Online Appendix: Measuring Quality for Use in Incentive Schemes:

The Case of "Shrinkage" Estimators

Nirav Mehta

University of Western Ontario *

April 2, 2019

B   Cutoff Model Proofs and Extensions

B.1   Direct Conditioning on Class Size

The difference in administrator's value from using different teacher-quality estimators derives from the assumption that the administrator chooses a cutoff policy based on only test score information. Such a one-dimensional policy is quite simple and, therefore, is of considerable clear policy relevance; this demonstrated by Table A.1, which documents existing incentive schemes and shows that none condition on class size. Moreover, when compared with a policy that may also explicitly condition on class sizes, a test-score-based cutoff may attenuate issues of class size manipulation for the sake of affecting the administrator's posterior about the quality of a particular teacher. However, allowing the administrator to explicitly take into account class size may still be of interest. This section shows how the theoretical results in Section 3 would be affected.

Now suppose the administrator, instead of only indirectly taking it into account when maximizing her utility, could instead explicitly condition on class size n_i. If n_i was a strictly monotonic function of teacher quality θ, then the administrator could achieve a perfect

classification of teachers by inverting $n(\theta)$—even if she ignored all teachers' test scores. A more realistic case would allow for multiple teacher qualities for at least one class size. Suppose that the distribution of teacher qualities for each class size was normally distributed. Because the administrator can explicitly condition on class size she can hold a separate cutoff-based classification problem for each class size level; denote the administrator's value from using the fixed effects and empirical Bayes estimators as $v_{CP,n}^{FE}(\kappa)$ and $v_{CP,n}^{EB}(\kappa)$, respectively. Then by Proposition 1 the administrator would obtain the same value for either estimator given the desired cutoff κ, i.e., $v_{CP,n}^{FE}(\kappa) = v_{CP,n}^{EB}(\kappa)$ for all (n, κ). Therefore, we can without loss of generality consider only the fixed-effects estimator, with optimal cutoff policy c_n^{*FE}. Further note that the administrator's expected objective would be at least as high if she is allowed to split her original objective into one objective for each class size; if the cutoff for $c_{n_1}^{*FE} = c_{n_2}^{*FE}$ for all class sizes n_1, n_2, then her value under the separate class size scheme would be the same as that from her original objective.

B.2   Administrator's Problem with Infinite Precision

We want to prove that as the variance of the measurement error tends to 0 (which implies $\sigma_{\bar{\epsilon}} \to 0$) all teachers will be correctly categorized, giving $v_{CP}^{FE}(\kappa) = v_{CP}^{EB}(\kappa) = 1$ for all desired κ. First, consider the fixed effects estimator. The administrator's utility for a teacher with true quality θ under the fixed effects estimator and cutoff policy c is

$$u_{CP}(\theta, \hat{\theta}^{FE}; c, \kappa) = \alpha 1\{\hat{\theta}^{FE} \geq c | \theta \geq \kappa\} + (1-\alpha)1\{\hat{\theta}^{FE} < c | \theta < \kappa\} \xrightarrow{p} \alpha 1\{\theta \geq c | \theta \geq \kappa\} + (1-\alpha)1\{\theta < c | \theta < \kappa\},$$
$$(1)$$

which is maximized at $c = \kappa$. The administrator's utility from using cutoff policy c under the empirical Bayes estimator, for the same teacher, is

$$plim_{\sigma_{\bar{\epsilon}} \to 0} u_{CP}(\theta, \hat{\theta}^{EB}; c, \kappa) = \alpha 1\{\hat{\theta}^{EB} \geq c | \theta \geq \kappa\} + (1-\alpha)1\{\hat{\theta}^{EB} < c | \theta < \kappa\}$$
$$= \alpha 1\{\lambda(\theta)\theta \geq c | \theta \geq \kappa\} + (1-\alpha)1\{\lambda(\theta)\theta < c | \theta < \kappa\}, \quad (2)$$

which is maximized at $c = \kappa/\lambda(F^{-1}(\kappa))$. The probabilities of the events in both (1) and (2) are all 1, giving an expected utility of 1 for all teacher qualities, which then integrates to a value of 1 for each estimator.

B.3 Proof of Proposition 2

Recall that we are considering first the case where $\kappa, c^{*EB} < 0$. Differentiating the administrator's value with respect to β_- and evaluating at $\beta_- = 0$, our goal is to show when

$$
\frac{\partial v_{CP}^{EB}}{\partial \beta_-}\bigg|_{\beta_-=0} = (1-\alpha) \int_{-\infty}^{\kappa} \frac{-c^{*EB}\theta}{(\delta_-)^2 \overline{\sigma_\epsilon}} \phi\left(\frac{\theta - \frac{c^{*EB}}{\delta_-}}{\overline{\sigma_\epsilon}}\right) \frac{\phi(\theta/\sigma_\theta)}{\sigma_\theta \Phi(\kappa/\sigma_\theta)} d\theta
$$

$$
+ \alpha \int_{\kappa}^{0} \frac{c^{*EB}\theta}{(\delta_-)^2 \overline{\sigma_\epsilon}} \phi\left(\frac{\theta - \frac{c^{*EB}}{\delta_-}}{\overline{\sigma_\epsilon}}\right) \frac{\phi(\theta/\sigma_\theta)}{\sigma_\theta (1 - \Phi(\kappa/\sigma_\theta))} d\theta < 0. \tag{3}
$$

The conjugate nature of the normal distribution (Bromiley (2003)) allows us to combine the above densities into one Gaussian density,[1] letting us write $\frac{1}{\overline{\sigma_\epsilon}}\phi\left(\frac{\frac{c^{*EB}}{\delta_-}-\theta}{\overline{\sigma_\epsilon}}\right) \times \frac{1}{\sigma_\theta}\phi(\frac{\theta}{\sigma_\theta}) = \frac{f_P(\theta)}{S_P}$, where S_P is a positive constant and

$$
f_P(\theta) = \frac{1}{\sigma_P} \phi\left(\frac{\theta - \mu_P}{\sigma_P}\right)
$$

$$
\mu_P = \frac{\frac{c^{*EB}}{\delta_-}\sigma_\theta^2 + 0 \times \overline{\sigma_\epsilon}^2}{\overline{\sigma_\epsilon}^2 + \sigma_\theta^2} = \frac{\frac{c^{*EB}}{\delta_-}\sigma_\theta^2}{\overline{\sigma_\epsilon}^2 + \sigma_\theta^2} = c^{*EB}
$$

$$
\sigma_P^2 = \frac{\overline{\sigma_\epsilon}^2 \sigma_\theta^2}{\overline{\sigma_\epsilon}^2 + \sigma_\theta^2},
$$

where the last equality on the second line follows because $\delta_- = \lambda(\theta)\big|_{\beta_-=0} = \frac{\sigma_\theta^2}{\overline{\sigma_\epsilon}^2 + \sigma_\theta^2}$. Dividing through by S_P, we have $\frac{\partial v_{CP}^{EB}}{\partial \beta_-}\bigg|_{\beta_-=0} < 0$ if

$$
\frac{\int_{-\infty}^{\kappa} \frac{c^{*EB}\theta}{(\delta_-)^2} f_P(\theta) d\theta}{\int_{\kappa}^{0} \frac{c^{*EB}\theta}{(\delta_-)^2} f_P(\theta) d\theta} > \left(\frac{\alpha}{1-\alpha}\right)\left(\frac{\Phi(\kappa/\sigma_\theta)}{1 - \Phi(\kappa/\sigma_\theta)}\right) \Rightarrow
$$

$$
\frac{\int_{-\infty}^{\kappa} \theta f_P(\theta) d\theta}{\int_{\kappa}^{0} \theta f_P(\theta) d\theta} > \left(\frac{\alpha}{1-\alpha}\right)\left(\frac{\Phi(\kappa/\sigma_\theta)}{1 - \Phi(\kappa/\sigma_\theta)}\right). \tag{4}
$$

---

[1] If $X_1 \sim F_1 = N\left(\mu_1, \sigma_1^2\right)$, with density f_1 and $X_2 \sim F_2 = N\left(\mu_2, \sigma_2^2\right)$, with density f_2, then $f_1(\theta) \times f_2(\theta) = \frac{f_P(\theta)}{S_P}$, where

$$
f_P(\theta) = \frac{1}{\sigma_P} \phi\left(\frac{\theta - \mu_P}{\sigma_P}\right)
$$

$$
\mu_P = \frac{\mu_1 \sigma_2^2 + \mu_2 \sigma_1^2}{\sigma_1^2 + \sigma_2^2}
$$

$$
\sigma_P^2 = \frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2},
$$

and S_P is a positive and constant scaling factor that depends on $\mu_1, \mu_2, \sigma_1, \sigma_2$ according to $S_p = \phi\left(\frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2 + \sigma_2^2}}\right)$.

