## APPENDIX B: ESTIMATED STANDARD ERRORS FOR THE ESTIMATED LOWER BOUNDS

Rule 8(a) of the Federal Rules of Civil Procedure provides in pertinent part that "[a] pleading that states a claim for relief must contain: . . . (2) a short and plain statement of the claim showing that the pleader is entitled to relief."[1] This rule plays a critical role in the adjudication of Rule 12(b)(6) motions to dismiss for failure to state a claim: failure to make a "short and plain statement of the claim showing that the pleader is entitled to relief" is the basis for dismissal under Rule 12(b)(6).

The null hypothesis of interest in this Note is that *Twombly/Iqbal* changed nothing about either judicial or party behavior. Under this hypothesis, both the negatively affected share and my lower bound on it will be zero, by construction, though my estimates generally would not be exactly zero due to sampling variation. My objective in this appendix is to (a) determine the sampling distribution of my estimated lower bounds when the null hypothesis holds, and (b) use this distribution to carry out classical hypothesis testing to assess whether the observed lower bound estimates are sufficiently far from zero to reject the null hypothesis. The theoretical discussion in Part 0 of this appendix is unavoidably both technical and somewhat lengthy. To keep it from being even longer, I have not tried to make this discussion accessible to those without advanced training in modern large-sample statistical theory; throughout, I assume readers' familiarity with asymptotic statistics, vector calculus, and basic matrix algebra. Readers who are uncomfortable with these mathematical areas should skip directly to Part (3), where I present the *z*-statistics necessary for conducting statistical inference.

### *A. Deriving the Asymptotic Variance of NASLB*

Determining the sampling distribution of the statistic NASLB involves some non-standard aspects because the data I use come from two separate data-collection procedures: one for the rate at which defendants file Rule 12(b)(6) MTDs in cases that plaintiffs have filed,

1. FED. R. CIV. P. 8(a).

and one for the rate at which defendants ultimately prevail on these motions. Because I do not have access to the underlying data, I cannot account for whatever dependence exists across these two collection procedures. I will simply assume that each MTD filing rate is independent of each defendant-prevails rate.

To determine the null distribution of my estimated lower bound, I suppose that the data generating process involves two sets of independently drawn cases. The first set involves M cases drawn independently from a population distribution whose support consists of four mutually exclusive categories:

1. a case is filed under the *Conley* regime and has a Rule 12(b)(6) MTD filed;
2. a case is filed under the *Conley* regime and does not have a Rule 12(b)(6) MTD filed;
3. a case is filed under the *Twombly/Iqbal* regime and has a Rule 12(b)(6) MTD filed;
4. a case is filed under the *Twombly/Iqbal* regime and does not have a Rule 12(b)(6) MTD filed.

Given a randomly drawn case, define the random variable x to take on a value in the set {1,2,3,4}, with the particular value depending on which of the four events just described occurs.[2] Define the probability that $x=1$ to be $m_c$ and the probability that $x=3$ to be $m_t$; let the probabilities that $x=2$ and $x=4$ be $p_c$ and $p_t$, respectively. Let M be the number of cases drawn from this process, and let $X=(X_1, X_2, X_3, X_4)$ be the vector giving the total number of cases falling in each of the four categories (so that $X_1+X_2+X_3+X_4=M$). Assuming the cases are draw independently, then X has a multinomial distribution with parameters $(M, m_c, p_c, m_t, p_t)$. Under our null hypothesis, we have $m_c=m_t=m$ and $p_c=p_t=p$. Since the four probabilities must sum to one, we then have $2(m+p)=1$, which implies that $(m+p)=½$. Since $(m+p)$ is the probability that a case would fall in either pleading regime under the null

2. Note that this approach allows the pleading regime itself to be randomly determined. An alternative approach would be to condition on the observed number of cases filed under each pleading regime, whether unadjusted or adjusted. The advantage of letting the numbers of cases filed in each regime be random is that it allows for the possibility that random chance, rather than party selection, explains why the number of MTDs actually filed under *Twombly/Iqbal* in the FJC's filing dataset is greater than the number actually filed under *Conley*.

hypothesis, it follows that any multinomial distribution that satisfies the null hypothesis has equal expected numbers of cases drawn from each pleading regime. In addition, since the null hypothesis implies that $m_c=m_t$, the same share of cases will have MTDs filed under each pleading regime when the null holds. Thus, the null is consistent with the absence of all types of party selection (i.e., the absence not only of plaintiff and settlement selection, but also of defendant selection).

