

Testing for Change in Procedural Standards, with Application to *Bell Atlantic v. Twombly*

[Forthcoming, *Journal of Legal Studies*]

William H.J. Hubbard*

whubbard@uchicago.edu

September 2012

Abstract

Quantifying change in legal standards—in the sense of change in the propensity of judges to decide cases a certain way—presents a vexing problem. In response to a change in the behavior of courts, plaintiffs and defendants will adjust their patterns of filing and settling cases. Models of the selection of disputes for litigation predict that when legal standards change, the rate at which plaintiffs prevail in litigation will not predictably change; if so, changes in legal standards cannot be measured with data on court outcomes. I consider both the Priest and Klein (1984) divergent expectations and Bebchuk (1984) asymmetric information models to develop a methodology for measuring changes in procedural standards in the presence of selection effects. I apply this methodology to *Bell Atlantic Corp. v. Twombly*, whose effects on dismissal

*William H.J. Hubbard is an Assistant Professor of Law at the University of Chicago Law School. I am grateful for comments from Amanda Agan, Douglas Baird, Gary Becker, Omri Ben-Shahar, Emily Buss, Joe Cecil, Lee Epstein, Jonathan Hall, Cheryl Hubbard, Louis Kaplow, Ashley Keller, Bill Landes, Ethan Lieber, Anup Malani, Tom Miles, Emily Oster, Mark Phillips, Richard Posner, Ashish Prasad, Luke Threinen, and Andrew Zuppann; participants at the University of Chicago Law School Work-In-Progress Workshop, the USC Gould School of Law CLEO Workshop, the 2011 Conference on Empirical Legal Studies, and the Harvard Law School Law and Economics Seminar; and an anonymous referee. I thank Lee Epstein, Bill Landes, and Richard Posner for providing me with their dataset of published opinions on motions to dismiss. Dylan Benson, Tirsit Berhanu, Liz Chao, Helen Chen, Jessica Dueck, and Georgia Sampedro provided invaluable assistance in preparing this dataset for analysis.

rates have been subject to considerable speculation and debate. I find that *Twombly* precipitated no significant change in dismissal rates, even after accounting for selection effects.

1 INTRODUCTION

How does one measure whether a legal standard has changed? There is of course a formal sense in which the law changes every time Congress passes a new statute or the Supreme Court decides a new case. This paper, however, concerns itself with legal change in the following sense: change in the outcomes of cases. It is this latter sense in which the law affects the expectations that individuals form about the risks of liability for different activities, and is thus the primary means by which the law influences human behavior.¹

This invites the question of whether, say, a Supreme Court opinion will actually change the outcomes of cases. In other words, given two cases, identical in every way except the date on which they are decided, would the one decided after the new rule have a different outcome than the one decided before? Answering this question empirically requires us to hold fixed the set of litigated cases before and after the potential law change. But in response to (for example) a new Supreme Court decision, plaintiffs and defendants may change their litigation strategy, settling cases that previously would have been litigated, and litigating cases that previously would have settled. Because the composition of the set of decided cases has changed, the rate at which plaintiffs or defendants prevail in litigation may not change, even after a sharp change in how courts decide cases. This leaves us with the problem of how to test for changes in legal standards.

Priest and Klein (1984) were the first to develop this argument. While their seminal

¹Holmes (1897) put it colorfully: “The reason why [law] is a profession, why people will pay lawyers to argue for them or to advise them, is that in societies like ours the command of public force is intrusted to the judges in certain cases People want to know under what circumstances and how far they will run the risk of coming against what is so much stronger than themselves, and hence it becomes a business to find out when this danger is to be feared.”

paper is most famous for its prediction that (subject to important qualifications) plaintiffs will prevail approximately 50 percent of the time in litigation, no less important is their point that (*not* subject to qualifications) a change in a legal standard will have no predictable effect on the rate at which plaintiffs win in court. Indeed, while the 50 percent prediction has been challenged on empirical and theoretical grounds (see, e.g., Shavell 1996), it has been broadly recognized since Priest and Klein (1984) that selection effects may make decided cases unrepresentative of all cases, and therefore changes in the rates at which plaintiffs prevail in litigation may not be reliable indicators of changes in the legal standards governing their cases.

Take *Bell Atlantic Corp. v. Twombly* (550 U.S. 544 [2007]). In *Twombly*, the Supreme Court addressed the standard under which a judge decides whether to dismiss a case for failure to state a claim. This rule applies in every federal civil case, and thus applies to approximately 250,000 newly filed cases per year. The previous standard for dismissing a case had been announced in *Conley v. Gibson* (355 U.S. 41 [1957]), and for fifty years, the Supreme Court had never altered the standard. But in *Twombly*, the Supreme Court unexpectedly “retired” language from *Conley* and announced a new standard that, by its terms, made it easier to dismiss a case.

There was an immediate uproar from the bar and legal academe. “No decision in recent memory has generated as much interest and is of such potentially sweeping scope as the Supreme Court’s decision in *Bell Atlantic Corp. v. Twombly*.” Smith (2009). Countless commentators criticized its (presumably) harmful effect on plaintiffs seeking relief in court.² Bills were introduced in Congress to overturn the decision. Notice Pleading Restoration Act of 2009, S. 1504, 111th Cong. (2009); Open Access to Courts Act of 2009, H.R. 4115, 111th Cong. (2009). Lawyers and academics testified that *Twombly* has had a “devastating impact” on plaintiffs, particularly in civil rights cases. See Schneider (2010).

But was *Twombly* in fact such a momentous legal change? Did it affect many cases?

²See Part 2.2 for citations.

Any cases? How can we know? If the Priest and Klein model is right, we might expect to see plaintiffs prevailing on a motion to dismiss (MTD) about 50 percent of the time before *Twombly*, and about 50 percent of the time after *Twombly*—even if *Twombly* radically changed the pleading standard. In a sample of nearly 13,000 published decisions on MTDs, this is exactly what I find. See Table 1. This could mean that *Twombly* had little effect, or it could mean that changes in case selection have masked the effect of *Twombly*. And if one cannot tell “change” from “no change,” then our enterprise, quantifying the effects of new legal rules, is doomed.

Well, not quite. In this paper, I will show that, at least under certain circumstances, empirical tests for a change in a legal standard can control for selection effects, allowing us to distinguish the finding of “a change in legal standard masked by selection effects” from “no change.” This paper uses the *Twombly* case to show how this methodology can work.

My first contribution is to highlight the fact that selection effects can occur at multiple stages of litigation. Although styled “The Selection of Disputes for Litigation,” Priest and Klein (1984) does *not* address the selection of *disputes* (e.g., car crashes or broken promises) for *litigation* (i.e., filed lawsuits). Instead, Priest and Klein present a model of the selection of *filed lawsuits* for *adjudication*.³ The literature on the Priest and Klein model has largely followed suit, either treating the set of filed lawsuits as fixed and exogenous (e.g., Priest and Klein 1984; Waldfogel 1995; Siegelman and Waldfogel 1999) or assuming that all suits with positive expected value are filed (e.g., Wittman 1985; Priest 1985; Siegelman and Donahue 1995).⁴ In other words, this literature tends to ignore selection of disputes into litigation, but addresses selection among filed cases.

In contrast, some recent empirical work studying the effects of various statutes and judicial decisions on case outcomes has acknowledged that the set of filed cases might change

³I will follow Waldfogel (1995) and call the resolution of a case through either dispositive motion or trial “adjudication.” In selection models, one can think of “settlement” and “adjudication” as the alternatives.

⁴By positive expected value, I mean the expected judgment award to the plaintiff exceeds the plaintiff’s expected litigation costs.

endogenously in response to a new statute or judicial decision. This work attempts to control for selection by limiting the data to cases filed before the new legal standard took effect. See, for example, Choi and Pritchard (2011). But this literature has essentially ignored selection effects among filed cases.

This paper addresses *both* the selection of disputes into lawsuits and the selection of lawsuits into adjudication. I do this in the context of an empirical examination of the effects of *Twombly* on the dismissal rates of federal cases.

My second contribution is to address selection effects in a way that incorporates *both* of the canonical models of selection into litigation: the divergent expectations (DE) framework in Priest and Klein (1984) and the asymmetric information (AI) framework introduced by Bebchuk (1984). I will show that the logic of the divergent expectations model suggests that there will be no predictable change in the rate at which plaintiffs prevail on MTDs; but by controlling for the selection of disputes into lawsuits and then accounting for the selection of lawsuits into litigated MTDs, one can identify a change in the legal standard even under the conditions of the Priest and Klein model. And while it is true, as Shavell (1996) famously observed, that under asymmetric information “any frequency of plaintiff victory at trial is possible,” I will show that the asymmetric information model *does* make testable predictions about how the frequency of plaintiff victory *changes* in response to a change in the legal standard. In short, I will argue that a fairly simple empirical methodology can test the prediction that *Twombly* increased the likelihood that a district court will dismiss a case, even given selection effects under either the divergent expectations or asymmetric information model.

With this methodology in hand, I test three hypotheses. As noted above, at the time it was decided, many commentators assumed that *Twombly* was a major change in the law of pleading and therefore would precipitate a rise in both the rate at which MTDs are granted and the total number of dismissals. I will call this view the *conventional hypothesis*. The

conventional hypothesis does not account for selection effects in litigation.