The top and bottom terms on the left side of (4) are expectations of truncated normal random variables, scaled by truncation probabilities, i.e.,

$$\int_{-\infty}^{\kappa} \theta f_P(\theta) d\theta = (F_P(\kappa) - F_P(-\infty)) \operatorname{E}_{F_P}[\theta | \theta < \kappa]$$

$$= (F_P(\kappa) - F_P(-\infty)) \left(\mu_P + \sigma_P \frac{\phi(\frac{-\infty - \mu_P}{\sigma_P}) - \phi(\frac{\kappa - \mu_P}{\sigma_P})}{\Phi(\frac{\kappa - \mu_P}{\sigma_P}) - \Phi(\frac{-\infty - \mu_P}{\sigma_P})} \right)$$

$$= \Phi\left(\frac{\kappa - \mu_P}{\sigma_P} \right) \left(\mu_P + \sigma_P \frac{-\phi(\frac{\kappa - \mu_P}{\sigma_P})}{\Phi(\frac{\kappa - \mu_P}{\sigma_P})} \right)$$

and

$$\int_{\kappa}^{0} \theta f_P(\theta) d\theta = (F_P(0) - F_P(\kappa)) \operatorname{E}_{F_P}[\theta | \kappa < \theta < 0]$$

$$= (F_P(0) - F_P(\kappa)) \left(\mu_P + \sigma_P \frac{\phi(\frac{\kappa - \mu_P}{\sigma_P}) - \phi(\frac{-\mu_P}{\sigma_P})}{\Phi(\frac{-\mu_P}{\sigma_P}) - \Phi(\frac{\kappa - \mu_P}{\sigma_P})} \right)$$

$$= \left(\Phi\left(\frac{-\mu_P}{\sigma_P} \right) - \Phi\left(\frac{\kappa - \mu_P}{\sigma_P} \right) \right) \left(\mu_P + \sigma_P \frac{\phi(\frac{\kappa - \mu_P}{\sigma_P}) - \phi(\frac{-\mu_P}{\sigma_P})}{\Phi(\frac{-\mu_P}{\sigma_P}) - \Phi(\frac{\kappa - \mu_P}{\sigma_P})} \right).$$

Putting the above expressions back into the comparison (4) and re-arranging, we have

$$\underbrace{\frac{\left(\mu_P + \sigma_P \frac{-\phi(\frac{\kappa - \mu_P}{\sigma_P})}{\Phi(\frac{\kappa - \mu_P}{\sigma_P})} \right)}{\left(\mu_P + \sigma_P \frac{\phi(\frac{\kappa - \mu_P}{\sigma_P}) - \phi(\frac{-\mu_P}{\sigma_P})}{\Phi(\frac{-\mu_P}{\sigma_P}) - \Phi(\frac{\kappa - \mu_P}{\sigma_P})} \right)}}_{\#1} \times \underbrace{\left(\frac{1 - \Phi\left(\frac{\kappa}{\sigma_\theta} \right)}{\Phi\left(\frac{\kappa}{\sigma_\theta} \right)} \right)}_{\#2} \times \underbrace{\frac{\Phi\left(\frac{\kappa - \mu_P}{\sigma_P} \right)}{\left(\Phi\left(\frac{-\mu_P}{\sigma_P} \right) - \Phi\left(\frac{\kappa - \mu_P}{\sigma_P} \right) \right)}}_{\#3} > \frac{\alpha}{1 - \alpha}. \quad (5)$$

Recall that we are considering the case where $\kappa < 0$ and $c^{*EB} < 0$.[2] Therefore, both the numerator and denominator of expression #1 are negative, with the numerator greater in absolute value than the denominator, which implies that expression #1 > 1. Moreover, $\kappa < 0$ implies that $\Phi(\kappa / \sigma_\theta) < 1/2$, which implies that expression #2 > 1 as well. For example, consider $\alpha \le 1/2$. In these cases, the expressions #1 and #2 would satisfy the above inequality. However, the expression #3 could potentially be less than 1.

Due to our well-known lack of a closed form expression for the standard normal CDF Φ, it is impossible to sign the above inequality in a purely analytical manner for all possible cases. Moreover, the inequality would be violated for extreme parameterizations, such as where the administrator only placed value on one type of error (making α go to either the

---

[2]I show below that similar conditions obtain when considering $\kappa > 0$ and $c^{*EB} > 0$.

corner of zero or one). However, it is still possible to show why Condition (5) would likely hold for a wide range of reasonable values of $(\alpha, \kappa, \sigma_\theta, \overline{\sigma_\epsilon}, c^{*EB})$.

To start, assume $\alpha = 1/2$, in which case what is left to be shown is that the left side of the inequality (5) is greater than one (in Appendix B.4, I use numerical methods to verify that estimator rankings consistent with (5) hold for a wide range of α). For simplicity, express the cutoff policy as a fraction of the desired cutoff κ, i.e., $c^{EB} = \gamma_c \kappa$, where $\gamma_c \in [0, 1]$ (i.e., c^{EB} will be the same sign as κ and no larger in absolute value than κ) and write $\mu_P = c^{*EB} = \gamma_c \kappa$. Because #1 is always greater than one, it is sufficient to satisfy the following condition:

$$1 < \frac{1 - \Phi\left(\frac{\kappa}{\sigma_\theta}\right)}{\Phi\left(\frac{\kappa}{\sigma_\theta}\right)} \times \frac{\Phi\left(\frac{\kappa - \mu_P}{\sigma_P}\right)}{\Phi\left(\frac{-\mu_P}{\sigma_P}\right) - \Phi\left(\frac{\kappa - \mu_P}{\sigma_P}\right)} \Rightarrow \frac{\Phi\left(\frac{-\gamma_c}{\sqrt{\gamma_\sigma}}\kappa_\sigma\right)}{\Phi\left(\frac{(1-\gamma_c)}{\sqrt{\gamma_\sigma}}\kappa_\sigma\right)} < \frac{1}{\Phi(\kappa_\sigma)}, \qquad (6)$$

where $\kappa_\sigma \equiv \kappa/\sigma_\theta$ expresses the desired cutoff in terms of standard deviations of teacher quality and $\gamma_\sigma \equiv \frac{\overline{\sigma_\epsilon^2}}{\overline{\sigma_\epsilon^2} + \sigma_\theta^2} \in (0, 1)$.

With the above simplifications, whether Condition (6) will be satisfied (which is sufficient for satisfying Condition (5) when $\alpha = 1/2$) depends on $(\gamma_\sigma, \gamma_c, \kappa_\sigma) \in (0, 1) \times [0, 1] \times (-\infty, 0]$. The following cases analyze when $(\gamma_\sigma, \gamma_c, \kappa_\sigma)$ satisfy Condition (6):

Case a: $\kappa_\sigma \to 0$. For all $(\gamma_\sigma, \gamma_c) \in (0, 1) \times [0, 1]$, Condition (6) will also be satisfied as $\kappa_\sigma \to 0$, as the left side tends to one and the right side tends to two.

Case b: $\kappa_\sigma < 0$. There are two relevant subcases:

Case b.i: $1 - \gamma_c \leq \sqrt{\gamma_\sigma}$.

Lemma 1. *Condition (6) will be satisfied if $1 - \gamma_c \leq \sqrt{\gamma_\sigma}$.*

Proof. The numerator of the left side of (6) will always be smaller than the numerator of the right for any finite κ_σ. If $1 - \gamma_c \leq \sqrt{\gamma_\sigma}$ then we have $\frac{(1-\gamma_c)}{\sqrt{\gamma_\sigma}}\kappa_\sigma \geq \kappa_\sigma \Rightarrow \Phi\left(\frac{(1-\gamma_c)}{\sqrt{\gamma_\sigma}}\kappa_\sigma\right) \geq \Phi(\kappa_\sigma)$, i.e. the denominator of the left will be at least as large as the denominator of the right. $\square$

Case b.ii: $1 - \gamma_c > \sqrt{\gamma_\sigma}$.

Lemma 2. *There exists a threshold $\hat{\gamma}_\sigma(\kappa_\sigma, \gamma_c) \in (0, 1)$ such that, given $(\kappa_\sigma, \gamma_c)$, Condition (6) is met for $\gamma_\sigma > \hat{\gamma}_\sigma(\kappa_\sigma, \gamma_c)$.*

Proof. Note that $(\gamma_\sigma, \gamma_c)$ only enter the left side of (6). The limit of the left side of Condition (6) as $\gamma_\sigma \to 0$ is $\Phi(\infty)/\Phi(-\infty) = 1/0 = \infty$, which is greater than the right side for any finite