The second part of the data generating process concerns cases in the FJC's grants data set. I assume that the N cases in this data set are generated independently from a population distribution whose support consists of the following four mutually exclusive categories:

1. a case with a Rule 12(b)(6) MTD adjudicated in one of the FJC's study periods is adjudicated under the *Conley* study period and has its defendant ultimately prevail;
2. a case with a Rule 12(b)(6) MTD adjudicated in one of the FJC's study periods is adjudicated under the *Conley* study period and has its plaintiff ultimately prevail;
3. a case with a Rule 12(b)(6) MTD adjudicated in one of the FJC's study periods is adjudicated under the *Twombly/Iqbal* study period and has its defendant ultimately prevail;
4. a case with a Rule 12(b)(6) MTD adjudicated in one of the FJC's study periods is adjudicated under the *Twombly/Iqbal* study period and has its plaintiff ultimately prevail.

Given a randomly drawn case, define the random variable y to take on a value in the set {1,2,3,4} depending on which of the four events just described occurs. Define the probability that y=1 to be $h_c$ and the probability that y=3 to be $h_t$; let the probabilities that y=2 and y=4 be $q_c$ and $q_t$, respectively. Let $Y=(Y_1, Y_2, Y_3, Y_4)$ be the vector giving the total number of cases falling in each of the four categories. Assuming the cases are drawn independently, then Y has a multinomial distribution with parameters ($N$, $h_c$, $q_c$, $h_t$, $q_t$). Under our null hypothesis, we have $h_c=h_t=h$ and $q_c=q_t=q$. Since the four probabilities must sum to one, we have $2(h+q)=1$, which implies that $(h+q)=½$. Since $(h+q)$ is the probability that a case in the FJC's grants data set would be adjudicated in either the *Conley* or the *Twombly/Iqbal* pleading regime under the null hypothesis, it again follows that any multinomial distribution that satisfies the null hypothesis has equal expected numbers of cases drawn from each pleading regime. In addition, since the null hypothesis implies $h_c=h_t$, the same share of cases will have defendants ultimately prevail under each pleading regime when the null holds. This shows that the null hypothesis again is consistent with the absence of all types of party selection, as well as with the absence of any difference in judicial behavior.

I now show that $NAS_{LB}$ has probability limit zero under the null hypothesis. Recall from footnote 157 of the main text that my lower bound can be written as

$$NAS_{LB} = \hat{g}_t - \hat{g}_c \frac{MTD_c}{MTD_t},$$

where $\hat{g}_t$ is the share of defendants who ultimately prevail among cases that are filed under the *Twombly/Iqbal* pleading regime, $\hat{g}_c$ is the same share among those filed under the *Conley* pleading regime, $MTD_c$ is the number of MTDs filed under the *Conley* regime, and $MTD_t$ is the number of MTDs filed under the *Twombly/Iqbal* regime.[3] By definition, we have $\hat{g}_t=Y_3/(Y_3+Y_4)$ and $\hat{g}_c=Y_1/(Y_1+Y_2)$. Dividing each numerator and each denominator by N yields $\hat{g}_t = \hat{h}_t/(\hat{h}_t + \hat{q}_t)$ and $\hat{g}_c = \hat{h}_c/(\hat{h}_c + \hat{q}_c)$, where hats denote sample frequencies corresponding to population probabilities. Dividing both the numerator and the denominator of the ratio ($MTD_c$ /$MTD_t$) by M shows that this ratio can be written as $(\hat{m}_c/\hat{m}_t)$, since $MTD_c=X_1$ and $MTD_t=X_3$, where $\hat{m}_c = X_1/M$ is the share of all cases in the filing data set that are filed under *Conley* and have MTDs filed, and analogously for $\hat{m}_t$.

Putting all of this together implies that I can write my lower bound as

$$NAS_{LB} = \frac{\hat{h}_t}{\hat{h}_t+\hat{q}_t} - \frac{\hat{h}_c}{\hat{h}_c+\hat{q}_c}\left(\frac{\hat{m}_c}{\hat{m}_t}\right). \qquad (B.1)$$

Every component of the right hand side of this equation is a sample frequency, and sample frequencies are consistent for their corresponding population probabilities. Therefore, each sample frequency converges in probability to its corresponding population probability, i.e., plim $\hat{h}_c = h_c$, plim $\hat{h}_t = h_t$, plim $\hat{q}_c = q_c$, plim $\hat{q}_t = q_t$, plim