Other commentators took the position that *Twombly* caused at most a modest change in how judges decide cases, and therefore its effects on dismissals would be modest. This is the null or *no change hypothesis*.

I consider the hypothesis that *Twombly* would have a large effect on the outcomes of cases if the mix of cases did not change in response to *Twombly*, but that observing this empirically is complicated by selection effects. I will show that despite selection effects, this hypothesis generates testable predictions. Given the empirical strategy that I develop herein, both the DE and AI selection models predict that if *Twombly* raised the bar for MTDs, then although the rate at which plaintiffs prevail on MTDs may not (predictably) change, the total number of dismissals (measured as a fraction of all filed cases) will rise. I call this the *selection/change hypothesis*. This new hypothesis has a “no change” version, but is identical to the original no change hypothesis: if *Twombly* wrought no (significant) change, then we will see no (significant) changes in the rates at which MTDs are granted or in the total number of dismissals.

To test these predictions, I bring two new datasets to bear. These have two important advantages over previously used datasets. First, they are very large. One contains information on outcomes for nearly 13,000 published district court opinions ruling on MTDs between May 21, 2006 and May 21, 2008 (a year before and after *Twombly*). The other is administrative data on over 250,000 federal civil lawsuits filed in the 2005–2008 period. Second, I have the filing date for every case in these datasets, allowing me to identify cases filed before *Twombly* but adjudicated after *Twombly*. This is crucial to my empirical strategy. Using these two datasets, I test the competing hypotheses, and find no support for the view that *Twombly* had a large effect. Instead, the data support the view that *Twombly* did not change the likelihood that district courts would dismiss cases for failure to state a claim.

In sum, this paper makes two contributions. First, and most broadly, it presents a

methodology for measuring changes in pretrial procedural standards that accounts for the ways in which actors in the legal system respond endogenously to changes in the legal environment, controlling for selection among disputes both pre- and post-filing. Second, and more narrowly, it helps resolve a contentious debate about one of the most controversial civil procedure decisions in a generation: *Bell Atlantic v. Twombly*.

The balance of this paper appears as follows: Part 2 describes *Twombly* and briefly addresses the large literature that it has spawned. Part 3 lists the central empirical predictions made about the effects of *Twombly*. It then develops a methodology for accounting for selection effects when testing these predictions. Part 4 describes the two data sets I use. Part 5 presents my estimation specifications and results. I find no evidence that *Twombly* marked a change in the law of pleading. The most “sweeping” Supreme Court decision “in recent memory,” it appears, has had little effect on the outcomes of cases. Part 6 concludes with a discussion of the limitations of my approach and a discussion of the ways in which *Twombly*, despite having no observable effect in the data, could nonetheless have been an important change with respect to certain types of cases.

2 THE STORY SO FAR

2.1 From *Conley* to *Twombly*

Civil actions in federal court begin with the filing of a complaint by the plaintiff. Once a plaintiff files a complaint, the defendant can file a MTD for “failure to state a claim upon which relief can be granted” under Federal Rule of Civil Procedure (“Rule”) 12(b)(6).⁵ If the court denies the MTD, the case proceeds to discovery and perhaps settlement or trial. If the court grants the motion, the case is over; the plaintiff’s only recourse is appeal to a

⁵Herein, MTD refers specifically to a Rule 12(b)(6) motion to dismiss; of course, there are other grounds for dismissal, such as lack of jurisdiction or failure to prosecute the claim, which are not relevant here.

higher court.⁶

For 50 years, the standard governing a Rule 12(b)(6) motion seemed unassailable. *Conley v. Gibson* (355 U.S. 41, 45–46 [1957]) described “the accepted rule that a complaint should not be dismissed for failure to state a claim unless it appears beyond doubt that the plaintiff can prove no set of facts in support of his claim which would entitle him to relief.” On its face, this statement seems to imply that a complaint would survive a MTD so long as the plaintiff did not plead facts that contradicted his legal claim. While perhaps never interpreted quite so literally, *Conley* reflected a liberal approach to the principle of “notice pleading” embodied by the Federal Rules. In the decades after *Conley*, “pleading standards were widely viewed as ‘well established’ and ‘relatively straightforward’” (Steinman 2010, p. 1295), so much so that “commentators lost interest in pleading” (Marcus 1986, p. 434).

And then, on May 21, 2007, the Supreme Court decided an antitrust dispute captioned *Bell Atlantic Corp. v. Twombly* (550 U.S. 544 [2007]). After reciting *Conley*’s famous “no set of facts” standard, the Court announced that “after puzzling the profession for 50 years, this famous observation has earned its retirement.” (544 U.S. at 563.) In its place, the Court required that a complaint plead facts “plausibly suggesting (not merely consistent with)” the plaintiff’s legal claim. (544 U.S. at 557.) It concluded: “we do not require heightened fact pleading of specifics, but only enough facts to state a claim to relief that is plausible on its face. Because the plaintiffs here have not nudged their claims across the line from conceivable to plausible, their complaint must be dismissed.” (544 U.S. at 570.)

The Supreme Court itself would reaffirm *Twombly* two years later. In *Ashcroft v. Iqbal* (556 U.S. 662 [2009]), the Court reiterated the rule from *Twombly* that “only a complaint that states a plausible claim for relief survives a motion to dismiss.” (Id. at 679.) The Court also noted that *Twombly* was not limited to antitrust cases. (Id. at 684.) This confirmed the conclusion that both courts and academics had already reached about the scope of *Twombly*:

⁶Here, and throughout the paper, I refer to dismissals *with prejudice* when describing the granting of a MTD.

it applied to all cases. See Bone (2009, p. 881); Hatamyar (2010, p. 555); Redish and Epstein (2008, p. 26).

2.2 The Aftermath

Twombly was a bombshell. The viability of *Conley* was not part of the question presented to the Supreme Court; it had not even been raised by the parties. *Twombly* struck “[s]eemingly without warning” (Hatamyar 2010, p. 554) and “sent shockwaves through the legal community—for academics, practitioners, and judges alike” (Steinman 2010, p. 1305 (citations omitted)).

The response to *Twombly* was immediate. Almost overnight, it became a fixture in judicial opinions,⁷ and after 50 years of near-dormancy, the scholarly literature on pleading exploded.⁸ As Reinert (2011, p. 3) notes, a near-consensus emerged “among academic observers that the *Iqbal/Twombly* pleading standard marks a sharp break with the past.” Most scholars view *Twombly* “as overturning fifty years of generous notice pleading practice, and critics attack it as a sharp departure from the ‘liberal ethos’ of the Federal Rules, favoring decisions ‘on the merits, by jury trial, after full disclosure through discovery.’” (Bone 2009, p. 875.)

Whether this development is a good thing or a bad thing has been the center of contentious debate. “Without question, *Bell Atlantic v. Twombly*[] ranks as one of the most controversial decisions of the United States Supreme Court in recent years.” (Epstein 2009, p. 2.) At base, normative arguments about *Twombly* stem from a single, empirical premise: rightly or wrongly, *Twombly* will cause more cases to be dismissed—a lot more cases. Miller

⁷*Twombly* has been cited in judicial opinions tens of thousands of times and is already one of the most cited decisions in the history of the United States. See, e.g., Steinman 2010.

⁸Stancil (2009, pp. 137–138) writes: “Until recently, the scholarly literature on pleading standards was remarkably thin, with only a few significant pieces written from the 1930s through the early 2000s. Widespread scholarly interest in pleading is a remarkably recent phenomenon, tracing its birth to the Supreme Court’s 2007 opinion in *Bell Atlantic Corp. v. Twombly*.” For a litany of citations, see Steinman (2010, pp. 1296–1298 and nn.10–14).

(2010, pp. 21) sums up that “the perception among many practicing attorneys and commentators is that the grant rate [for MTDs] has increased.”⁹

Although *Twombly* was an antitrust case, the greatest concerns raised have been about its effect on civil rights plaintiffs. Reinert (2011) and others note the historical importance of federal courts to civil rights claimants, as well as the particular difficulty that individuals with claims such as employment discrimination may have pleading specific facts about the defendant’s motives. This has led to predictions that *Twombly* will disproportionately increase dismissals of civil rights claims (see, e.g., Ward 2008; Miller 2010) and that “the shift is undeniably big” (Schneider 2010).

Of course, there have been dissenting voices. Some scholars have argued that, as a doctrinal matter, *Twombly* does not represent a sharp break with precedent (Epstein 2009; Steinman 2010), or is only a modest change in doctrine (Bone 2009). These commentators go on to predict that, consequently, the observed effect of *Twombly* will not be dramatic.

What is remarkable about this debate is that despite widespread disagreement about the meaning of *Twombly* as a matter of legal doctrine, all sides of the debate (implicitly) agree that whether or not *Twombly* worked a major change in the law will ultimately be resolved as an *empirical* matter: we can measure its effect on the law of civil procedure by measuring how it changes the rate at which cases are dismissed.

2.3 The Evidence to Date

A small literature has emerged that attempts to test whether *Twombly* has led to an increase in MTDs being granted, with particular attention to civil rights claims. These studies are virtually the first attempts ever to study pleading and MTDs empirically. Chiorazzi et al.