5

κ_σ. As $\gamma_\sigma \to 1$, the above sufficient condition $1 - \gamma_c \le \sqrt{\gamma_\sigma}$ is satisfied for $\gamma_c \in (0,1)$ because $1 - \gamma_c \le 1 \Leftrightarrow \gamma_c \ge 0$, meaning the left side is less than the right side. Because the left side is continuous in γ_σ, there exists a $\hat\gamma_\sigma(\kappa_\sigma, \gamma_c)$ by the Intermediate Value Theorem such that $G(\kappa_\sigma, \gamma_c, \hat\gamma_\sigma(\kappa_\sigma, \gamma_c)) \equiv \dfrac{\Phi\left(\frac{-\gamma_c}{\sqrt{\hat\gamma_\sigma(\kappa_\sigma, \gamma_c)}}\kappa_\sigma\right)}{\Phi\left(\frac{(1-\gamma_c)}{\sqrt{\hat\gamma_\sigma(\kappa_\sigma, \gamma_c)}}\kappa_\sigma\right)} - \dfrac{1}{\Phi(\kappa_\sigma)} = 0$. The solution to G will be unique if the left side of (6) is monotonically decreasing in γ_σ. The derivative of the left side with respect to γ_σ is

$$
\frac{\partial\left[\frac{\Phi\left(\frac{-\gamma_c}{\sqrt{\gamma_\sigma}}\kappa_\sigma\right)}{\Phi\left(\frac{(1-\gamma_c)}{\sqrt{\gamma_\sigma}}\kappa_\sigma\right)}\right]}{\partial\gamma_\sigma} = 2\frac{\kappa_\sigma}{\sqrt{\gamma_\sigma}}\left[\frac{\gamma_c\phi\left(\frac{-\gamma_c}{\sqrt{\gamma_\sigma}}\kappa_\sigma\right)\Phi\left(\frac{(1-\gamma_c)}{\sqrt{\gamma_\sigma}}\kappa_\sigma\right) + (1-\gamma_c)\phi\left(\frac{(1-\gamma_c)}{\sqrt{\gamma_\sigma}}\kappa_\sigma\right)\Phi\left(\frac{-\gamma_c}{\sqrt{\gamma_\sigma}}\kappa_\sigma\right)}{\Phi\left(\frac{(1-\gamma_c)}{\sqrt{\gamma_\sigma}}\kappa_\sigma\right)^2}\right],
$$

$$(7)$$

which is negative because the term in brackets is positive and $\kappa_\sigma < 0$. $\qquad\square$

Lemma 3. *The threshold $\hat\gamma_\sigma(\kappa_\sigma, \gamma_c)$ is decreasing in γ_c.*

Proof. Having established that there exists a cutoff $\hat\gamma_\sigma$ solving G in Lemma 2, we can implicitly differentiate G around the solution to obtain

$$
\frac{\partial\hat\gamma_\sigma}{\partial\gamma_c} = -\frac{\frac{\partial G}{\partial\gamma_c}}{\frac{\partial G}{\partial\gamma_\sigma}} = -\frac{\dfrac{\partial\left[\frac{\Phi\left(\frac{-\gamma_c}{\sqrt{\gamma_\sigma}}\kappa_\sigma\right)}{\Phi\left(\frac{(1-\gamma_c)}{\sqrt{\gamma_\sigma}}\kappa_\sigma\right)}\right]}{\partial\gamma_c}}{\dfrac{\partial\left[\frac{\Phi\left(\frac{-\gamma_c}{\sqrt{\gamma_\sigma}}\kappa_\sigma\right)}{\Phi\left(\frac{(1-\gamma_c)}{\sqrt{\gamma_\sigma}}\kappa_\sigma\right)}\right]}{\partial\gamma_\sigma}}.
$$

The derivative in the denominator, (7), has been shown to be negative, which means that $\frac{\partial\hat\gamma_\sigma}{\partial\gamma_c}$ will have the same sign as the derivative in the numerator (because the entire expression is multiplied by negative one):

$$
\frac{\partial\left[\frac{\Phi\left(\frac{-\gamma_c}{\sqrt{\gamma_\sigma}}\kappa_\sigma\right)}{\Phi\left(\frac{(1-\gamma_c)}{\sqrt{\gamma_\sigma}}\kappa_\sigma\right)}\right]}{\partial\gamma_c} = \frac{\kappa_\sigma}{\sqrt{\gamma_\sigma}}\left[\frac{\phi\left(\frac{(1-\gamma_c)}{\sqrt{\gamma_\sigma}}\kappa_\sigma\right)\Phi\left(\frac{-\gamma_c}{\sqrt{\gamma_\sigma}}\kappa_\sigma\right) - \phi\left(\frac{-\gamma_c}{\sqrt{\gamma_\sigma}}\kappa_\sigma\right)\Phi\left(\frac{(1-\gamma_c)}{\sqrt{\gamma_\sigma}}\kappa_\sigma\right)}{\Phi\left(\frac{(1-\gamma_c)}{\sqrt{\gamma_\sigma}}\kappa_\sigma\right)^2}\right].
$$

The denominator of the term in the brackets is positive iff

$$
\phi\left(\frac{(1-\gamma_c)}{\sqrt{\gamma_\sigma}}\kappa_\sigma\right)\Phi\left(\frac{-\gamma_c}{\sqrt{\gamma_\sigma}}\kappa_\sigma\right) > \phi\left(\frac{-\gamma_c}{\sqrt{\gamma_\sigma}}\kappa_\sigma\right)\Phi\left(\frac{(1-\gamma_c)}{\sqrt{\gamma_\sigma}}\kappa_\sigma\right) \Leftrightarrow \frac{\phi\left(\frac{(1-\gamma_c)}{\sqrt{\gamma_\sigma}}\kappa_\sigma\right)}{\Phi\left(\frac{(1-\gamma_c)}{\sqrt{\gamma_\sigma}}\kappa_\sigma\right)} > \frac{\phi\left(\frac{-\gamma_c}{\sqrt{\gamma_\sigma}}\kappa_\sigma\right)}{\Phi\left(\frac{-\gamma_c}{\sqrt{\gamma_\sigma}}\kappa_\sigma\right)},
$$

which is satisfied because each side represents the reverse hazard function for the normal

distribution $r(x) = \frac{\phi(x)}{\Phi(x)}$, which is decreasing (see Theorem 17 of Chechile (2011)), and we have $\frac{(1-\gamma_c)}{\sqrt{\gamma_\sigma}}\kappa_\sigma < 0 < \frac{-\gamma_c}{\sqrt{\gamma_\sigma}}\kappa_\sigma$. $\qquad\square$

Putting the above together, although it is not possible to obtain explicit conditions that exactly characterize the set of parameters satisfying the desired inequality, it is possible to show that the inequality would be satisfied for a wide range of parameters. Table B.1 summarizes the cases considered above. First, all $(\gamma_c, \gamma_\sigma) \in (0,1) \times [0,1]$ satisfy the inequality as $\kappa_\sigma \to 0$ (case a). Intuitively, as $\kappa_\sigma \to 0$, the countervailing decrease in Type I errors vanishes, meaning the increase in Type II errors will necessarily dominate. For $\kappa_\sigma < 0$ (case b), we know that there exists a cutoff value of γ_σ for each $(\kappa_\sigma, \gamma_c)$ such that the inequality will be satisfied, and that this cutoff is decreasing in γ_c (making it more likely to hold). If γ_c is large enough (case b.i) then κ_σ does not need to be considered to verify whether the inequality will be satisfied. Otherwise (case b.ii), whether the inequality will be satisfied depends on $(\gamma_\sigma, \gamma_c, \kappa_\sigma)$. Intuitively, as $1 - \gamma_c$ increases (relative to $\sqrt{\gamma_\sigma}$) the weights $\phi\left(\frac{\theta - \frac{c^{*EB}}{\delta_-}}{\sigma_{\bar{\epsilon}}}\right)$ in (3) increase more for larger values of $|\theta|$, causing the increase in Type II errors to be larger than the decrease in Type I errors.

<div align="center">Table B.1: Summary of cases</div>

Case	Parameterization	Will (5) be satisfied?	Proof
a	$\kappa_\sigma \to 0$	Yes	Direct (see page 5 of text)
b.i	$\kappa_\sigma < 0$ and $1 - \gamma_c \le \sqrt{\gamma_\sigma}$	Yes	Lemma 1
b.ii	$\kappa_\sigma < 0$ and $1 - \gamma_c > \sqrt{\gamma_\sigma}$	Depends on $(\gamma_\sigma, \gamma_c, \kappa_\sigma)$	Lemma 2

Therefore, it is instructive to compute the "worst-case" cutoff where $\gamma_c = 0$. Setting $\gamma_c = 0$ and rearranging (6), we obtain

$$\frac{\Phi(0)}{\Phi\left(\frac{\kappa_\sigma}{\sqrt{\gamma_\sigma}}\right)} < \frac{1}{\Phi(\kappa_\sigma)} \Leftrightarrow \gamma_\sigma > \left(\frac{\kappa_\sigma}{\Phi^{-1}\left(\frac{\Phi(\kappa_\sigma)}{2}\right)}\right)^2,$$

where the sign reversal in the last inequality is due to $\Phi^{-1}\left(\frac{\Phi(\kappa_\sigma)}{2}\right) < 0$. Take for example $\kappa_\sigma = -1$, meaning the administrator wishes to identify the bottom 16% teachers. In this case, $\hat{\gamma}_\sigma(\kappa_\sigma = -1, \gamma_c = 0) \approx 0.50$. Several empirical studies find that test scores are comprised of roughly equal parts signal and noise (see, e.g., Staiger and Rockoff (2010), which means we can use $\gamma_\sigma = 1/2$ as a rough approximation. This, combined with the fact that $\hat{\gamma}_\sigma$ would only decrease as γ_c increased from its lower bound of zero, makes it very likely that Condition (6) would be satisfied.