3. The main text actually uses the subscripts “TI" and “Conley”, but in this Appendix I use “t” and “c” for short. In addition, the main text uses g, rather than $\hat{g}$, to refer to the share of cases in which defendants prevail. In this appendix I use the “hats” to emphasize that the quantities involved are estimates rather than population parameters. In addition, I

$\hat{m}_c = m_c$, and plim $\hat{m}_t = m_t$. By the Slutsky theorem, the probability limit of a continuous function of random variables is the continuous function of the probability limits of the random variables.[4] Therefore, the probability limit of $NAS_{LB}$ is $[h_t/(h_t+q_t) - h_c/(h_c+q_c)\times(m_c/m_c)]$. Recall that under the null hypothesis, $h_c=h_t=h$, $q_c=q_t=q$, and $m_c=m_t=m$. It thus follows that the probability limit of $NAS_{LB}$ is $[h/(h+q) - h/(h+q)\times(m/m)]= [h/(h+q) - h/(h+q)]$, which is zero, as I claimed above.

Next, I use the delta method to show that $NAS_{LB}$ has an asymptotically normal distribution.[5] Since $\hat{h}_t + \hat{q}_t = 1 - (\hat{h}_c + \hat{q}_c)$ by construction, we can re-write $NAS_{LB}$ as

$$NAS_{LB} = \frac{\hat{h}_t}{1-(\hat{h}_c+\hat{q}_c)} - \frac{\hat{h}_c}{\hat{h}_c+\hat{q}_c}\left(\frac{\hat{m}_c}{\hat{m}_t}\right). \qquad (B.2)$$

Let $\theta=(m_c, m_t, h_c, h_t, q_c)'$, and let $\hat{\theta} = (\hat{m}_c, \hat{m}_t, \hat{h}_c, \hat{h}_t, \hat{q}_c)'$ be its sample analog. It will be useful to distinguish between the parts of $\theta$ and $\hat{\theta}$ that involve parameters from the filing and grants datasets. Thus, I define - $\theta_g=(h_c, h_t, q_c)'$, $\hat{\theta}_g = (\hat{h}_c, \hat{h}_t, \hat{q}_c)'$, $\theta_m= (m_c,m_t)'$, and $\hat{\theta}_m = (\hat{m}_c, \hat{m}_t)'$, so that $\theta = (\theta'_m, \theta'_g)'$ and $\hat{\theta} = (\hat{\theta}'_m, \hat{\theta}'_g)'$. Below I will work with $\sqrt{N}(\hat{\theta} - \theta)$, which is complicated slightly by the fact that the relevant sample size for the filing dataset is M, rather than N. Observe that $\sqrt{N}(\hat{\theta}_m - \theta_m) = \sqrt{N/M} \times \sqrt{M}(\hat{\theta}_m - \theta_m)$. I will assume that $\lim_{N\to\infty}(N/M)=\lambda$ exists; that is, I assume that as the number of grants-dataset cases grows large, the ratio of this number to the number of filings-dataset cases converges to some finite constant, $\lambda$. By the product rule,[6] $\sqrt{N}(\hat{\theta}_m - \theta_m)$. has the same large-N behavior as $\sqrt{\lambda}\sqrt{M}(\hat{\theta}_m - \theta_m)$. Therefore, $\sqrt{N}(\hat{\theta} - \theta)$ has the same large-sample behavior as $\left[\lambda\sqrt{M}(\hat{\theta}_m - \theta_m), \sqrt{N}(\hat{\theta}_g - \theta_g)\right]$, which will be straightforward to work with.

We can write the parameter $NAS_{LB}$ as a function $f(\hat{\theta})$ that maps from the vector of sample parameters, $\hat{\theta}$, into the set of real numbers, i.e.,

$$f(\hat{\theta}) = \frac{\hat{h}_t}{1-(\hat{h}_c+\hat{q}_c)} - \frac{\hat{h}_c}{\hat{h}_c+\hat{q}_c}\left(\frac{\hat{m}_c}{\hat{m}_t}\right). \qquad (B.3)$$

---

4. *See* WILLIAM H. GREENE, ECONOMETRIC ANALYSIS, SIXTH EDITION, 1045 (2008) (Theorem D.12).

5. *See* Greene, Econometric Analysis, *supra* note 3, at 1055-1056 (Theorems D.21 and D.21A).

6. *See Id*., at 1049 (Theorem D.16, relation (D-13)).