⁹In most of the literature, this prediction is treated as so obvious that it goes unstated; but many make it explicit. See Ward 2008, p. 916 (“Post-*Twombly*, a defendant would predict a higher rate of success on [a Rule 12(b)(6)] motion.”); Hannon 2008, p. 1814 (“Generally, any substantive alteration to the pleading standard would have an effect on the dismissal rate under 12(b)(6).”).

(1988) compile a massive annotation of 316 empirical studies in civil procedure, but note (at p. 120) that “there are no studies in pleading per se, perhaps because of the advent of modern notice pleading.”¹⁰ Indeed, for the period before 2008, I know of only one empirical study directly addressing the use of MTDs (Willging 1989).¹¹

Hannon (2008) is the first empirical study of *Twombly*. Hannon looks at a sample of 3,287 district court opinions published on Westlaw that cite either *Twombly* or *Conley* in the context of a MTD. He concludes that, at least among civil rights cases, the grant rate for MTDs rose after *Twombly*. Five subsequent academic studies have followed Hannon’s methodology and reached similar results, although they varied the dates covered and used smaller samples of cases. Hatamyar (2010) and Seiner (2009, 2010) conclude that at least in the context of civil rights claims, *Twombly* is leading to more dismissals. Hatamyar Moore (2012) and Brescia (2012) are less sure of the effect of *Twombly*, but make stronger claims about the effect of *Iqbal*.

These studies on the effects of *Twombly* are quick to acknowledge that inferences drawn from them must be qualified by the limitations of their data and methodology. First, the estimated effects are relatively small and none are statistically significant.¹² Second, any study relying on opinions ruling on MTDs can at best quantify only the share of MTDs granted, not the overall rate at which filed cases are dismissed. Third, these studies rely on opinions published in Westlaw or LexisNexis, when in fact only a fraction of all district court orders are published. See Hoffman, Izenman and Lidicker (2007); Lizotte (2007). It is not known the extent to which the selection of decisions for publication may affect the representativeness of samples drawn from Westlaw or LexisNexis, although Hatamyar Moore (2012) provides evidence that the effect may be small. Fourth, a related concern is that focusing only on cases citing *Conley* or *Twombly*, rather than all cases ruling on MTDs,

¹⁰By way of comparison, they found *five* studies on the use of telephonic conferences in civil litigation.

¹¹A second study, Connolly and Lombard (1980), contains some data related to MTDs.

¹²Hannon (2008) reports one statistically significant result, but this is due to a specification error. A corrected regression on the same data (not reported, on file with author) yields no significant effect.

introduces an additional selection bias.¹³

The most authoritative examination of the data to date is Cecil et al. (2011a), a study by researchers at the Federal Judicial Center (FJC) of thousands of docket records from a sample of federal district courts. Cecil et al. compare cases from 2005–2006 with cases from 2009–2010, bracketing an interval during which both *Twombly* and *Iqbal* were decided. Thus, they could not separately study *Twombly* and *Iqbal*. Using a dataset of MTD filings, they find that MTDs were filed in a larger share of cases in 2009–2010 (6.2 percent) than in 2005–2006 (4.0 percent). Using a second database of rulings on MTDs, they find no change in the rate at which MTDs are granted. Because of differences between the two samples, however, they do not estimate whether the fraction of all filed cases that were dismissed had risen. A follow-up study (Cecil et al. 2011b) confirms these earlier results. The huge data set in Cecil et al. (2011a) and the striking result of no change in the rate at which MTDs are granted has generated considerable interest—and controversy.¹⁴

None of these studies finds a statistically significant increase in the rate at which MTDs are granted after *Twombly*. This is consistent with no change in the legal standards applied by district courts after *Twombly*. It is also consistent, however, with *Twombly* inaugurating a higher standard for MTDs, and the selection of cases responding endogenously in ways that leave unchanged the rate at which MTDs are granted. These studies do not address the question of how to measure the effect of *Twombly* in the presence of selection effects.

Gelbach (2012) is the only study other than mine to explicitly account for selection effects in estimating the effects of *Twombly* or *Iqbal* on dismissal rates. Gelbach develops a unique model of pre-trial litigation, and uses this model to identify the ways in which selection effects can alter the observed effect of a change in the legal standard for MTDs. Combining

¹³District courts ruling on a MTD may cite other Supreme Court cases, precedent from the circuit courts, or even other district court opinions. Given the age of *Conley* during the sample periods in the studies above, district court opinions choosing to cite *Conley* directly, rather than any of the hundreds or thousands of more recent (and equally controlling) precedents, are unlikely to be a representative sample of rulings prior to *Twombly*. For cases citing *Twombly*, this bias may be less of a problem, given its recent vintage.

¹⁴See Cecil 2012 and sources cited therein for criticisms of the study and responses thereto.

this model with the data presented in Cecil et al. (2011a), Gelbach derives bounds on the fractions of dismissed cases that were affected by *Twombly* and *Iqbal*. By design, Gelbach (2012) does not address the DE or AI models empirically and, due to the design of the Cecil et al. (2011a) study, cannot separately study *Twombly* and *Iqbal*.

3 ACCOUNTING FOR SELECTION EFFECTS

Both the praise and the scorn directed at *Twombly* are a consequence of the belief that *Twombly* has, in fact, raised the bar for pleading. But has it? The competing accounts in Part 2 offer contrasting predictions. The conventional view proceeds from the premise that *Twombly* has raised the pleading standard and reaches these conclusions:

Prediction 1. *Twombly* will (significantly) increase the fraction of MTDs that are granted.

Prediction 2. *Twombly* will (significantly) increase the fraction of filed cases that are dismissed.

As discussed above, the conventional account suggests that the effects of *Twombly* will be largest among civil rights cases, where evidence of elements such as discriminatory intent may not be available to plaintiffs absent discovery. Because discovery is generally not available until after a ruling on a MTD, the argument goes, many civil rights plaintiffs will be left without the ability to show that their claim is “plausible” as required by *Twombly*. Thus, for these predictions the conventional account would add the following corollary:

Corollary. The effects above will be (significantly) greater for civil rights cases.

As noted above, the conventional view does not account for selection effects. Below, I explain that selection effects can render uninformative a null result for Prediction 1. Both the DE and AI models predict that the rate at which MTDs are granted will not predictably change, even if the legal standard changes. Prediction 2, however, remains a viable test for the effect of *Twombly*, even in the presence of selection effects under both the DE and AI

selection models, so long as one controls for pre-filing selection using the empirical strategy outlined below. Yet none of the studies cited above test Prediction 2 to assess the effect of *Twombly*.

3.1 Selection Effects in the DE Model

Although Priest and Klein (1984) were skeptical that changes in legal standards could be detected from observed changes in the outcomes of litigated cases, an extension of their model overcomes the selection problem. The key is to recognize that the logic of Priest and Klein should operate not only among filed lawsuits, but among disputes *before* they become filed lawsuits. Priest and Klein (1984, p. 8) assumed that the distribution of case quality among filed lawsuits was arbitrary, but when one applies the logic of the selection model to all disputes (whether filed lawsuits or not), we should expect that among all disputes, the disputes that are “close” are least likely to settle and most likely to become filed lawsuits. In the parlance of the Priest and Klein model, these are the cases with case quality close to Y^* , the threshold case quality for a plaintiff to win at trial. After filing, of course, settlement is still possible, and among filed lawsuits, we would again expect to see the “really close” cases make it to trial, while the discovery process moves the parties toward settling the not-as-close cases.¹⁵

This innovation in the Priest and Klein model gives us some empirical traction, at least when considering the effect of procedural standards such as the threshold for factual sufficiency of a pleading addressed in *Twombly*. The reason is that the standard for a plaintiff to overcome a MTD (call this S^*) is *lower* than the standard for a plaintiff to prevail at trial (Y^*). Raising the standard for pleading (as *Twombly* seems to do) moves S^* *closer* to Y^* . Because of the selection of disputes into filed lawsuits, most lawsuits are close to Y^* . Thus, if one can hold the set of filed lawsuits constant, then raising S^* means moving from a part

¹⁵An earlier version of this paper presented this model of selection into litigation, followed by selection into trial, in a somewhat more formal way. See Hubbard (2012).

of the case quality distribution with fewer cases to a part with more cases.

Further, by the logic of the Priest and Klein model, not all cases will have MTDs—instead, only cases where the MTD is “close” will be litigated. The rest will either settle, be dropped, or will move on to an answer and discovery. Therefore, a rise in S^* should lead to an increase in the total number of MTDs granted, because there are now more cases close to S^* . See Figure 1. By the same token, the rate at which MTDs are granted may be close to 50 percent, regardless of whether S^* is lower or higher. Hence, looking at MTD grant rates may not reveal whether S^* has changed.

In sum, if we can hold constant the set of filed cases, the DE model, as I develop it above, generates the prediction that if *Twombly* increased the pleading standard, MTD grant rates will not (predictably) change, but the fraction of all filed cases that are dismissed will rise. The only challenge that remains is to hold the set of filed cases constant. To do this, my empirical strategy is to look only at cases filed before *Twombly*, comparing those that were decided before *Twombly* with those decided after (while holding case duration fixed).¹⁶ I elaborate on my empirical methodology in Parts 4 and 5 below.