Figure B.1: Cutoff $\hat{\gamma}_\sigma$ when $\kappa_\sigma = -1$ and $\kappa_\sigma = -2$
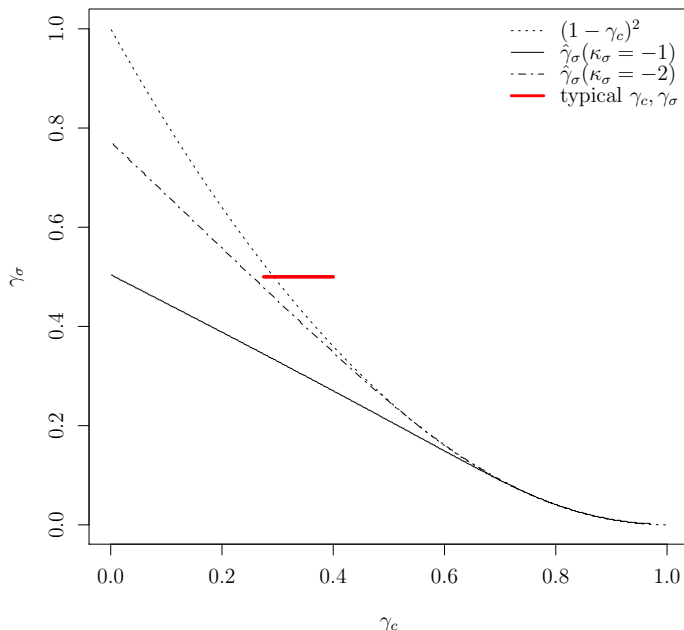
Figure B.1 plots the sufficient condition $\gamma_\sigma \geq (1 - \gamma_c)^2$ (dotted curve) and the κ_σ-dependent $\hat{\gamma}_\sigma$ when κ_σ is -1 (solid curve) and -2 (dot-dashed curve). Any value of γ_σ above a scenario-specific curve would satisfy Condition (6) for that scenario. For example, the result that all values of γ_σ would satisfy the inequality as $\gamma_c \to 1$ can be seen as the sufficient condition goes down to zero when $\gamma_c \to 1$. As κ_σ increases in absolute value the cutoff $\hat{\gamma}_\sigma$ increases, where $\hat{\gamma}_\sigma$ approaches the sufficient condition $(1 - \gamma_c)^2$ as $|\kappa_\sigma| \to \infty$.

Across a wide variety of class size and teacher quality scenarios and desired cutoffs (κ), the typical value of γ_c is in the range of 0.275 to 0.4. For example, the 25th and 75th percentiles of γ_c are, respectively, 0.286 and 0.30 when using the calibrated relationship between class size and teacher quality for reading in the LAUSD. The red segment shows the range of typical values of γ_c, coupled with the typical value of $\gamma_\sigma \approx 1/2$. Much of the red segment lies in the "sufficiency" part of the plot, meaning Condition (6) would be satisfied for any non-positive κ_σ. We can see that the typical values of $(\gamma_c, \gamma_\sigma)$ would also easily satisfy Condition (6) when $\kappa_\sigma = -1$ and $\kappa_\sigma = -2$.

It is important to remember that Condition (6) is only a sufficient condition for the derivative of the value under empirical Bayes with respect to β_- being negative, as it ignores the first term in (5). For example, consider $\gamma_c = 0.275$ and $\gamma_\sigma = 0.5$. Setting γ_c to the lowest value in the typical range is conservative as it makes (6) harder to satisfy. The left side

8

of (6) decreases in κ more quickly than the right side of (6), meaning a low (i.e., extreme negative) desired cutoff would make it harder to satisfy the inequality.[3] Choosing $\kappa = -5\sigma_\theta$, i.e., the desired cutoff was five standard deviations below the mean teacher quality, the left side numerically evaluates to 1.91. Notably, this extreme cutoff is far beyond any existing or proposed teacher incentive scheme. Thus, we can be confident that the desired inequality would be satisfied for reasonable parameter values.

Now suppose that $\kappa > 0$ and $c^{*EB} > 0$. For the negative-quadratic $n(\theta)$ to result in lower value from using empirical Bayes for all κ, the desired inequality would be $\left.\frac{\partial v_{CP}^{EB}}{\partial \beta_+}\right|_{\beta_+=0} > 0$. The (sufficient) analogue of the sufficient Condition (6) in this case would be

$$\frac{\Phi(\kappa_\sigma)}{1-\Phi(\kappa_\sigma)} > \frac{\Phi\left(\frac{(1-\gamma_c)}{\sqrt{\gamma_\sigma}}\kappa_\sigma\right) - \Phi\left(\frac{-\gamma_c}{\sqrt{\gamma_\sigma}}\kappa_\sigma\right)}{1-\Phi\left(\frac{(1-\gamma_c)}{\sqrt{\gamma_\sigma}}\kappa_\sigma\right)} \Leftrightarrow \frac{1-\Phi(-\kappa_\sigma)}{\Phi(-\kappa_\sigma)} > \frac{\Phi\left(\frac{-\gamma_c}{\sqrt{\gamma_\sigma}}(-\kappa_\sigma)\right) - \Phi\left(\frac{(1-\gamma_c)}{\sqrt{\gamma_\sigma}}(-\kappa_\sigma)\right)}{\Phi\left(\frac{(1-\gamma_c)}{\sqrt{\gamma_\sigma}}(-\kappa_\sigma)\right)}$$

$$\Leftrightarrow \frac{1}{\Phi(-\kappa_\sigma)} > \frac{\Phi\left(\frac{-\gamma_c}{\sqrt{\gamma_\sigma}}(-\kappa_\sigma)\right)}{\Phi\left(\frac{(1-\gamma_c)}{\sqrt{\gamma_\sigma}}(-\kappa_\sigma)\right)},$$

which, since $-\kappa_\sigma < 0$, is equivalent to Condition (6). Thus, the conditions on $(\gamma_c, \gamma_\sigma)$ are identical, given $|\kappa_\sigma|$.

B.4   Asymmetric Type I and Type II Weights

The administrator's preferred estimator in the cutoff-based model is not very sensitive to α being close to 1/2. Figure B.2 plots the ratio of the administrator's value under fixed effects and empirical Bayes, by class size scenario $n(\theta)$ and desired cutoff κ, for different values of the Type I error weight. Figure B.2a shows the ratio in administrator's value when $\alpha = 1/4$, i.e., the administrator values Type I errors one-third as much as she values Type II errors. Figure B.2b shows the value ratio when $\alpha = 2/3$, i.e., the administrator values Type I errors twice as much as Type II errors. In both plots, we can see that the relative ranking of the estimators is the same as it was under the symmetric weight, $\alpha = 1/2$, scenario.
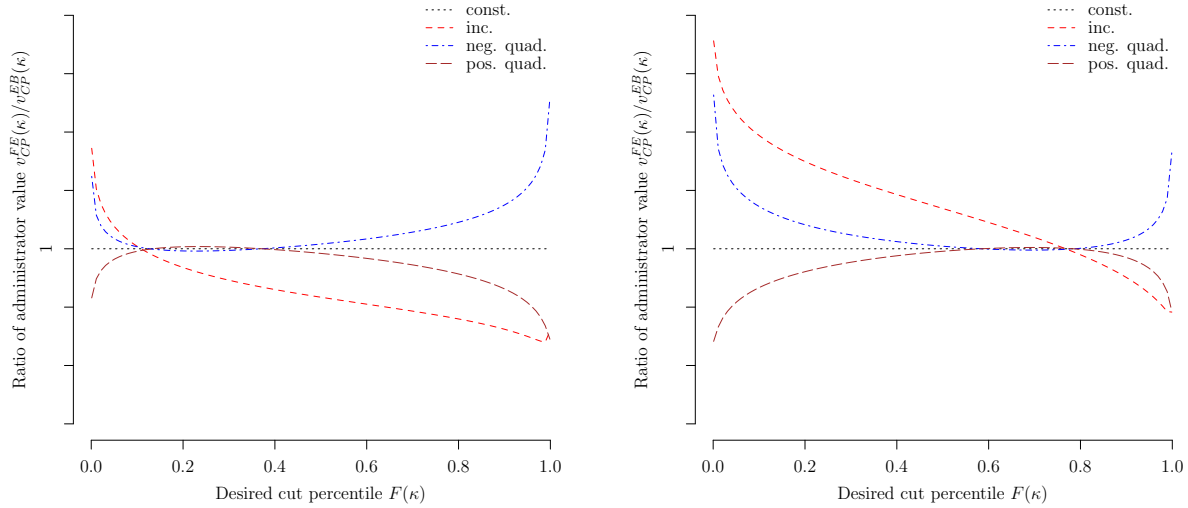
B.5   Proposition 1

This section proves that fixed effects and empirical Bayes return the same value when the administrator's problem is symmetric.