Note that the probability limit of $NAS_{LB}$ is $f(\theta)=[h_t/(h_t+q_t) - h_c/(h_c+q_c)\times(m_c/m_c)]$, which we saw via the application of the Slutsky theorem above. It is clear by inspection that the function f is both continuous and twice-differentiable in the vector θ, provided that none of the denominators in (B.3) is identically zero. Therefore, we can use Taylor's theorem to obtain the following representation of $NAS_{LB}$:

$$NAS_{LB} = f(\hat{\theta}) = f(\theta) + (\hat{\theta} - \theta)' \nabla f(\theta^*), \quad \text{(B.4)}$$

where $\theta^*$ lies on the hyperplane between θ and $\hat{\theta}$. Subtracting f(θ) from both sides of this equation and multiplying by the square root of N, the grants data set sample size, then yields

$$\sqrt{N}[f(\hat{\theta}) - f(\theta)] = \sqrt{N}(\hat{\theta} - \theta)' \nabla f(\theta^*). \quad \text{(B.5)}$$

Again using the product rule, we see that the right hand side of (B.5) has the same large-sample behavior as $\sqrt{N}(\hat{\theta} - \theta)' \operatorname{plim} \nabla f(\theta^*)$. Because $\hat{\theta}$ is consistent for θ, and because $\theta^*$ lies on the hyperplane between $\hat{\theta}$ and θ, $\theta^*$ must also be consistent for θ. Therefore, $\operatorname{plim} \theta^* = \theta$. Again applying the Slutsky Theorem, we have $\operatorname{plim} \nabla f(\theta^*) = \nabla f(\operatorname{plim} \theta^*)$, so $\operatorname{plim} \nabla f(\theta^*) = \nabla f(\theta)$. Therefore, the right hand side of (B.5) has the same large-sample behavior as $\sqrt{N}(\hat{\theta} - \theta)' \nabla f(\theta)$, whose large-sample behavior is the same as $\left(\sqrt{\lambda}\sqrt{M}(\hat{\theta}_m - \theta_m)', \sqrt{N}(\hat{\theta}_g - \theta_g)\right) \nabla f(\theta)$, in light of the discussion *supra* concerning $\sqrt{N}(\hat{\theta} - \theta)$. The vectors $\hat{\theta}_m$ and $\hat{\theta}_g$ involve only sample proportions, which are sample means, so we can apply a central limit theorem to obtain the result that both $\sqrt{M}(\hat{\theta}_m - \theta_m)'$ and $\sqrt{N}(\hat{\theta}_g - \theta_g)$ are asymptotically normal. Thus, as N grows, the distribution of $\left[\sqrt{\lambda}\sqrt{M}(\hat{\theta}_m - \theta_m), \sqrt{N}(\hat{\theta}_g - \theta_g)\right] \nabla f(\theta)$ converges to a normal distribution with mean vector zero and variance matrix equal to

$$\nabla f(\theta)' \begin{bmatrix} \lambda\Omega_m & 0 \\ 0 & \Omega_g \end{bmatrix} \nabla f(\theta), \quad \text{(B.6)}$$

where $\Omega_m$ is the variance matrix for $\sqrt{M}(\hat{\theta}_m - \theta_m)'$ and $\Omega_g$ is the variance matrix for $\sqrt{N}(\hat{\theta}_g - \theta_g)$. The zero sub-matrices allow us to re-write (B.6) as

$$\lambda \nabla_m' \Omega_m \nabla_m + \nabla_g' \Omega_g \nabla_g, \quad \text{(B.7)}$$

where $\nabla_m$ is the 2-by-1 column vector of partial derivatives of the function f with respect to $m_c$ and $m_t$ and $\nabla_g$ is the 3-by-1 column vector of partial derivatives of the function f with respect to the elements of the vector $\theta_g$.

Recall that $\hat{\theta}_m = (\hat{m}_c, \hat{m}_t)'$, and that the $\hat{m}_c$ and $\hat{m}_t$ parameters are the fraction of filing-dataset cases that have Rule 12(b)(6) MTDs filed under the *Conley* and *Iqbal* regimes, respectively. Thus, $V(\hat{m}_c) = m_c(1 - m_c)/M$ and $V(\hat{m}_t) = m_t(1 - m_t)/M$, with $\mathrm{Cov}(\hat{m}_c, \hat{m}_t) = -m_c m_t/M$. Note that the M in each denominator is eliminated when we work with $\sqrt{M}(\hat{\theta}_m - \theta_m)$. In addition, under the null hypothesis we have $m_c=m_t=m$. Therefore, under the null hypothesis, the matrix $\Omega\Omega_m$ is given by

$$\Omega_m = \begin{bmatrix} m(1-m) & -m^2 \\ -m^2 & m(1-m) \end{bmatrix}. \tag{B.8}$$