3.2 Selection Effects in the AI Model

AI models of selection into litigation, beginning with Bebchuk (1984), tend to assume that the defendant is the party with private information, and I make that assumption here.¹⁷ This seems especially appropriate given that much of the criticism of *Twombly* has focused on its effect on civil rights and discrimination cases, in which it is plausible that the defendant

¹⁶Note that if actors anticipate a future legal change, their behavior may begin to change in response to the new legal rule *before* the rule is even implemented. This will have the effect of attenuating any measured effect of *Twombly*, by shrinking the difference in outcomes before and after that decision. Fortunately, this appears not to be a concern here. First, as noted above, *Twombly* was a surprise; the standard for pleading was not even raised by the parties to the case. Second, the data reveal no noticeable short-run shifts in filings or dismissals in advance of *Twombly*.

¹⁷If I instead assume that the plaintiff has private information, a similar argument follows, except that the empirical prediction will be an effect on the fraction of filed cases that are dismissed *or settled before resolution of a MTD*. The empirical results under such a specification (not reported) are less precise, but not qualitatively different from those reported herein.

begins the case with considerable private information about its (allegedly discriminatory) motives.

The basic prediction of the AI model is that defendants will settle with plaintiffs who have strong cases and litigate with plaintiffs who have weak cases.¹⁸ Relevant to my analysis is the set of cases in which, before *Twombly*, either (1) the parties settled before the time to resolve a MTD or (2) a MTD was either granted or denied. Assuming that *Twombly* was a change in the legal standard, how does a higher pleading standard affect this set of cases? There are two effects:

First, a higher standard means that some cases that would have survived a MTD will now be dismissed. This will tend to (1) raise the share of MTDs that are granted and (2) raise the total number of cases that are dismissed.

Second, the settlement rate may fall. If *Twombly* reduces the likelihood of success for all plaintiffs equally, there will be no effect on the settlement rate; see Proposition 4 in Bebchuk (1984). But one might expect *Twombly* to expand the range of case quality by reducing the likelihood of success for weak cases (which are close to the margin for dismissal) but leaving strong cases relatively unaffected. If so, the settlement rate will fall; see Proposition 5 in Bebchuk (1984). In this latter case, note that more cases will have MTDs, and the cases that no longer settle will be relatively strong cases, and thus relatively likely to survive a MTD, compared to those previously subject to MTDs. This will tend to (1) reduce the share of MTDs that are granted but (2) raise the total number of cases that are dismissed.

Hence, the effect of the new pleading standard on the rate at which MTDs are granted may be ambiguous, because its two effects push in opposite directions. But the predicted effect on the fraction of all filed cases that are dismissed is unambiguous: it should rise.

In sum, the DE and AI models generate the same predictions, given my empirical method-

¹⁸The AI model assumes a single take-it-or-leave-it settlement offer by the plaintiff. Presumably, this could occur either before or after filing of a lawsuit. I can remain agnostic about timing, as I am holding the set of filed cases fixed in my empirical methodology and focusing only on the possibility of selection after filing but before resolution of a motion to dismiss.

ology. I refer to these shared predictions as the selection/change hypothesis. Table 2 summarizes the predictions of the hypotheses in this paper. Note that the null hypothesis for the both the conventional and selection/change hypotheses is the same: no change. I now turn to the data with which I will test these predictions.

4 DATA

I have compiled two large datasets of federal district court cases. First, I have compiled a data set of over 250,000 cases filed from May 2005–May 2008 using data files collected and published by the Administrative Office of the United States Courts (AO). I will refer to this data as the “AO Data.” My processing of the AO data is described in detail in the Data Appendix. And I note that all of the results reported herein are robust to the changes in the criteria for inclusion in the data set.¹⁹ Summary statistics for the AO Data appear in Table 3. One important fact jumps from this table. In the aggregate, the rates of dismissal before and after *Twombly* appear to be identical. If *Twombly* has changed courts’ willingness to dismiss cases, its effect is being masked by other factors.

Second, Lee Epstein, William M. Landes, and Richard A. Posner have created a dataset of information on all federal district court opinions ruling on a Rule 12(b)(6) motion that were published on Westlaw during the two-year period around *Twombly* (May 21, 2006 to May 21, 2008). After processing this data (which I describe in detail in the Data Appendix), this dataset contains 12,717 unique observations. I will refer to this dataset as the “Opinions Data.” Summary statistics for the Opinions Data appear in Table 1. Again, the summary statistics reveal little change in dismissal rates after *Twombly*.

These two data sets are notable for several reasons. First, they form the largest collection of cases brought to bear on the question of whether *Twombly* has had an effect. Indeed, their huge sample sizes allow me to draw inferences with considerable precision. Second, while the

¹⁹All unreported results mentioned herein are available from the author.

Opinions Data, like earlier studies, can use rulings on MTDs to test Prediction 1, the AO Data can test Prediction 2, which has not yet been tested.²⁰ Third, while the Opinions Data relies on published opinions, and thus may be subject to publication bias, the AO Data is taken from a census of all cases. Fourth, unlike the samples of published opinions in earlier studies, the Opinions Data includes all rulings on MTDs, regardless of whether *Conley* or *Twombly* was cited by the judge.

Finally, although other empirical studies of legal change have used either administrative data or published opinions, as I do here, prior studies in general were not able to configure their data in a way that permitted drawing inferences about changes in a legal standard in the presence of selection effects. To control for selection effects in the composition of filed cases, my empirical strategy focuses on cases that are filed under the old standard but decided under the new standard. This requires sorting cases by filing date, and identifying those cases where were filed before the (apparent) legal change but in which a MTD was decided after the legal change.

Datasets of published opinions, however, inevitably are organized by the date of the opinion, not the date of case filing. (It is not even clear that a set of published opinions, standing alone, *could* be organized by filing date.) The AO data likewise is organized by the (fiscal) year in which each case was terminated. But the AO data, unlike published opinions, does provide filing dates for all cases. Thus, by combining all available annual datasets of terminations with the most recent data on pending cases, I create a database constituting a census of all cases filed on any given date during 1978–2008.

To organize my dataset of published opinions by filing date, I merge it with my administrative data. This requires matching each opinion to a case in the AO data, keeping in mind that a case may have more than one published opinion (or, more likely, none at all), and that, conversely, a single opinion may decide motions pending in several (or dozens or

²⁰Recall that Prediction 1 states that *Twombly* will increase the fraction of MTDs that are granted; and Prediction 2 states that *Twombly* will increase the fraction of filed cases that are dismissed.

hundreds of) related cases. To do this, I note that each case filed in the federal system is assigned a docket number, which is essentially a seven-digit serial number. Each federal district court is subdivided into one or more offices, and within each office, the docket number is a unique identifier for every case ever filed. For every federal civil case, its district, office, and docket numbers together form a unique ID. Equally important, every published district court opinion contains in its case caption the name of its district, the office, and the docket number(s) of the case. By collecting and processing this information, I match filing dates to my opinion data.²¹

Despite their large size and unique configuration, these data sets have their limitations. First, neither the Opinions Data nor the AO Data provides information on how many MTDs were filed. A rise in the number of MTDs filed would be another likely effect of *Twombly*. Cecil et al. (2011a) report a higher rate of filings of MTDs in 2010 than 2006, which suggests that *Twombly* or *Iqbal* may have had this kind of effect.

Second, the *disposition* field in the AO Data does not distinguish between the two main types of motions by which a case can be terminated on the merits before trial: the Rule 12(b)(6) MTD and the Rule 56 motion for summary judgment.²² Both fall into the disposition category “Judgment on motion before trial.”²³

To address this concern, I divide all cases coded as “Judgment on motion before trial” into two groups, based on the duration of the case at the time of termination and the prevailing party. I deem cases with durations of at least 45 and less than 225 days, and for which the prevailing party was the defendant (or prevailing party information was missing), to be terminated on a Rule 12(b)(6) motion. For these cases, I create the dummy variable *dismissal* and set it equal to one. Cases of the same duration, but in which the plaintiff

²¹These and other data processing steps are described in greater detail in the Data Appendix.

²²Nor does the AO Data allow one to distinguish between dismissals based on the factual insufficiency of the complaint (what *Twombly* addressed) and on the legal insufficiency of the complaint (which was untouched by *Twombly*).

²³This category may also capture Rule 12(c) motions, which are essentially equivalent to Rule 12(b)(6) motions, but appear to be much less frequent.

prevailed, may be Rule 12(c) dispositions or other types of judgments (such as defaults). I deem cases with durations of 225 days or more to be terminated on a Rule 56 motion. I chose these boundaries to correspond approximately to the minimum and maximum amounts of time from filing of a complaint to disposition of a MTD in what might be a typical case.²⁴ The Data Appendix details this further. These precise boundaries are, of course, arbitrary, but the results below are not sensitive to adjustments to these bounds.

5 RESULTS

I now present my estimation specifications and regression results for tests of the predictions summarized in Table 2.