---

[3]This is illustrated in Figure B.1, which shows that $\hat{\gamma}_\sigma$ increases as κ_σ decreases from -1 to -2. Indeed, as can be seen by inspecting G in Lemma 2, $\hat{\gamma}_\sigma \to (1-\gamma_c)^2$ as $\kappa_\sigma \to -\infty$.

Figure B.2: Difference between administrator's value under fixed effects and empirical Bayes, by class size scenario and desired cut point and weight on Type I error, α

(a) $v_{CP}^{FE}(\kappa)/v_{CP}^{EB}(\kappa)$ when $\alpha = 1/4$ (b) $v_{CP}^{FE}(\kappa)/v_{CP}^{EB}(\kappa)$ when $\alpha = 2/3$

Definition 1. *The administrator's problem is symmetric if $\alpha = 1/2$, $n(\theta)$ is symmetric around the population mean of teacher quality, and the administrator's desired cutoff is $\kappa = 0$.*

Proposition 1. *The administrator receives the same value from both estimators when the problem is symmetric.*

Proof. Because $n(\theta)$ is symmetric about $\theta = 0$ and $\theta_i \sim F = \mathrm{N}\left(0, \sigma_\theta^2\right)$, the distribution of θ is symmetric around its population mean of 0. The optimal c^{*EB} solves

$$\int_0^\infty \frac{1}{\lambda(n(\theta))\sigma_{\bar{\epsilon}}(n(\theta))} \phi\left(\frac{c^{*EB}/\lambda(n(\theta)) - \theta}{\sigma_{\bar{\epsilon}}(n(\theta))}\right) \frac{\phi(\theta/\sigma_\theta)}{\sigma_\theta \cdot 1/2} d\theta = \int_{-\infty}^0 \frac{1}{\lambda(n(\theta))\sigma_{\bar{\epsilon}}(n(\theta))} \phi\left(\frac{c^{*EB}/\lambda(n(\theta)) - \theta}{\sigma_{\bar{\epsilon}}(n(\theta))}\right) \frac{\phi(\theta/\sigma_\theta)}{\sigma_\theta \cdot 1/2} d\theta.$$

At $c^{*EB} = 0$, the expression becomes

$$\int_0^\infty \frac{1}{\lambda(n(\theta))\sigma_{\bar{\epsilon}}(n(\theta))} \phi\left(\frac{-\theta}{\sigma_{\bar{\epsilon}}(n(\theta))}\right) \frac{\phi(\theta/\sigma_\theta)}{\sigma_\theta \cdot 1/2} d\theta = \int_{-\infty}^0 \frac{1}{\lambda(n(\theta))\sigma_{\bar{\epsilon}}(n(\theta))} \phi\left(\frac{-\theta}{\sigma_{\bar{\epsilon}}(n(\theta))}\right) \frac{\phi(\theta/\sigma_\theta)}{\sigma_\theta \cdot 1/2} d\theta,$$

which holds because of the symmetry of $\phi(\cdot)$, $n(\cdot)$, and $\lambda(\cdot)$ (through its dependence on n, which is symmetric). Therefore, $c^{*EB} = 0$ solves the administrator's problem when empirical Bayes is used. Because $\lambda(n(\theta)) = 1$, $\forall \theta$ when the fixed effects estimator is used, $c^{*FE} = 0$ must also solve the administrator's problem when fixed effects is used, meaning

the administrator's objective is equivalent under both estimators. $\square$

B.6 Functional Form of $\lambda(\cdot)$

The functional form for $\lambda(\theta)$ was chosen to simplify exposition; another way to provide intuition for result in Proposition 2 is to consider the effect of an infinitesimal "bump" β_b to the shrinkage weight, via

$$\lambda(\theta) = \begin{cases} \overline{\lambda} & \theta \neq \theta_b \\ \overline{\lambda} + \beta_b & \theta = \theta_b. \end{cases}$$

If $\theta_b < \kappa$ then

$$\left. \frac{\partial v_{CP}^{EB}}{\partial \beta_b} \right|_{\beta_b=0} = (1-\alpha) \frac{-c^{*EB}}{(\overline{\lambda})^2 \overline{\sigma_{\bar{\epsilon}}}} \phi \left(\frac{\theta_b - \frac{c^{*EB}}{\overline{\lambda}}}{\overline{\sigma_{\bar{\epsilon}}}} \right) \frac{\phi(\theta_b/\sigma_\theta)}{\sigma_\theta \Phi(\kappa/\sigma_\theta)} > 0,$$

i.e., lowering the weight $\lambda(\cdot)$ for lower values of θ, as would occur with a negative-quadratic $\lambda(\cdot)$, would lower the administrator's value and increasing $\lambda(\cdot)$ for lower values of θ, as would occur with a positive-quadratic $\lambda(\cdot)$, would increase the administrator's value.

B.7 Mechanical Heteroskedasticity

The theoretical results in Section 3 were obtained under the assumption that all teachers faced the same distribution of mean measurement error, i.e., $\sigma_{\bar{\epsilon}}(\theta) = \overline{\sigma}_{\bar{\epsilon}}$. There are two main sources of heteroskedasticity in the errors on teacher quality measures. The first could be thought of as "essential heteroskedasticity", where the test score errors for different teachers had different variances, meaning the standard deviation of the measurement error for a single test score for teacher i could be written as $\sigma_{\epsilon,i}$. Such heteroskedasticity could be present even in the case of constant class sizes. The second is what I will refer to as "mechanical heteroskedasticity", which is due to nonconstant class sizes (which would emerge, e.g., in the scenario of Proposition 2).

Essential heteroskedasticity Consider the environment of Boyd et al. (2013), which allows for heteroskedasticity in terms of two observed variables: the grade level and the student level (in terms of prior achievement). In principle, either or both forms of heteroskedasticity could be accommodated by my framework by conducting the comparison at the appropriate level, e.g., assigning a bonus to the κ-quantile quality teacher among 4th-grade teachers with students who had low prior achievement. More generally, if the error variance depends on known variables then conducting comparisons given those variables will result in

homoskedastic error terms.

Mechanical heteroskedasticity The other main form of heteroskedasticity is the mechanical heteroskedasticity that would emerge in the case of nonconstant class sizes. That is, the mean score for teacher i, e.g., in the cutoff model, $\overline{y}_i = \theta_j + \frac{\sum_j \epsilon_{ji}}{n_i}$, is naturally heteroskedastic when $n(\theta)$ is nonconstant, where $\sigma_{\overline{\epsilon}}(\theta) = \sigma_\epsilon / \sqrt{n(\theta)}$.

I chose the simple and tractable homoskedastic environment to most clearly illustrate the theoretical results. This mechanical heteroskedasticity is accounted for by the indirect inference algorithm described in Appendix D.2, and the numerical model solutions and quantitative results allow for it. Nevertheless, it is worthwhile to show how the theoretical results from the cutoff model could still hold in the presence of mechanical heteroskedasticity. Proposition 1 follows trivially, since the constant $n(\theta)$ precludes there being any mechanical heteroskedasticity, but Proposition 2 is less obvious.

I show here why Proposition 2 would still likely hold in the presence of mechanical heteroskedasticity. Consider as a starting point that $\left.\frac{\partial v_{CP}^{EB}}{\partial \beta_-}\right|_{\beta_-=0} < 0$ in the homoskedastic case. With heteroskedastic errors, we have

$$
\left.\frac{\partial v_{CP}^{EB}}{\partial \beta_-}\right|_{\beta_-=0} = (1-\alpha) \int_{-\infty}^{\kappa} \frac{-c^{*EB}\theta}{(\delta_-)^2 \sigma_{\overline{\epsilon}}(\theta)} \phi\left(\frac{\theta - \frac{c^{*EB}}{\delta_-}}{\sigma_{\overline{\epsilon}}(\theta)}\right) \frac{\phi(\theta/\sigma_\theta)}{\sigma_\theta \Phi(\kappa/\sigma_\theta)} d\theta
$$
$$
+ \alpha \int_{\kappa}^{0} \frac{c^{*EB}\theta}{(\delta_-)^2 \sigma_{\overline{\epsilon}}(\theta)} \phi\left(\frac{\theta - \frac{c^{*EB}}{\delta_-}}{\sigma_{\overline{\epsilon}}(\theta)}\right) \frac{\phi(\theta/\sigma_\theta)}{\sigma_\theta(1 - \Phi(\kappa/\sigma_\theta))} d\theta.
$$

