29	3.9%
2011	50	6.7%
2012	38	5.1%
2013	12	1.6%
2014	25	3.3%
2016	1	0.1%
2018	2	0.3%
Unknown (magistrates)	82	11.0%

Table 7: Keyword Cluster Assembly

Keywords	*Cluster*
annual leave	benefits
benefits	benefits

Keywords	*Cluster*
insurance	benefits
retirement	benefits
social security	benefits
control	control
duration	duration
job length	duration
length of job	duration
permanency	duration
tenure	duration
in business	independence
opportunity for loss	independence
opportunity for profit	independence
facilities	location
location	location
method of payment	method of payment
regular business	regular business
skill	skill
specialist	skill
supervision	supervision
assistant	supervision
termination	permanency
permanence	permanency
equipment	tools
supplies	tools
Tools	tools
1099	tax treatment
tax treatment	tax treatment
w2	tax treatment
w-2	tax treatment
work day	tax treatment
workday	tax treatment
clear and unambiguous	written contract
contract	written contract
contract language	written contract
independent contractor agreement	written contract
intention	written contract
Signed	written contract
subjective belief	written contract
subjective intent	written contract
written agreement	written contract
written contract	written contract

NOTE*:* Keywords were stemmed and lemmatized in order to capture all singular/plural forms and other variations in the counts.

Table 8: Machine Learning Classifiers: Outcome Classification

Method	*Sensitivity*	*Specificity*	*Accuracy*	*True positives*	*True negatives*	*False positives*	*False negatives*
cite	30.0%	90.2%	67.5%	17	85	9	40
logistic	57.7%	73.1%	67.3%	33	69	25	24
logistic.cite	65.3%	75.6%	71.7%	37	71	23	20
dT	41.9%	83.5%	67.8%	24	78	16	33
dT.cite	46.2%	83.0%	69.1%	26	78	16	31
rf	60.0%	86.2%	76.3%	34	81	13	23
rf.cite	58.7%	84.2%	74.5%	33	79	15	24
lda	47.0%	83.9%	70.0%	27	79	15	30
lda.cite	61.2%	80.4%	73.2%	35	76	18	22
qda	67.3%	60.9%	63.3%	38	57	37	19
qda.cite	68.0%	71.3%	70.0%	39	67	27	18
knn	51.3%	75.3%	66.2%	29	71	23	28
knn.cite	57.9%	77.7%	70.2%	33	73	21	24

NOTE: Table 2 and the surrounding text contain an explanation of the classifiers contained in this table.

Table 9: Random Forest Variable Importance

Variable	*Mean Decrease in Gini Coefficient*
Written contract	39.71369185
Factor: benefits	32.08165643
Factor: control	29.53894009
Factor: permanency	28.74444579
FLSA case	27.03592345
Factor: supervision	23.86619853
Factor: skill	23.51876859
Factor: tools	22.98105695
Factor: independence	17.43266658
Factor: duration	16.40272228
Factor: location	15.21166107
Factor: tax treatment	13.43705324
Factor: regular business	12.53838706
Factor: method of payment	7.523325496
Occupation: other service	6.388093706
Occupation: transportation/delivery	5.470640836
Occupation: direct sales/marketing	3.950765087

Variable	*Mean Decrease in Gini Coefficient*
Occupation: installer/technician	3.292881436
Occupation: construction/agriculture/manufacturing	3.263834581
Occupation: doctor/nurse/teacher/coach	3.246682583
Occupation: exotic dancer	3.166820799
Pro se plaintiff	3.106009936
Occupation: cleaning/maintenance	1.738089157
Occupation: security guard	0.724805095

Figure 4: Probability of Independent Contractor Ruling, by Occupation (All Judges), With 95% Confidence Intervals