5.1 Prediction 1: MTD Grant Rate

The first prediction of the conventional view is that the percentage of MTDs that are granted will rise after *Twombly*. To test Prediction 1, I use the Opinions Data. MTDs decided before May 22, 2007 represent the control group; those decided after May 22, 2007 represent the treatment group. I employ a linear probability model with the following specification:

$$Granted_i = \alpha + \beta \cdot Twombly_i + \gamma \cdot Civil_Rights_i + \delta_i + \eta \cdot Duration_i + \varepsilon_i \quad (1)$$

Granted is an indicator for whether the MTD was granted in whole. *Twombly* is an indicator for the opinion being issued after *Twombly*; $\hat{\beta}$ therefore estimates the effect of *Twombly*. *Civil_Rights* is an indicator for a civil rights case, δ is a vector of circuit fixed effects, and

²⁴As a default, the defendant has 20 days to file a MTD (from the date the complaint is served; see Rule 12(a)(1)), and I assume that a court may take between one and six months to rule. This time range is consistent with findings of a study of docket records in eight district courts, which found that average times to ruling varied widely across districts, but the district averages all fell in the range of 63 to 176 days. See IAALS (2009); Kourlis and Singer (2009). This range also generates predicted rates of dismissal and summary judgment that are consistent with studies on these rates by the FJC. See Willging (1989); Cecil et al. (2007).

Duration is a measure of the duration of the case from filing of the complaint to the date of the opinion. I also run regressions restricted only to civil rights cases, as well as logit regressions (reporting marginal effects). All standard errors are clustered at the district level.

To control for the selection of disputes into filed cases, I include only cases filed on or before May 21, 2007. Thus, all cases in this sample were filed before *Twombly*. Truncating the sample by filing date, however, means that some of the most-quickly-resolved MTDs under *Twombly* are dropped from the sample. This will make the group of cases decided after *Twombly* older, on average, than those decided before. Because the time it takes a court to rule on a motion to dismiss may be correlated with the ruling on the MTD, I include *Duration* as a control to address this source of bias.²⁵

Table 4 present my regression results for all cases. Table 5 presents results for civil rights cases only. Nowhere is the coefficient on *Twombly* statistically significant, and the point estimates are all very small in magnitude relative to the overall rate of granting MTDs, which is around 50 or 60 percent.²⁶ My results are fairly precise. In Table 4, I report the lowest positive effect of *Twombly* that the data reject at the 5 percent level.²⁷ For all cases, the hypothesis that *Twombly* increased the rate at which MTDs are granted by 2.21 percent is rejected in every specification. For civil rights cases, the hypothesis that *Twombly* increased the rate at which MTDs are granted by 3.11 percent is rejected in every specification. These results cast doubt on the conventional view.

²⁵This is the same strategy used by Choi and Pritchard (2011). Regressions employing other methods of addressing potential bias from a truncated sample (not reported) yield very similar results.

²⁶A difference-in-difference specification in which *Civil_Rights* is interacted with *Twombly* to measure the differential effect of *Twombly* on civil rights cases also finds no effect.

²⁷Here I employ a one-tailed test. The question posed is whether the effect of *Twombly* is at least as great as the null supposes.

5.2 Prediction 2: Dismissals as a Fraction of All Filings

The second prediction of the conventional view is that dismissals (cases terminated on a MTD) will rise as a share of all cases after *Twombly*. The selection/change hypothesis also predicts a rise in dismissals. To test Prediction 2, I use the AO Data. I employ a linear probability model with the following specification:

$$Dismissed_i = \alpha + \beta \cdot Twombly_i + \gamma_i + \delta_i + \varepsilon_i \quad (2)$$

Dismissed is an indicator for whether the case was dismissed as defined in Part 4 above. *Twombly* is an indicator for the case being in the later group (i.e., litigation of a MTD occurring after *Twombly*); $\hat{\beta}$ therefore estimates the effect of *Twombly*. There are two sets of fixed effects: γ_i is a vector of nature-of-suit fixed effects, and δ_i is a vector of circuit fixed effects.²⁸ I also run regressions restricted only to civil rights cases, as well as logit regressions (reporting marginal effects). All standard errors are clustered at the district level.

I include cases filed April 6, 2006 to May 21, 2006 and cases filed April 6, 2007 to May 21, 2007. Thus, all cases in this sample were filed before *Twombly*; this sample holds the selection of disputes into filed cases fixed. The former group litigated their MTDs before *Twombly*, while the latter group litigated their cases after *Twombly*. (May 21 is 45 days after April 6; as noted above, a MTD will likely not be filed, briefed, and decided in less than 45 days.) Thus, in the latter group of cases, the litigation of MTDs may have endogenously responded to *Twombly*. Comparing these two groups allows me to measure the effect of *Twombly* on dismissals.

Table 6 presents my results for Prediction 2, including all case types. I present separate results for civil rights cases in Table 7. None of the estimated effects of *Twombly* are statistically significant, and the point estimates are extremely small in magnitude. As Tables

²⁸Equation (2) has a full vector of nature-of-suit fixed effects, while Equation (1) has only a civil rights nature-of-suit dummy. This is because the Opinions Data contains only the single nature-of-suit category. See the Data Appendix for details on the nature-of-suit categories.

6 and 7 indicate, my results are quite precise. In all specifications, I can reject a null hypothesis of a 0.4 percent effect—i.e., the hypothesis that at least 1 in 250 cases was affected by *Twombly*.²⁹ In sum, the empirical test of Prediction 2 rejects all hypotheses based on *Twombly* marking a change in courts’ willingness to dismiss cases, whether accounting for selection or not.

6 CONCLUSION

There is little doubt that the cases adjudicated by courts comprise a selected sample of all disputes; because of this, we might worry that a change in a legal standard as employed by the courts will have no predictable effect on observable outcomes in litigation.

The *Twombly* case presents a perfect example of this problem. While traditional doctrinal analysis of the case has led scholars to divergent views on whether it marks a significant change in the law governing pleading, both sides of the debate agree that its effect on the law should be inferable from its effect on practice in the district courts. A number of studies find no significant change in MTD grant rates following *Twombly*, but this null result may be an artifact of selection effects.

In this paper, I show that it is possible both to account for selection effects in litigation and to empirically measure the effects of *Twombly* on dismissals. I find fairly precise zeros for the effects of *Twombly* on both the grant rate of MTDs and the overall rate of dismissals among filed cases. These results support the view that *Twombly* effected no (significant) change in courts’ willingness to dismiss cases, even after accounting for selection effects under both the DE and AI models.³⁰

²⁹One potential concern with this estimation approach is that it is possible that in the wake of *Twombly*, courts began to take longer to decide MTDs. By fixing the time window for a dismissal (at 45–224 days), I may have biased estimates of the effect of *Twombly* downwards. If this is happening, we should expect to see post-*Twombly* dismissal rates rise relative to pre-*Twombly* rates if we lengthen the time window for dismissals—but when I do so, I find no evidence of this pattern.

³⁰Willing and Lee (2010) find qualitative evidence consistent with this conclusion. They survey both plaintiffs’ and defense attorneys across a range of practice areas and report that “[m]ost interviewees indicated

To this result I must attach three qualifications. First, it is possible that the effects of *Twombly* will show up gradually over time. Rather than directly changing the behavior of district courts, *Twombly* may have touched off a period of evolution in the treatment of pleadings by district courts, which eventually led to plaintiffs facing a significantly higher bar to pleadings. The empirical methodology in this paper can only detect the immediate effects of *Twombly*.³¹ Nonetheless, I can say that the effect of *Twombly*, if it was significant, was not immediate. This suggests something about the Supreme Court’s control over the “lower” courts’ treatment of pleadings: at best, the lower courts’ response (if any) to *Twombly* was not so much a quick shift to a higher standard, as it was a period in which the lower courts negotiated a new standard for pleadings among themselves. Indeed, the *Iqbal* decision might be seen as part of this process as well.

Second, it is quite possible that while the effect of *Twombly* on broad categories of civil litigation is negligible, *Twombly* may have had an impact in a small subset of cases—indiscernable to the empirical analysis used in this paper but perhaps important nonetheless. The opinion in *Twombly* expressed grave concerns about the cost of litigating low-merit, high-stakes class actions. Perhaps if *Twombly* tipped the balance in favor of granting a MTD in only a handful of cases that, by some criteria, fell into this category, this was exactly the effect the Supreme Court intended. Of course, it is also possible that *Twombly* has had dramatic effects on very different types of cases; in either event, such an effect, if it exists, would have to be concentrated among a very small number of cases.

Third, and relatedly, throughout this paper I have sometimes referred to changes in the “legal standard,” and sometimes referred to changes in the “outcomes of cases,” treating these two concepts as interchangeable. And for purposes of my empirical analysis, they are interchangeable, as they are observationally equivalent in the data. But one can distinguish in principle between (1) *Twombly* having no effect on the pleading standard applied by

that they had not seen any impact of [*Twombly* and *Iqbal*] in their practice.”

³¹Further, the results in this paper do not address the impact, if any, of the *Iqbal* decision.

district courts and (2) *Twombly* changing the pleading standard, but this change in the pleading standard having no (discernible) effect on the rate at which district courts dismiss cases.

My sense is that both of these phenomena have occurred with respect to *Twombly*. Scholars have documented that long before *Twombly*, the lower courts were requiring factual pleadings that went beyond the threadbare allegations seemingly anticipated by *Conley*. Marcus 1986; Fairman 2003. Thus, it is possible that *Twombly* was little more than validation of what most courts were already doing. And there is a strong case to be made that even before *Twombly*, the vast majority of complaints already contained factual detail well above any heightened “plausibility” standard introduced by *Twombly*; if so, even if the legal standard applied by district courts did become stricter, the change would have virtually no effect on the outcomes of MTDs across all cases. The potential reasons for this are manifold. Future work will explore additional factors, but I note three factors here.