If in the homoskedastic error case we have $\left.\frac{\partial v_{CP}^{EB}}{\partial \beta_-}\right|_{\beta_-=0} < 0$, then by Proposition 2 we have

$$
(1-\alpha) \int_{-\infty}^{\kappa} h(\theta) \frac{\phi(\theta/\sigma_\theta)}{\sigma_\theta \Phi(\kappa/\sigma_\theta)} d\theta > \alpha \int_{\kappa}^{0} h(\theta) \frac{\phi(\theta/\sigma_\theta)}{\sigma_\theta(1 - \Phi(\kappa/\sigma_\theta))} d\theta,
$$

where $h(\theta) = \frac{c^{*EB}\theta}{(\delta_-)^2 \sigma_{\overline{\epsilon}}} \phi\left(\frac{\theta - \frac{c^{*EB}}{\delta_-}}{\overline{\sigma}_{\overline{\epsilon}}}\right) \geq 0$. We can then examine the effect of heteroskedasticity by seeing how $h(\theta)$ changes when the average measurement error variance can now depend on θ, via $n(\theta)$, i.e., when $h(\theta) = \frac{c^{*EB}\theta}{(\delta_-)^2 \sigma_{\overline{\epsilon}}(\theta)} \phi\left(\frac{\theta - \frac{c^{*EB}}{\delta_-}}{\sigma_{\overline{\epsilon}}(\theta)}\right)$. To operationalize this, note that the relationship between θ and $\sigma_{\overline{\epsilon}}(\theta)$ will have the opposite sign as that between θ and $\lambda(\theta)$ (which has the same sign as the relationship between θ and $n(\theta)$). Thus, I consider

$$
\sigma_{\overline{\epsilon}}(\theta) = \max\{\underline{\sigma}, \delta_\sigma - \beta_\sigma \theta\}, \forall \theta \leq 0,
$$

where $\underline{\sigma} > 0$. Note that, as in the parameterization of $\lambda(\cdot)$ for Proposition 2, although it has

12

Figure B.3: How $h(\theta)$ is affected when β_σ is increased from zero
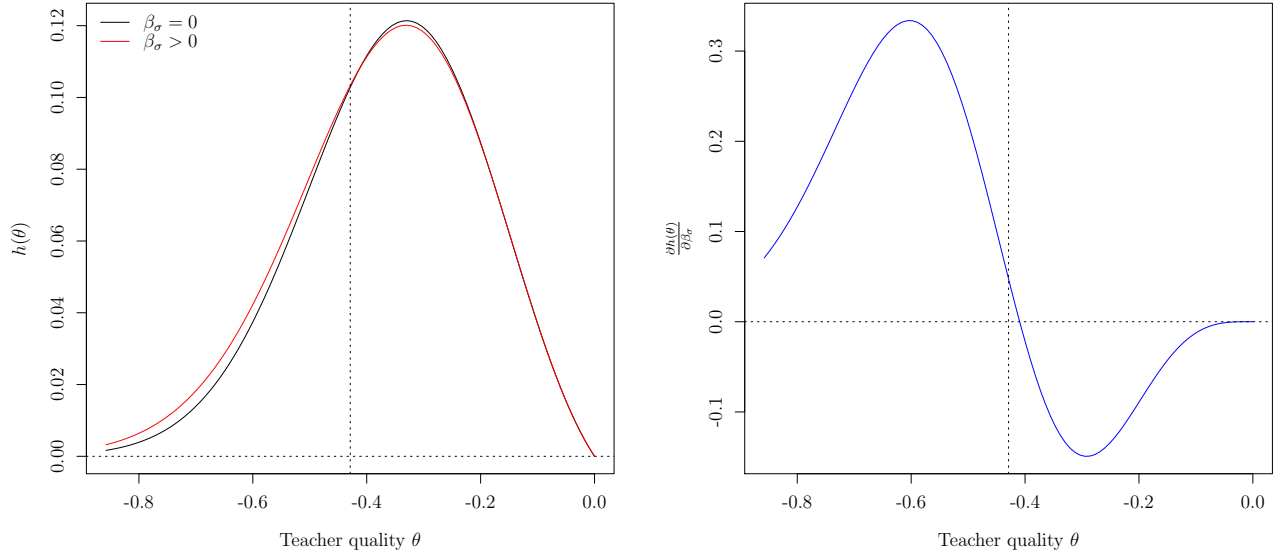
been included for completeness, the max operator is obviated by evaluating the derivative at $\beta_\sigma = 0$.

Differentiating h with respect to β_σ, at $\beta_- = \beta_\sigma = 0$ (and, for simplicity, setting $\delta_- = 1$), we have

$$
\left.\frac{\partial h}{\partial \beta_\sigma}\right|_{\beta_-=\beta_\sigma=0} = \underbrace{\frac{c^{EB}\theta^2}{\delta_\sigma}}_{<0}\left(1 - \frac{(\theta - c^{EB})^2}{\delta_\sigma^2}\right)\underbrace{\frac{1}{\delta_\sigma}\phi\left(\frac{c^{EB} - \theta}{\delta_\sigma}\right)}_{>0}.
$$

The middle term is positive for values of θ that are very close to c^{EB}, resulting in a reduction in $h(\theta)$. For $\theta < c^{EB} - \delta_\sigma$, the middle term will be negative, resulting in an increase in $h(\theta)$.[4] Given that the starting point where $h(\theta)$ was such that $\left.\frac{\partial v_{CP}^{EB}}{\partial \beta_-}\right|_{\beta_-=0} < 0$, the increases in $h(\theta)$ for larger, negative values of θ and decreases in $h(\theta)$ for θ values close to c^{EB} mean the inequality will likely continue to hold, i.e., $\left.\frac{\partial v_{CP}^{EB}}{\partial \beta_-}\right|_{\beta_-=0} < 0$ even with $\beta_\sigma \neq 0$. Intuitively, $h(\theta)$ "stretches" to the left, as shown in Figure B.3. The left panel shows how $h(\theta)$ changes when we increase β_σ from zero, and the right panel plots the derivative of $h(\theta)$ with respect to β_σ, evaluated at $\beta_\sigma = \beta_- = 0$. In each panel, the vertical dotted line denotes κ, which has been set to $-2\sigma_\theta$. Thus, the presence of mechanical heteroskedasticity would, if anything, likely widen the difference in value obtained under the two estimators.

---

[4]The middle term would also be negative for $\theta > c^{EB} + \delta_\sigma$, which does not substantially affect the argument as we are only considering $\theta \leq 0$.

C Extensions to Hidden Type Model HT-0

C.1 Model HT-1

Now allow $T > 2$ and let output depend on teacher experience $x_{i(j,t),t}$ according to $q_{jit} = \beta_0 + \theta_{i(j,t)} + e(x_{i(j,t),t})$, where $e(x_{it})$ represents output, net of β_0 and teacher quality, for a teacher with $t - 1$ periods of prior experience.

The optimal hiring policy ψ_h is unchanged. Consider the retention decision for teachers in period $t = T$, for teachers with the same experience, $x_{it} = x_t$. Such a policy need not only apply to teachers' first years of experience; Wiswall (2013) shows that teacher quality also changes after the first few years of experience. Let $\hat{q}_{H_{it}}$ be the sample mean of teacher i's output signals realized before period t. The retention decision ψ_r still has a reservation value property, which now depends on the mean of each teacher's entire history of signals, $\hat{q}_{H_{it}}$, where the threshold now depends on the period, i.e., $\underline{q}_t = \mu - \left(\frac{\chi + e(x_t)}{\rho_t} \right)$, where $\rho_t = \sigma_\theta^2 / (\sigma_\theta^2 + \frac{\sigma_\epsilon^2}{n|H_t|})$. The reservation signal $\underline{q}_t$ is decreasing in x_t if there are productivity gains to experience and increasing in ρ_t, due to the higher precision about teachers' true quality. Note that solution to this problem would be the same as that from HT-0, setting the replacement cost (in HT-0) to $\chi_t \equiv \chi + e(x_t)$ and using the relevant ρ_t, and that considering instead periods $t < T$ would change the desired threshold quality, which could be modeled by suitably adjusting the replacement cost χ from the static model HT-0. Therefore, this sequence of per-period reservation signals can then be mapped to the cutoff-based model via a sequence of cutoff-based problems, one for each period of experience, as was done for Model HT-0. Also, note that a similar transformation to the one above could be performed to adapt Model HT-2 (see Section C.2) to also allow for an effect of experience on output.

C.2 Model HT-2

This model augments HT-0 to allow class size to depend on teacher quality, i.e., $n_i = n(\theta)$. As in HT-0, consider the administrator's problem in the second period. As in the cutoff model, the administrator must now integrate over the distribution of class sizes when choosing their reservation signal. I first derive expected teacher quality, given the quality signal $\hat{q} = \hat{\theta}^{EB} = \lambda(n(\theta))\hat{\theta}^{FE}$ is greater than cutoff rule $\underline{q}$, where $g(\cdot)$ denotes the density function

of the argument(s) it takes:

$$\mathrm{E}\left[\theta|\hat{\theta}^{EB} \geq \underline{q}\right] = \int_{-\infty}^{\infty} \theta g(\theta|\hat{\theta}^{EB} \geq \underline{q})d\theta = \int_{-\infty}^{\infty}\int_{\underline{q}}^{\infty} \theta g(\theta, \hat{\theta}^{EB}|\hat{\theta}^{EB} \geq \underline{q})d\hat{\theta}d\theta$$

$$= \int_{-\infty}^{\infty}\int_{\underline{q}}^{\infty} \theta \frac{g(\theta, \hat{\theta}^{EB})}{\Pr\{\hat{\theta}^{EB} \geq \underline{q}\}}d\hat{\theta}^{EB}d\theta = \frac{1}{\Pr\{\hat{\theta}^{EB} \geq \underline{q}\}}\int_{-\infty}^{\infty}\int_{\underline{q}}^{\infty} \theta g(\theta, \hat{\theta}^{EB})d\hat{\theta}^{EB}d\theta$$