Differentiating (B.3) partially and evaluating at the true population parameters shows that the partial derivatives of the function f with respect to $m_c$ and $m_t$ are

$$\frac{\partial f}{\partial m_c} = -\frac{g_c}{m_t} \text{ and } \frac{\partial f}{\partial m_t} = \frac{g_c m_c}{m_t^2},$$

and imposing the null hypothesis's requirement that $g_c=g_t=g$ and $m_c=m_t=m$, we have

$$\frac{\partial f}{\partial m_c} = -\frac{g}{m} \text{ and } \frac{\partial f}{\partial m_t} = \frac{g}{m},$$

so

$$\nabla_m = \frac{-g}{m}(1, \quad -1)'$$

under the null hypothesis. Therefore, under the null hypothesis the first part of (B.7) equals[7]

7. To reach the second equality, I note that using the basic algebra of matrix multiplication,

$$(1, \quad -1)\Omega_m \begin{pmatrix} 1 \\ -1 \end{pmatrix} = 2m(1-m) + 2m^2 = 2m.$$

When we multiply this result by $\lambda(g/m)^2$ , the result is $V_m = 2\lambda g^2/m$, as stated in (B.9).

$$\begin{aligned} V_m \equiv \lambda \nabla'_m \Omega_m \nabla_m &= \lambda \left(\frac{g}{m}\right)^2 (1, \quad -1)\Omega_m \begin{pmatrix} 1 \\ -1 \end{pmatrix} \\ &= \frac{2\lambda g^2}{m}. \end{aligned} \tag{B.9}$$

Because each element of $\theta_g$ involves a proportion, the analysis for $\Omega_g$ is similar to that for $\Omega_m$. After imposing the null hypothesis, the covariance matrix of $\sqrt{N}(\hat{\theta}_g - \theta_g)$ can be shown to equal

$$\Omega_g = \begin{bmatrix} h(1-h) & -h^2 & -hq \\ -h^2 & h(1-h) & -hq \\ -qh & -qh & q(1-q) \end{bmatrix}. \tag{B.10}$$

Differentiating (B.3) partially and again evaluating at the true population parameters yields

$$\begin{aligned} \frac{\partial f}{\partial h_c} &= \frac{h_t}{[1-(h_c+q_c)]^2} - \frac{q_c}{[h_c+q_c]^2}\left(\frac{m_c}{m_t}\right) &&=^{H_0} \quad 4(h-q), \\ \frac{\partial f}{\partial h_t} &= \frac{1}{1-(h_c+q_c)} &&=^{H_0} \quad 2, \\ \frac{\partial f}{\partial q_c} &= \frac{h_t}{[1-(h_c+q_c)]^2} - \frac{-h_c}{[h_c+q_c]^2}\left(\frac{m_c}{m_t}\right) &&=^{H_0} \quad 8h, \end{aligned} \tag{B.11}$$

where the notation "$=^{H_0}$" indicates that an equality holds under the null hypothesis, though not generally otherwise.[8] Thus, $\nabla_g = (4[h-q], \quad 2, \quad 8h)'$ under the null.

Unlike $\lambda V_m = \nabla'_m \Omega_m \nabla_m$, which we have seen equals $(2\lambda g^2/m)$ under the null hypothesis, the term

$$V_g \equiv \nabla'_g \Omega_g \nabla_g \tag{B.12}$$

does not generally simplify neatly. Therefore, I simply observe that under the null hypothesis, the limiting variance of $\sqrt{N}[f(\hat{\theta}) - f(\theta)]$, which we have seen is also the limiting variance of $\sqrt{N}(NAS_{LB})$, is

$$\lim_{N\to\infty} V\left(\sqrt{N}[f(\hat{\theta}) - f(\theta)]\right) = \lim_{N\to\infty} V\left(\sqrt{N}[NAS_{LB}]\right) = \frac{2\lambda g^2}{m} + V_g. \tag{B.13}$$

8. Recall that under the null hypothesis, we have $m_c$=$m_t$=m, $h_c$=$h_t$=h, and h+q=½.

To estimate the variance of $NAS_{LB}$ involves three further (entirely conventional) steps in the argument. First, I assume that N is large enough so that the actual variance of $\sqrt{N}(NAS_{LB})$ given a finite sample size N is well approximated by the limit that appears in (B.13). Second, observe that the variance of the product of (i) a non-stochastic scalar and (ii) a random variable equals the product of (iii) the square of the scalar and (iv) the variance of the random variable. Under the just-made assumption that the formula in (B.13) is an acceptable approximation to $V[\sqrt{N}(NAS_{LB})]$ for finite N, then, we can approximate $V(NAS_{LB})$ using

$$V(NAS_{LB}) \cong N^{-1}\left[\frac{2\lambda g^2}{m} + \nabla_g' \Omega_g \nabla_g\right]. \quad \text{(B.14)}$$