One. Plaintiffs draft detailed complaints for strategic reasons. A complaint is the plaintiff’s first opportunity to present a persuasive narrative to the court, and thus the plaintiff’s first opportunity to sway the opinion of the judge. See *Am. Nurses Ass’n v. Illinois*, 783 F.2d 716, 723–24 (7th Cir. [1986]); Willging and Lee (2010). And barring a successful motion to dismiss, a defendant is required to admit or deny every allegation of the complaint. Rule 8(b). A factually rich complaint may yield damaging admissions or elicit denials that undermine the defendant’s credibility.

Two. Many plaintiffs, especially individual plaintiffs, are represented by attorneys who are working on a contingency basis. Attorneys working on contingency have a strong incentive to screen cases for merit at the outset of their representation, and cases lacking a plausible factual basis for the legal claim are unlikely to attract the attention of plaintiff’s attorneys, no matter the pleading standard.

Three. Rule 11 requires that “factual contentions have evidentiary support,” lest the

pleading party or its attorneys face sanctions. By pleading detailed facts, a plaintiff could pre-empt any threat of a motion for Rule 11 sanctions. Pleading parties thus have an incentive to plead facts in support of legal claims even though, prior to *Twombly*, there was no requirement to do so. See Picker (2007).

With these caveats, this paper strongly rejects the view that *Twombly* constitutes a major change in how district courts have applied the law of pleading.

DATA APPENDIX

AO Data Processing

The AO collects information on every civil case terminated in the federal court system, including the following: the circuit, district, and office in which it was filed; the docket number under which it was filed; the date of filing and the date (if any) of termination; the “nature of suit,” which is the category of legal claim (e.g., personal injury, breach of contract) identified by the plaintiff when filing the suit; and the disposition of the suit (e.g., judgment on a jury verdict). This administrative data are compiled and published as database files by the FJC and made available through the Inter-University Consortium for Political and Social Research (ICPSR)³². See FJC 2005–2009.³³

The first year for which complete data is available is fiscal year 1979 (beginning July 1, 1978); the most recent year available is fiscal year 2009 (ending September 30, 2009). The AO has also released data on all civil cases pending as of December 31, 2008. Together, these datasets constitute a census of all federal civil cases filed or terminated from July 1, 1978

³²See <http://www.icpsr.umich.edu/icpsrweb/ICPSR/>.

³³The AO uses this data to generate aggregate statistics about total case filings, terminations, and other information, which it publishes annually. See <http://www.uscourts.gov/Statistics.aspx>. This data have been used in a large number of papers over the years, although most studies have relied on the published aggregate statistics rather than the underlying microdata. Eisenberg and Schlanger (2003, nn.2–17) and Hadfield (2005, n. 5) collect dozens such citations.

to December 31, 2008 (and all cases terminated through September 30, 2009), amounting to about 8 million observations.

To process this data for analysis, I ensure all codes for district, office number, and docket number are numeric values, replacing alphanumeric values with unused, unique, numerical values when necessary. These codes are used to generate identifiers for each unique federal civil case. I generate dummy variables identifying various categories of suits. See Table A1. As reported in Table A1, a number of categories are excluded from the reported specifications. These include categories outside of (what might be called) the “typical” forms of civil litigation described in the main text. For example, I exclude habeas corpus petitions, which are essentially criminal matters, and social security cases, which are essentially appeals from administrative proceedings. See Hubbard (2012) for more details.

Further, reported results exclude cases with a *pro se* or *in forma pauperis* party. Shortly after *Twombly*, the Supreme Court issued a per curiam opinion that appeared to exclude *pro se* cases from the application of *Twombly*. See *Erickson v. Pardus* (551 U.S. 89 [2007]) (per curiam). Reported results also exclude categories of cases involving unusual spikes in filings, such as cases in the Eastern District of Louisiana (due to Hurricane Katrina) and nature of suit categories involving “mass filings” (i.e., more than 1000 cases filed in a single day in a single district). I exclude cases in multi-district litigation (MDL) proceedings; these cases are often terminated (including dismissed) in batches of hundreds or thousands. I exclude fraud (including securities fraud) cases, as these cases involved explicitly heightened pleading standards even before *Twombly*. See Rule 9(b); 15 USC §78u-4(b). Finally, I exclude cases involving the review of arbitral awards (*disposition* = 15, 16 or *progress* = 13), appeals to a district court (*disposition* = 19, 20), and bankruptcy stays (variously coded, depending on year), all of which are procedurally unique.

For computing dismissal rates, I exclude all observations in which the disposition is a transfer to a new district (including MDL transfers). I exclude cases for which no opportunity

arose for a MTD; given my definition of *dismissal*, these are cases with *duration* ≤ 45 . For the same reason, I drop cases that were disposed of on grounds that are procedurally antecedent to a Rule 12(b)(6) motion, such as dismissals for want of jurisdiction, dismissals for want of prosecution, default judgments, remands of removed cases to state court, and remands to U.S. agencies.

I also account for double counting of cases in the raw data. If a single case is closed and reopened (this can happen multiple times) or transferred to another district (including for an MDL proceeding), that case will appear as multiple observations in the AO data, even though there has been only a single opportunity for the case to be terminated because of a Rule 12(b)(6) MTD. To address double counting, I de-duplicate case records in the data. See Hubbard (2012) for additional details on recoding of cases and de-duplication.

Notably, in more than 99 percent of cases with multiple observations, time-invariant variables, such as jurisdictional basis and nature of suit code, had no discrepancies across observations for the same case. This suggests a high degree of reliability in the data. Also, although some studies have found high error rates in the coding of some fields in the AO Data (see Eisenberg and Schlanger 2003; Hadfield 2004), the accuracy of the fields describing case characteristics essential to this paper (such as the jurisdiction, dates of filing and disposition, etc.) has never been questioned, and Hadfield (2005) finds the key outcome field, *disposition*, to be fairly reliable.³⁴

Opinions Data Processing

The Opinions Data is derived from a data set collected by Lee Epstein, William M. Landes, and Richard A. Posner. Each observation in this data set represents a published judicial opinion ruling on a MTD. For each observation, the data set provides the Westlaw citation

³⁴Based on audits using docket records of a sample of cases, Hadfield (2005) finds that the “Judgment on motion before trial” code corresponds to final judgments on motions before trial (presumably either Rule 12(b)(6) or Rule 56 motions) in 85–90 percent of cases and the “Judgement: Settled” code corresponds to settlements in approximately 95 percent of cases.

of the opinion, the month and year of the opinion, the district in which the case was decided, an indicator for whether the case was a civil rights case (based on subjective coding criteria), an indicator for whether the opinion granted the MTD in its entirety, as well as a number of other fields not used in my analysis.

To prepare this data for use, I remove duplicate observations (i.e., observations referring to the same opinion) from the data set and then divide the remaining observations among myself and my research assistants by random assignment. We then retrieve from Westlaw each opinion and record the exact date of the opinion, as well the district, office, and docket number. The date and district information is cross-checked against the corresponding fields in the existing data set and discrepancies are resolved. I drop all additional duplicate observations identified during this process. When two duplicate observations contain different codes for whether the MTD was granted in full or for whether the case was a civil rights case, I drop both. This happened in less than 1 percent of cases.

This yields a data set of 12,717 unique opinions. Summary statistics for this data set appear in the main text. This configuration of the data set makes it most comparable to previous studies on the effect of *Twombly* on the rate at which MTDs are granted.

In order to check the reliability of the coding in this dataset and to generate results more comparable to results based on the AO Data, I reorganize the data set by unique case identifier (given by district, office, and docket number) rather than Westlaw citation. Some cases had more than one opinion on Westlaw; for these cases, information about the multiple opinions was collapsed into a single observation. Conversely, some opinions dealt with more than one case (this happened in the case of consolidated and MDL cases). For these opinions, multiple observations were created, each with the same opinion-specific information, and one observation assigned to each unique case.

Once in this configuration, this data set could be merged with the AO Data by matching on unique case identifier. Of the 13,031 observations generated by this process, more than

98 percent (all but 242) were successfully matched with observations in the AO Data. By matching the Opinions Data to the AO Data, I can check for consistency across those codes that are common to both datasets; despite the fact that the Opinions Data was coded by hand, approximately 98 percent of all observations in the matched Opinions Data have identical values for those fields that also appear in the AO Data. This indicates a high level of reliability, particularly for a hand-collected and hand-coded data set.

I then drop all observations excluded from all specifications using the AO Data (see Table A1) and all observations ($n = 83$) in which there were opinions published both before and after *Twombly*. Matching the Opinions Data with the AO Data also allows me to compare the date on which a MTD is granted in the Opinions Data to the date on which a case is terminated in the AO Data. Using this, I define dismissal as termination of a case within 90 days of the granting of a MTD.

With this configuration of the Opinions Data, I isolate those cases filed before *Twombly* and compute the duration from case filing to the date of the opinion. (This requires dropping 132 cases with more than one opinion and 6 cases in which the computed duration was negative.) This subset of the data contains 10,150 observations and regression results using this data are reported in Tables 4 and 5 in the main text.