$$= \frac{1}{\Pr\{\hat{\theta}^{EB} \geq \underline{q}\}}\int_{-\infty}^{\infty}\int_{\underline{q}}^{\infty} \theta g(\hat{\theta}^{EB}|\theta)d\hat{\theta}^{EB}f(\theta)d\theta$$

$$= \frac{1}{\Pr\{\hat{\theta}^{EB} \geq \underline{q}\}}\int_{-\infty}^{\infty}\int_{\underline{q}/\lambda(n(\theta))-\theta}^{\infty} \theta g(\bar{\epsilon}|\theta)d\bar{\epsilon}f(\theta)d\theta$$

$$= \frac{1}{\Pr\{\hat{\theta}^{EB} \geq \underline{q}\}}\int_{-\infty}^{\infty} \theta \left[1 - \Phi\left[\frac{\underline{q}/\lambda(n(\theta)) - \theta}{\sigma_{\bar{\epsilon}}(\theta)}\right]\right]f(\theta)d\theta,$$

where

$$\Pr\{\hat{\theta}^{EB} \geq \underline{q}\} = \int_{-\infty}^{\infty} \Pr\{\hat{\theta}^{EB} \geq \underline{q}|\theta\}f(\theta)d\theta = \int_{-\infty}^{\infty} \Pr\left\{\bar{\epsilon} \geq \frac{\underline{q}}{\lambda(n(\theta))} - \theta\right\}f(\theta)d\theta$$

$$= \int_{-\infty}^{\infty} 1 - \Phi\left[\frac{\frac{\underline{q}}{\lambda(n(\theta))} - \theta}{\sigma_{\bar{\epsilon}}(\theta)}\right]f(\theta)d\theta.$$

We then have

$$\Pr\{\hat{\theta}^{EB} < \underline{q}\} = 1 - \Pr\{\hat{\theta}^{EB} \geq \underline{q}\} = \int_{-\infty}^{\infty} \Phi\left[\frac{\frac{\underline{q}}{\lambda(n(\theta))} - \theta}{\sigma_{\bar{\epsilon}}(\theta)}\right]f(\theta)d\theta.$$

As before, these quantities when using fixed effects are obtained by setting $\lambda(\cdot) = 1$:

$$\mathrm{E}\left[\theta|\hat{\theta}^{FE} \geq \underline{q}\right] = \frac{1}{\Pr\{\hat{\theta}^{FE} \geq \underline{q}\}}\int_{-\infty}^{\infty}\int_{\underline{q}-\theta}^{\infty} \theta \left[1 - \Phi\left[\frac{\underline{q} - \theta}{\sigma_{\bar{\epsilon}}(\theta)}\right]\right]f(\theta)d\theta$$

$$\Pr\{\hat{\theta}^{FE} \geq \underline{q}\} = \int_{-\infty}^{\infty} 1 - \Phi\left[\frac{\underline{q} - \theta}{\sigma_{\bar{\epsilon}}(\theta)}\right]f(\theta)d\theta$$

$$\Pr\{\hat{\theta}^{FE} < \underline{q}\} = \int_{-\infty}^{\infty} \Phi\left[\frac{\underline{q} - \theta}{\sigma_{\bar{\epsilon}}(\theta)}\right]f(\theta)d\theta.$$

The administrator's value from using the empirical Bayes estimator, using the result that

she will use a cutoff signal policy, is then

$$v_{HT2}^{EB}(\chi) = \max_{\underline{q}} \left\{ -\Pr\{\hat{\theta}^{EB} < \underline{q}\}\chi + \Pr\{\hat{\theta}^{EB} \geq \underline{q}\} \, \mathrm{E}\left[\theta | \hat{\theta}^{EB} \geq \underline{q}\right] \right\}$$

$$= \max_{\underline{q}} \left\{ -\Pr\{\hat{\theta}^{EB} < \underline{q}\}\chi + \Pr\{\hat{\theta}^{EB} \geq \underline{q}\} \frac{1}{\Pr\{\hat{\theta}^{EB} \geq \underline{q}\}} \int_{-\infty}^{\infty} \int_{\underline{q}/\lambda(n(\theta))-\theta}^{\infty} \theta \left[1 - \Phi\left[\frac{\underline{q}/\lambda(n(\theta)) - \theta}{\sigma_{\bar{\epsilon}}(\theta)} \right] \right] f(\theta)d\theta \right\}$$

$$= \max_{\underline{q}} \left\{ -\Pr\{\hat{\theta}^{EB} < \underline{q}\}\chi + \int_{-\infty}^{\infty} \theta \left[1 - \Phi\left[\frac{\underline{q}/\lambda(n(\theta)) - \theta}{\sigma_{\bar{\epsilon}}(\theta)} \right] \right] f(\theta)d\theta \right\}$$

$$= \max_{\underline{q}} \left\{ \int_{-\infty}^{\infty} (-\chi)\Phi\left[\frac{\frac{\underline{q}}{\lambda(n(\theta))} - \theta}{\sigma_{\bar{\epsilon}}(\theta)} \right] f(\theta)d\theta + \int_{-\infty}^{\infty} \theta \left[1 - \Phi\left[\frac{\underline{q}/\lambda(n(\theta)) - \theta}{\sigma_{\bar{\epsilon}}(\theta)} \right] \right] f(\theta)d\theta \right\}, \tag{8}$$

and the administrator's value from using fixed effects is

$$v_{HT2}^{FE}(\chi) = \max_{\underline{q}} \left\{ \int_{-\infty}^{\infty} (-\chi)\Phi\left[\frac{\underline{q} - \theta}{\sigma_{\bar{\epsilon}}(\theta)} \right] f(\theta)d\theta + \int_{-\infty}^{\infty} \theta \left[1 - \Phi\left[\frac{\underline{q} - \theta}{\sigma_{\bar{\epsilon}}(\theta)} \right] \right] f(\theta)d\theta \right\}. \tag{9}$$

Intuitively, under either estimator the administrator will either replace a teacher with some probability, which in expectation reduces her objective by an expected value of χ, or retains the teacher with the complementary probability, in which case her objective is increased by that teacher's quality θ.

Because $n(\theta)$ is not constant (as it was in HT-0), the reliability of signals varies by teacher and the analytical characterization of the administrator's reservation signal from Model HT-0 no longer obtains. The estimator-specific reservation signals, $\underline{q}^{*EB}$ and $\underline{q}^{*FE}$, are respectively obtained by numerically solving (8) and (9).

The ranking of the administrator's utility from HT-2, by class size scenario $n(\theta)$, is the same as her ranking under the cutoff-based model.

Proposition 2. *In Model HT2, the administrator's preferred estimator depends on the relationship between teacher quality and class size. In particular, when the relationship between teacher quality and class size is negative-(positive-)quadratic, the administrator will prefer the fixed effects (empirical Bayes) estimator.*

Proof. I will focus on teachers with $\theta < 0$, since below-average teachers will be most affected by the reservation policy, which will generally be negative.[5] As in the cutoff model, parameterize the empirical Bayes weights using

$$\lambda(\theta) = \begin{cases} \max\{\underline{\lambda}, \delta_- + \beta_-\theta\} & \text{if } \theta < 0 \\ \max\{\underline{\lambda}, \delta_+ + \beta_+\theta\} & \text{if } \theta \geq 0, \end{cases}$$

---

[5]Figure C.2 shows that estimator rankings are similar for the negative-quadratic and increasing scenarios, and between the positive-quadratic and decreasing scenarios.

where $\underline{\lambda} > 0$, and assume homoskedastic errors, i.e., $\sigma_{\bar{\epsilon}}(\theta) = \overline{\sigma_{\bar{\epsilon}}}$ for all θ.[6] Differentiating the administrator's value with respect to β_- and evaluating the derivative at $\beta_- = 0$ (as will be made clear below, this is convenient because the environment in which the derivative is evaluated is consistent with the one studied in HT-0, the constant class size scenario studied by Proposition 4), we obtain

$$\frac{\partial v_{HT2}^{EB}}{\partial \beta_-}\bigg|_{\beta_-=0} = \int_{-\infty}^{0} \frac{\underline{q}^{*EB}}{\delta_-^2}\theta\,[\chi+\theta]\frac{1}{\overline{\sigma_{\bar{\epsilon}}}}\phi\left(\frac{\frac{\underline{q}^{*EB}}{\delta_-}-\theta}{\overline{\sigma_{\bar{\epsilon}}}}\right)\frac{1}{\sigma_\theta}\phi\left(\frac{\theta}{\sigma_\theta}\right)d\theta, \qquad (10)$$

because $\frac{\partial v_{HT2}^{EB}}{\partial \underline{q}^{*EB}} \times \frac{\partial \underline{q}^{*EB}}{\partial \beta_-} = 0$ due to the Envelope Theorem. As in the proof of Proposition 2, combine the normal densities into one density, $f_P(\theta)/S_P$, where S_P is a positive constant (see footnote 1 for details). The expression (10) becomes