Third and finally, I replace the population parameters in this formula—λ, g, h, and q—with consistent estimates. This final step is appropriate, for large N, because of the Slutsky theorem. Thus when we replace population parameters in the right hand side of (B.13) with consistent estimators, the result is a consistent estimator of the limiting variance of $\sqrt{N}(NAS_{LB})$. It then follows that when we replace the population parameters on the right hand side of (B.14) with consistent estimators, the resulting formula will be a good estimator for $V(NAS_{LB})$. Thus, the actual estimator I use for the variance of $NAS_{LB}$ is:

$$\hat{V} \equiv N^{-1}\left[\frac{2\hat{\lambda}\hat{g}^2}{\hat{m}} + \hat{\nabla}_g' \hat{\Omega}_g \hat{\nabla}_g\right], \quad \text{(B.15)}$$

where $\hat{\nabla}_g' = (4[\hat{h} - \hat{q}], \;\; 2, \;\; 8\hat{h})$ and $\hat{\Omega}_g$ is the matrix that results from replacing h and q with $\hat{h}$ and $\hat{q}$ in (B.10). The estimated standard error that I use for $NAS_{LB}$, $\hat{\sigma}$, is then the square root of (B.15).

### *B. Estimating the Parameters λ, m, h, and q*

Next I discuss the estimators I use for λ, m, h, q, and g. Recall that λ is the limit value of the ratio of N to M. The obvious estimator for this limit is the ratio of actual N to actual M, so I use $\hat{\lambda} \equiv (M/N)$ to estimate λ. The bold rows in Appendix A Table 1 report the number of cases in the 2005/2006 and 2009/2010 time periods in each of the three case categories I consider. For example, there are 3795 employment discrimination cases in the earlier period and 3871 in the later period, yielding a total of N=7666 such cases. I report this value and the values for the total other cases and civil rights cases categories in the first row of Appendix B Table 1. Turning to M, the second row of Appendix B Table 1 reports the total numbers of cases from the FJC's grants data set for each of my three case categories. The value of M is the sum of the value in Appendix A Table 4's "Total" column for 2006 and the value in the corresponding column for 2010. For example, there are a total of 92 employment discrimination cases in the grants data set's 2006 period and 113 such cases in its 2010 period, so M equals 205 for employment discrimination cases. The third row of Appendix B Table 1 reports the estimated value of $\hat{\lambda}$ for each case category; these values range from 0.013 to 0.048.

Next consider m. Under the null hypothesis, $m=m_c=m_t$, where $m_c$ is the population probability that a case drawn for the filing data set will wind up (i) having an MTD filed (ii) under the *Conley* regime, and $m_t$ is the corresponding probability when we substitute *Iqbal* for *Conley* in (ii). Observe that the sum $(m_c+m_t)$ is the population probability that a case drawn for the filing data set has an MTD filed under *either* pleading regime. Under the null hypothesis, this sum equals 2m. Therefore, m equals one-half the probability that a case drawn for the filing data set has an MTD filed under one pleading regime or the other. This quantity can be estimated consistently using one-half the share of all cases in the filing data set that have MTDs filed.

I reported the number of MTDs filed by case category in the two columns of Appendix A Table 3, one for the 2005/2006 period and the other for the 2009/2010 period. Respectively, these columns correspond to the variables $X_1$ and $X_3$ I defined at the beginning of the current appendix. I report their sum in the fourth row of Appendix B Table 1, which shows that there are 611 MTDs filed in the employment discrimination category, 830 filed in my civil rights category, and 3209 MTDs filed in the total other cases category. Appendix B Table 1's fifth row then reports the corresponding estimates of m, which equal these $X_1$+ $X_3$ figures divided by twice the filing data set sample size. The resulting estimates of m are 0.040 for employment discrimination cases, 0.057 for civil rights cases, and 0.020 for total other cases.