REFERENCES

- Bebchuk, Lucian Arye. 1984. Litigation and Settlement under Imperfect Information. *RAND Journal of Economics* 15(3):404–415.
- Bone, Robert G. 2009. *Twombly*, Pleading Rules, and the Regulation of Court Access. *Iowa Law Review* 94:873–936.
- Brescia, Raymond H. 2012. The *Iqbal* Effect: The Impact of New Pleading Standards in Employment and Housing Discrimination Litigation. *Kentucky Law Journal* 100:235–287.
- Cecil, Joe S. 2012. Of Waves and Water: A Response to Comments on the FJC Study *Motions to Dismiss for Failure to State a Claim After Iqbal*. Working paper. Available at <http://dx.doi.org/10.2139/ssrn.2026103>.

Cecil, Joe S., George W. Cort, Margaret S. Williams, and Jared J. Bataillon. 2011a. *Motions to Dismiss for Failure to State a Claim After Iqbal*. Washington, D.C.: FJC.

—. 2011b. *Update on Resolution of Rule 12(b)(6) Motions Granted with Leave to Amend*. Washington, D.C.: FJC.

Cecil, Joe S., Rebecca N. Eyre, Dean Miletich, and David Rindskopf. 2007. A Quarter-Century of Summary Judgment Practice in Six Federal District Courts. *Journal of Empirical Legal Studies* 4:861–907.

Choi, Stephen J., and A.C. Pritchard. 2011. The Supreme Court’s Impact on Securities Class Actions: An Empirical Assessment of *Tellabs*. *Journal of Law, Economics, & Organization* Advance Access published Aug. 25, 2011. doi:10.1093/jleo/ewr014.

Chiorazzi, Michael, and Barbara A. Baccari, Karen R. Cashion, Christopher R. Hart, Donald M. Nielsen, Charles M. North, and William T. O’Neil. Empirical Studies in Civil Procedure: A Selected Annotated Bibliography. *Law & Contemporary Problems* 51(3):87–207.

Connolly, Paul R. J., and Patricia A. Lombard. 1980. *Judicial Controls and the Civil Litigation Process: Motions*. Washington, D.C.: FJC.

Eisenberg, Theodore. 1991. The Relationship between Plaintiff Success Rates before Trial and at Trial. *Journal of the Royal Statistical Society, Series A* 154(1):111–116.

Eisenberg, Theodore, and Margo Schlanger. 2003. The Reliability of the Administrative Office of the U.S. Courts Database: An Initial Empirical Analysis. *Notre Dame Law Review* 78:1455–1496.

Epstein, Richard A. 2009. *Twombly*, After Two Years: The Procedural Revolution in Antitrust That Wasn’t. *GCP: The Online Magazine for Global Competition Policy* July 2009(2).

Fairman, Christopher M. 2003. The Myth of Notice Pleading. *Arizona Law Review* 45:987–1065.

Federal Judicial Center. 2005a. Federal Court Cases: Integrated Database, 1970–2000 (Computer file ICPSR08429-v7.) Ann Arbor, MI: Inter-university Consortium for Political and Social Research.

—. 2005b. Federal Court Cases: Integrated Database, 2001 (Computer file ICPSR03415-v3.) Ann Arbor, MI: Inter-university Consortium for Political and Social Research.

—. 2005c. Federal Court Cases: Integrated Database, 2002 (Computer file ICPSR04059-v3.) Ann Arbor, MI: Inter-university Consortium for Political and Social Research.

—. 2005d. Federal Court Cases: Integrated Database, 2003 (Computer file ICPSR04026-v5.) Ann Arbor, MI: Inter-university Consortium for Political and Social Research.

—. 2006a. Federal Court Cases: Integrated Database, 2004 (Computer file ICPSR04348-

- v2.) Ann Arbor, MI: Inter-university Consortium for Political and Social Research.
- . 2006b. Federal Court Cases: Integrated Database, 2005 (Computer file ICPSR04382-v1.) Ann Arbor, MI: Inter-university Consortium for Political and Social Research.
- . 2007. Federal Court Cases: Integrated Database, 2006 (Computer file ICPSR04685-v2.) Ann Arbor, MI: Inter-university Consortium for Political and Social Research.
- . 2009a. Federal Court Cases: Integrated Database, 2007 (Computer file ICPSR22300-v2.) Ann Arbor, MI: Inter-university Consortium for Political and Social Research.
- . 2009b. Federal Court Cases: Integrated Database, 2008 (Computer file ICPSR25002-v2.) Ann Arbor, MI: Inter-university Consortium for Political and Social Research.
- Gelbach, Jonah B. 2012. Note. Locking the Doors to Discovery? Assessing the Effects of *Twombly* and *Iqbal* on Access to Discovery. *Yale Law Journal* 121:2270–2345.
- Hadfield, Gillian K. 2004. Where have all the trials gone? Settlements, non-trial adjudications and statistical artifacts in the changing disposition of federal civil cases. *Journal of Empirical Legal Studies* 1:705–734.
- . 2005. Exploring Economic and Democratic Theories of Civil Litigation: Differences between Individual and Organizational Litigants in the Disposition of Federal Civil Cases. *Stanford Law Review* 57:1275–1327.
- Hannon, Kendall W. 2008. Much Ado About *Twombly*? A Study of the Impact of *Bell Atlantic Corp. v. Twombly* on 12(b)(6) Motions. *Notre Dame Law Review* 83:1811–1846.
- Hatamyar, Patricia W. 2010. The Tao of Pleading: Do *Twombly* and *Iqbal* Matter Empirically? *American University Law Review* 59:553–633.
- Hatamyar Moore, Patricia W. 2012. An Updated Quantitative Study of *Iqbal*'s Impact on 12(b)(6) Motions. *University of Richmond Law Review* 46:603–657.
- Hoffman, David A., Alan J. Izenman and Jeffrey R. Lidicker. 2007. Docketology, District Courts, and Doctrine. *Washington University Law Review* 85:681–751.
- Holmes, Oliver W. 1897. The Path of the Law. *Harvard Law Review* 10(8):457–78.
- Hubbard, William H.J. 2012. The Problem of Measuring Legal Change, with Application to *Bell Atlantic v. Twombly*. U of Chicago Law & Economics, Olin Working Paper No. 575.
- Institute for the Advancement of the American Legal System (IAALS). 2009. *Civil Case Processing in the Federal District Courts*. Denver, CO: IAALS.
- Kourlis, Rebecca Love and Jordan M. Singer. 2009. Managing Towards the Goals of Rule 1. *Federal Courts Law Review* 2009:1–18.
- Lizotte, Brian N. 2007. Publish or Perish: The Electronic Availability of Summary Judg-

ments by Eight District Courts. *Wisconsin Law Review* 2007:107–149.

Marcus, Richard L. 1986. The Revival of Fact Pleading Under the Federal Rules of Civil Procedure. *Columbia Law Review* 86(3):433–494.

Miller, Arthur R. 2010. From *Conley* to *Twombly* to *Iqbal*: A Double Play on the Federal Rules of Civil Procedure. *Duke Law Journal* 60(1):1–130.

Picker, Randal C. 2007. *Twombly*, *Leegin*, and the Reshaping of Antitrust. *Supreme Court Review* 2007:161–202.

Priest, George L. 1985. Reexamining the Selection Hypothesis: Learning from Wittman's Mistakes. *Journal of Legal Studies* 19:215–243.

Priest, George L., and Benjamin Klein. 1984. The Selection of Disputes for Litigation. *Journal of Legal Studies* 13:1–55.

Redish, Martin H. and Lee Epstein. 2008. *Bell Atlantic v. Twombly* and the Future of Pleading in the Federal Courts: A Normative and Empirical Analysis. Northwestern Univ. School of Law Public Law and Legal Theory Working Paper No. 10-16.

Reinert, Alexander A. 2011. The Costs of Heightened Pleading. *Indiana Law Review* 86:119–184.

Schneider, Elizabeth M. 2010. The Changing Shape of Federal Civil Pretrial Practice: The Disparate Impact on Civil Rights and Employment Discrimination Cases. *University of Pennsylvania Law Review* 158:517–570.

Seiner, Joseph A. 2010. Pleading Disability. *Boston College Law Review* 51:95–149.

—. 2009. The Trouble with *Twombly*: A Proposed Pleading Standard for Employment Discrimination Cases. *University of Illinois Law Review* 2009:1011–1060.

Shavell, Steven. 1996. Any Frequency of Plaintiff Victory at Trial Is Possible. *Journal of Legal Studies* 25:493–501.

Siegelman, Peter, and John J. Donahue III. 1995. The Selection of Employment Discrimination Disputes for Litigation: Using Business Cycle Effects to Test the Priest-Klein Hypothesis. *Journal of Legal Studies* 24:427–462.

Siegelman, Peter, and Joel Waldfogel. 1999. Toward a Taxonomy of Disputes: New Evidence Through the Prism of the Priest/Klein Model. *Journal of Legal Studies* 28:101–130.

Smith, Douglas G. 2009. The *Twombly* Revolution? *Pepperdine Law Review* 36:1063–1100.

Stancil, Paul. 2009. Balancing the Pleading Equation. *Baylor Law Review* 61:90–173.

Steinman, Adam N. 2010. The Pleading Problem. *Stanford Law Review* 62:1293–1360.

- Waldfogel, Joel. 1995. The Selection Hypothesis and the Relationship between Trial and Plaintiff Victory. *Journal of Political Economy* 103(2):229–260.
- Ward, Ettie. 2008. The After-Shocks of *Twombly*: Will We “Notice” Pleading Changes? *St. John’s Law Review* 82:893–918.
- Willging, Thomas E. 1989. *Use of Rule 12(b)(6) in Two Federal District Courts*. Washington, D.C.: FJC.
- Willging, Thomas E. and Emery G. Lee, III. 2010. *In Their Words: Attorney Views About Costs and Procedures in Federal Civil Litigation*. Washington, DC: FJC.
- Wittman, Donald. 1985. Is the Selection of Cases for Trial Biased? *Journal of Legal Studies* 14:185–214.

Tables

Table 1: Summary Statistics for Opinions Data

	(1) Pre- <i>Twombly</i>	(2) Post- <i>Twombly</i>
Civil Rights		
Opinions	2,230	2,379
Dismissal Rate	0.569	0.583
Non-Civil Rights		
Opinions	3,851	4,257
Dismissal Rate	0.466	0.476
Total		
Opinions	6,081	6,636
Dismissal Rate	0.504	0.515

Table 2: Summary of Predictions from Competing Hypotheses

	Conventional	No Change	Selection/Change
1. MTD Grant Rate	+	0	?
2. Dismissals as Fraction of Filed Cases	+	0	+

Note: “+” indicates significantly positive change in response to the treatment. “0” indicates little or no change. “?” indicates no clear prediction.

Table 3: Summary Statistics for Administrative Data

	Year Beginning May 22		
	2005	2006	2007
All Cases			
Filings	85,606	83,770	82,682
Dismissal Rate	0.014	0.015	0.013
Civil Rights			
Filings	22,762	21,485	21,193
Dismissal Rate	0.016	0.016	0.013
Shares			
Tort	0.068	0.066	0.066
Contract	0.260	0.269	0.269
Civil Rights	0.266	0.256	0.256

Table 4: Opinions Data Regression Results, All Cases

Model	(1) Linear	(2) Linear	(3) Linear	(4) Logit
<i>Twombly</i>	− 0.0031 (0.0129)	− 0.0040 (0.0128)	− 0.0038 (0.0127)	− 0.0037 (0.0130)
Can reject at $p=0.05$ effect of	0.0221	0.0211	0.0210	0.0217
<i>Duration</i>	− 0.0194* (0.0090)	− 0.0270** (0.0095)	− 0.0273** (0.0091)	− 0.0285** (0.0097)
<i>Civil_Rights</i>			0.1024** (0.0197)	0.1049** (0.0200)
Circuit Dummies		Y	Y	Y
Constant	0.5280** (0.0235)	0.7342** (0.0148)	0.6864** (0.0149)	
<i>N</i>	10,150	10,150	10,150	10,150

Note: All errors are clustered at the district level. Logit coefficients are reported as marginal effects, evaluated at sample mean. Reported values for the constant in regressions with circuit fixed effects reflect the mean in the omitted circuit (DC).

* Significant at 5% level.

** Significant at 1% level.

Table 5: Opinions Data Regression Results, Civil Rights Cases

Model	(1) Linear	(2) Linear	(3) Logit
<i>Twombly</i>	− 0.0082 (0.0201)	− 0.0075 (0.0193)	− 0.0076 (0.0197)
Can reject at $p=0.05$ effect of	0.0311	0.0303	0.0311
<i>Duration</i>	− 0.0408** (0.0149)	− 0.0437** (0.0150)	− 0.0449** (0.0158)
Circuit Dummies		Y	Y
Constant	0.6146** (0.0346)	0.7649** (0.0257)	
N	3,614	3,614	3,614

Note: All errors are clustered at the district level. Logit coefficients are reported as marginal effects, evaluated at sample mean. Reported values for the constant in regressions with circuit fixed effects reflect the mean in the omitted circuit (DC).

* Significant at 5% level.

** Significant at 1% level.

Table 6: AO Data Regression Results, All Cases

Model	(1) Linear	(2) Linear	(3) Linear	(4) Logistic
<i>Twombly</i>	− 0.0023 (0.0022)	− 0.0024 (0.0022)	− 0.0024 (0.0022)	− 0.0021 (0.0020)
Can Reject at $p = 0.05$ Effect of	0.0013	0.0013	0.0013	0.0011
NOS Dummies			Y	Y
Circuit Dummies		Y	Y	Y
Constant	0.0187** (0.0020)	0.0187** (0.0014)	0.0208** (0.0029)	
N	17,223	17,223	17,223	17,223

Note: All errors are clustered at the district level. Logit coefficients are reported as marginal effects, evaluated at sample mean. Reported values for the constant in regressions with circuit fixed effects reflect the mean in the omitted circuit (DC). “NOS” stands for “Nature of Suit.”

* Significant at 5% level.

** Significant at 1% level.

Table 7: AO Data Regression Results, Civil Rights Cases

Model	(1) Linear	(2) Linear	(3) Logit
<i>Twombly</i>	− 0.0027 (0.0041)	− 0.0027 (0.0040)	− 0.0021 (0.0030)
Can Reject at $p = 0.05$ Effect of	0.0040	0.0039	0.0029
Circuit Dummies		Y	Y
Constant	0.0192** (0.0034)	0.0192** (0.0029)	
N	4,853	4,853	4,853

Note: All errors are clustered at the district level. Logit coefficients are reported as marginal effects, evaluated at sample mean. Reported values for the constant in regressions with circuit fixed effects reflect the mean in the omitted circuit (DC).

* Significant at 5% level.

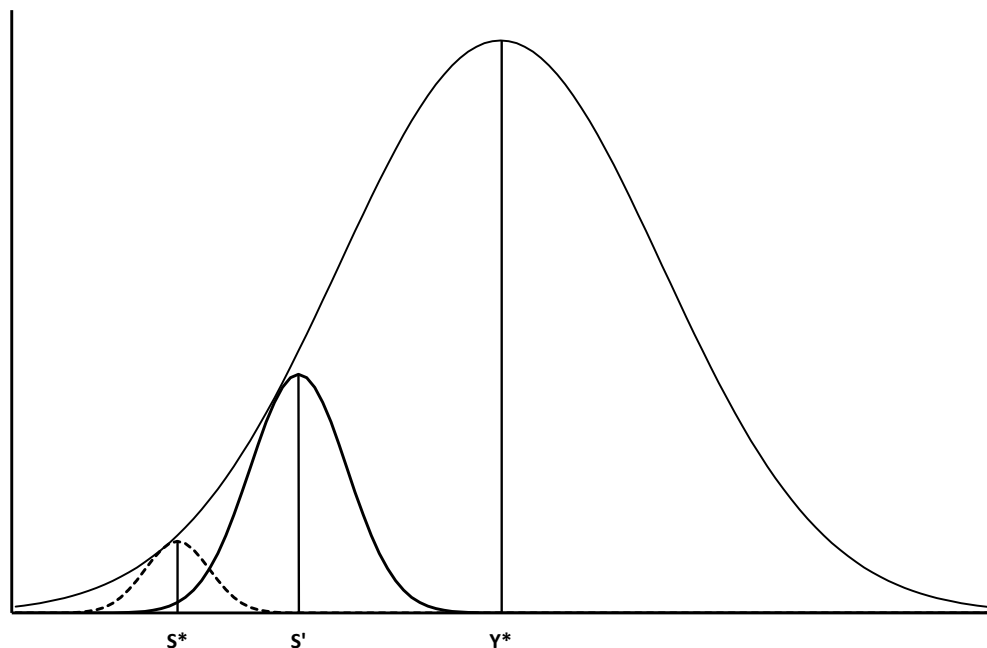
** Significant at 1% level.

Table A1: Nature of Suit Recodes

Recode Category	AO Nature of Suit Codes
Contract	100–196 excl. 150–153, 191
Real Property	220; 230; 290
Torts (excl. Fraud)	310–362 excl. 330, 340, and 360
Fraud (incl. Securities)	371; 850
Antitrust	410
Civil Rights	440–446
Other	191; 210; 330; 720–791 excl. 750; 820–840; 875–890; 895; 950; 990
Mass Filing Categories	340; 360; 365–370; 380; 710
Other Categories Excluded in Reported Specifications	240; 245; 385; 422–430; 450–690 excl. 520; 860–871 excl. 862; 891–894; 970
Excluded in All Specifications	150–153; 400; 520; 750; 810; 862; 900–940, 992

Figures

Figure 1: Motions to Dismiss and Dismissals
under DE Model, Holding Filed Lawsuits Fixed



Note: The thin solid curve represents the distribution of case quality Y among filed cases. Y^* is the threshold for plaintiff success at trial. The thick dashed curve represents the distribution of filed cases with a MTD, given pleading standard S^* . S' indicates a new, higher pleading standard. The thick solid curve represents the distribution of filed cases with MTDs, given S' . Cases below the relevant pleading standard are dismissed if a MTD is litigated. (Heights of thick curves are exaggerated for clarity. Even near S' , only a fraction of filed cases will have litigated motions to dismiss.)