Now I turn to estimating h, g, and q. Recall that $h_c$ is the population probability that a case drawn for the grants data set will have its motion to dismiss adjudicated under the *Conley* regime and granted, and $h_t$ is the analogous probability for the *Twombly/Iqbal* regime. Observe that ($h_c$ + $h_t$) is the population probability that a case drawn for the grants data set will have its MTD granted under one or the other of the pleading regimes. Since $h=h_c=h_t$ under the null hypothesis, it follows that this population probability equals 2h under the null. Consequently, the sample fraction of MTDs that are granted is a consistent estimator for 2h, so I estimate h using $\hat{h} \equiv [Y_1 + Y_3]/2N$, where I defined $Y_1$ and $Y_3$ above as the numbers of MTDs actually granted under the *Conley* and *Iqbal* pleading regimes, respectively. In addition, recall that g is the population probability, under the null hypothesis, that an MTD will actually be granted, regardless of the pleading regime under which it is adjudicated. Thus, we have g=2h, and g can be estimated consistently using $2\hat{h}$.

I reported the values of $Y_1$ and $Y_3$ in the second and fifth columns, respectively, of Appendix A Table 4. I report the *sum* of these figures in the sixth row of Appendix B Table 1. This row shows that defendants prevailed in a total of 125 of the MTDs in the grants data set for employment discrimination cases, 224 of the civil rights cases' MTDs, and 560 of the total other cases' MTDs. The seventh row of Appendix B Table 1 then reports the ratio of each category's number of MTDs granted to the number of MTDs adjudicated in the grants data set for that category, *i.e.*, M from the second row. These ratios are 0.610 for employment discrimination cases, 0.646 for civil rights cases, and 0.558 for total other cases.[9] The eighth row of Appendix B Table 1 reports the estimated value of h, which is just half the estimate for g: 0.305 for employment discrimination cases, 0.323 for civil rights cases, and 0.279 for total other case. Finally, to estimate q, recall that under the null hypothesis, (h+q)=½. Therefore, a consistent estimate of q may be calculated as $\hat{q} = \frac{1}{2} - \hat{h}$. I report these estimates in the final row of Appendix B Table 1.

9. For each category, notice that the estimate of g lies between the share of defendants who prevailed in the FJC's *Conley* and *Iqbal* periods (reported, *e.g.*, in Table 4 of the main text). This regularity occurs because the estimates of g can be shown to equal weighted averages of the two defendant-prevails rates. The same relationship holds if we compare twice the estimate of m to the MTD filing rates reported in Table 3 of the main text, since $2\hat{m}$ is a weighted average of these MTD filing rates. In each case, the weight applied to each *Conley* rate is the share of observations that fall in the *Conley* period, and analogously for the *Iqbal* rates.

### *C. Estimated Standard Errors for the Estimated Lower Bounds*

The first row of Appendix B Table 2 reports my estimates of $\hat{V}_m$, the part of the variance in $\sqrt{N}(NAS_{LB})$ that is due to variation in MTD filings, for each of the three case categories I consider. These estimates are 0.502 for employment discrimination cases, 0.703 for civil rights cases, and 0.405 for total other cases. The second row reports the estimates of $\hat{V}_g$, the part of the variance in my estimated lower bound that is due to variation in the rate at which defendants prevail in MTDs that appear in the grants data set. These estimates are quite similar across case category: 0.952 for employment discrimination cases, 0.915 for civil rights cases, and 0.987 for total other cases. The table's third row reports the sum $\hat{V} = \hat{V}_g + \hat{V}_m$, which is my estimate of the limiting total variance, $V[\sqrt{N}(NAS_{LB})]$. These estimates also are relatively similar: 1.454 for employment discrimination cases, 1.618 for civil rights cases, and 1.392 for total other cases.

The table's fourth row reports the number of cases in the grants data set, N, for each case category, repeated from Appendix B Table 1; there are 205 employment discrimination cases, 347 civil rights cases, and 1003 total other cases. The fifth row then reports the ratio of $\hat{V}$ to N, which is my estimate of the asymptotic variance of $NAS_{LB}$. These estimates are 0.0071 for employment discrimination cases, 0.0047 for civil rights cases, and 0.0014 for total other cases. The substantial relative differences in these estimated variances arise because $\hat{V}/N$ depends directly, and negatively, on the number of cases in the grants data set; there are nearly 70% more employment discrimination cases than civil rights cases, and there are nearly three times as many total other cases as employment discrimination cases.

The sixth row of Appendix B Table 2 reports my estimates of $\hat{\sigma} \equiv \sqrt{\hat{V}/N}$, the estimated standard error of $NAS_{LB}$ under the null hypothesis that *Twombly* and *Iqbal* changed nothing about pleading, including party behavior. These estimated standard errors are 0.084 for employment discrimination cases, 0.069 for civil rights cases, and 0.037 for total other cases. The table's seventh row reports the negatively affected share estimates (these are repeated from Table 6 of the main text), which are 0.154 for employment discrimination cases, 0.181 for civil rights cases, and 0.215 for total other cases.

Dividing the negatively affected share estimates by their estimated standard errors yields the conventional *z*-statistic. These statistics are

1.833 for employment discrimination cases, 2.263 for civil rights cases, and 5.811 for total other cases. Under the null hypothesis, this statistic should have an approximately standard normal distribution. Therefore, formal hypothesis testing can be based on whether these statistics are unusually far from zero, where the metric for "unusual" is given by the quantiles of the standard normal distribution.

The final row of Appendix B Table 2 reports one-sided *p*-values based on these *z* statistics, for testing the null hypothesis of no change in pleading against the alternative hypothesis of pleading changes consistent with the economic model of pleading I discuss in the main text.[10] These *p*-values are 0.033 for employment discrimination cases, 0.004 for civil rights cases, and zero to four digits for total other cases. Thus, all three of the negatively affected share estimates are statistically significant at conventional levels.[11]

---

10. I use one-sided *p*-values rather than two-sided because the negatively affected share can never be negative. Therefore, the relevant alternative hypothesis concerns not the possibility that $NAS_{LB} \neq 0$, but rather that $NAS_{LB} > 0$. As such, the appropriate *p*-value is found by determining the probability that a random draw from the standard normal distribution would exceed the reported *z* statistic (rather than the probability that the absolute value of such a random draw would exceed the reported *z* statistic, as with a two-sided *p*-value).

11. One might worry about multiple comparisons here. Since there are three separate *z*-statistics here, the probability that at least one of them exceeds the traditional critical value for a given significance level is greater than that significance level. Equivalently, the probability of observing at least one traditional (i.e., marginal) *p*-value less than or equal to a given significance level is greater than that significance level. Thus, if my *greatest z* statistic were, say, 1.833, then it would be inappropriate to conclude that any of the $NAS_{LB}$ estimates were significant. But since the *least z* statistic has that value, there is little question that these estimates are significant.

**Appendix B, Table 1.**

ESTIMATES OF PARAMETERS USED IN ESTIMATING $V_M$ AND $V_G$

| | EMPLOYMENT DISCRIMINATION | CIVIL RIGHTS | TOTAL OTHER |
|---|---|---|---|
| M | 7666 | 7242 | 79,200 |
| N | 205 | 347 | 1003 |
| $\hat{\lambda} = \frac{N}{M}$ | 0.027 | 0.048 | 0.013 |
| Number of MTDs Filed: $X_1 + X_3$ | 611 | 830 | 3209 |
| $\hat{m} = \frac{X_1 + X_3}{2N}$ | 0.040 | 0.057 | 0.020 |
| Number of Defendants Prevailing: $Y_1 + Y_3$ | 125 | 224 | 560 |
| $\hat{g} = \frac{Y_1 + Y_3}{M}$ | 0.610 | 0.646 | 0.558 |
| $\hat{h} = \frac{\hat{g}}{2}$ | 0.305 | 0.323 | 0.279 |
| $\hat{q} = \frac{1}{2} - \hat{h}$ | 0.195 | 0.177 | 0.221 |

**Appendix B, Table 2.**

VARIANCE ESTIMATES AND *Z*-RATIOS FOR THE ESTIMATED LOWER BOUND

| | EMPLOYMENT DISCRIMINATION | CIVIL RIGHTS | TOTAL OTHER |
|---|---|---|---|
| $\hat{V}_m = \frac{2\hat{\lambda}\hat{g}^2}{\hat{m}}$ | 0.502 | 0.703 | 0.405 |
| $\hat{V}_g = \hat{\nabla}'_g \hat{\Omega}_g \hat{\nabla}_g$ | 0.952 | 0.915 | 0.987 |
| $\hat{V} = \hat{V}_m + \hat{V}_g$ | 1.454 | 1.618 | 1.392 |
| N | 205 | 347 | 1003 |
| $\frac{1}{N}\hat{V}$ | 0.0071 | 0.0047 | 0.0014 |
| $\hat{\sigma} \equiv \sqrt{\frac{\hat{V}}{N}}$ | 0.084 | 0.069 | 0.037 |
| $NAS_{LB}$ | 0.154 | 0.181 | 0.215 |
| $\hat{z} = \frac{NAS_{LB}}{\hat{\sigma}}$ | 1.833 | 2.623 | 5.811 |
| One-sided $p$-value | 0.033 | 0.004 | 0.000 |