$$\frac{\partial v_{HT2}^{EB}}{\partial \beta_-}\bigg|_{\beta_-=0} = \frac{1}{S_P}\int_{-\infty}^{0}\underbrace{\frac{\underline{q}^{*EB}}{\delta_-^2}\theta\,[\chi+\theta]}_{m(\theta)}f_P(\theta)d\theta,$$

where $f_P(\theta) = \frac{1}{\sigma_P}\phi\left(\frac{\theta-\mu_P}{\sigma_P}\right)$, $\mu_P = \frac{\underline{q}^{*EB}}{\delta_-}\frac{\sigma_\theta^2}{\overline{\sigma_{\bar{\epsilon}}}^2+\sigma_\theta^2} = \underline{q}^{*EB}$, $\sigma_P^2 = \sigma_\theta^2\gamma_\sigma$, and $\gamma_\sigma = \frac{\overline{\sigma_{\bar{\epsilon}}}^2}{\overline{\sigma_{\bar{\epsilon}}}^2+\sigma_\theta^2}$. The function $m(\theta)$ is strictly concave (in particular, negative quadratic) in θ and has zeros at $\theta = -\chi$ and $\theta = 0$. Therefore, we have

$$\frac{\partial v_{HT2}^{EB}}{\partial \beta_-}\bigg|_{\beta_-=0} < 0 \Leftrightarrow \int_{-\infty}^{-\chi}-m(\theta)f_P(\theta)d\theta > \int_{-\chi}^{0}m(\theta)f_P(\theta)d\theta.$$

Observe that $-m(-\chi-a) > m(-\chi+a)$ for $0 < a < \chi$ and $-m(-\chi-a) \geq 0$ for $a > 0$. Therefore, if f_P were centered around a value no greater than $-\chi$, i.e., $\mu_P \leq -\chi$, then by the symmetry of the normal distribution we would have $\int_{-\infty}^{-\chi}-m(\theta)f_P(\theta)d\theta > \int_{-\chi}^{0}m(\theta)f_P(\theta)d\theta$ because $-m(\theta)f_P(-\chi-\theta) > m(\theta)f_P(-\chi+\theta)$ for $\theta > 0$ and, thus, $\frac{\partial v_{HT2}^{EB}}{\partial \beta_-}\bigg|_{\beta_-=0} < 0$. Note that the measurement errors have been assumed to be homoskedastic and that $\lambda(\cdot)$ is constant because we are evaluating the derivative at $\beta_- = 0$, which means the optimal fixed effects policy from Model HT-0 (see (8)) would obtain when using fixed effects: $\underline{q}^{*FE} = \frac{-\chi}{\rho}$, where $\rho = \frac{\sigma_\theta^2}{\sigma_\theta^2+\overline{\sigma_{\bar{\epsilon}}}^2} < 1$. Then, by Proposition 4 we have $\underline{q}^{*EB} = \rho\underline{q}^{*FE} = -\chi$, which satisfies $\mu_P \leq -\chi$, and the desired inequality is obtained.[7] Figure C.1 illustrates this scenario, where the solid curve is $|m(\theta)|$, which is $-m(\theta)$ for $\theta \leq -\chi$ and $m(\theta)$ for $\theta \in [-\chi, 0]$, and the

---

[6]The numerical solutions presented later in this section and the quantitative results in Section 5.3 allow for mechanical heteroskedasticity in $\sigma_{\bar{\epsilon}}(\theta)$.

[7]Although not as straightforward to solve for explicitly, $\underline{q}^{*EB}$ is very close to $-\chi$ across a wide variety of $n(\theta)$, even when allowing for both mechanical heteroskedasticity in $\sigma_{\bar{\epsilon}}(\theta)$ and the effect of class size in $\lambda(\theta)$.

Figure C.1: $|m(\theta)|$ and $f_P(\theta)$
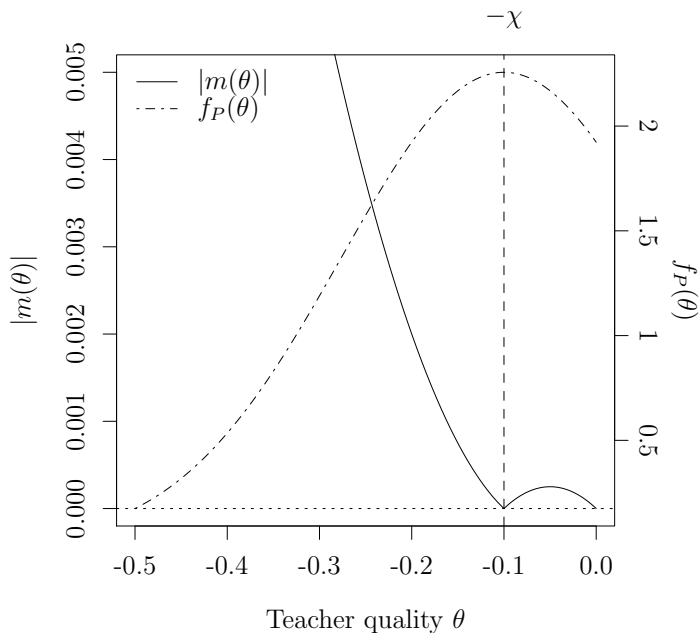
dot-dashed curve is f_P, drawn with $\mu_P = -\chi$. □

To illustrate this theoretical result, Figure C.2 plots the ratio of value from the FE over EB estimators for a wide range of replacement costs χ,[8] across different class size scenarios: constant, increasing, decreasing, negative quadratic, and positive quadratic. The constant class size scenario (dotted black line) represents a special case of HT-2 where $n(\theta) = n$, which is simply model HT-0. Unsurprisingly, then, we obtain the same value for all replacement costs χ. Under the negative quadratic scenario (dot-short-dashed blue curve) the administrator would obtain higher value from using fixed effects for every χ. This is the same result as was obtained for a wide range of parameterizations of the cutoff-based model. Also, as in the cutoff-based model, the estimator ranking is reversed under the positive-quadratic class size scenario (long-dashed brown curve); i.e., she would prefer to use empirical Bayes instead of fixed effects. The rankings of increasing (short-dashed red) and negative-quadratic (dot-short-dashed blue) scenarios are similar. As argued above, most of the difference in value comes from the part of the teacher quality distribution at highest risk

---

[8]Wiswall (2013) reports that teachers with 30 years of experience have value-added that is one standard deviation higher than new teachers and 0.75 standard deviations higher than teachers with five years of experience; this implies a 0.25 sd difference acquired in the first five years of experience. Therefore, I set $\chi = 0.25\sigma_\theta = 0.054$ for the baseline quantitative results and for this figure use a range for the replacement cost running from zero to 0.30, over five times this value.

Figure C.2: Ratio of values of FE over EB, across class size scenarios
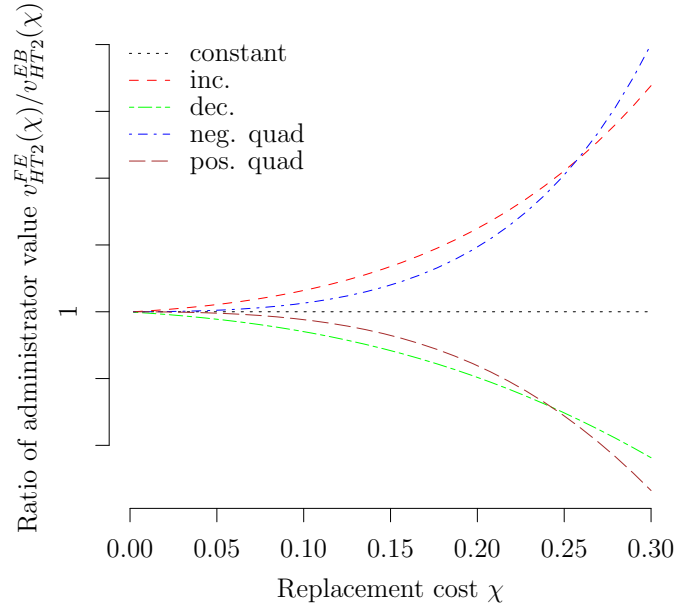
of being replaced—those with below-average qualities. By the same reasoning, the rankings of decreasing (dot-long-dashed green) and positive-quadratic (long-dashed brown) scenarios are also similar.[9]

As with model HT-1, an environment with multiple periods could be modeled by suitably adjusting the desired threshold quality. For example, adding more periods could be accommodated by decreasing the replacement cost, as the administrator would have a relatively higher gain from replacing when there are more periods of output. Because they range from a cost of zero to several times the estimated difference in value-added between a teacher with five years experience and no experience, Figure C.2 then likely also characterizes estimator rankings for multi-period environments.

The takeaway from this section is that (i) the administrator's preferred estimator depends on the class size scenario $n(\theta)$, (ii) though the difference in values from using either estimator depends on other model parameters (T, χ), the preferred estimator does not, and (iii) the administrator would prefer the same estimator in HT-2 as she would in the cutoff model.

---

[9]This figure takes into account the mechanical heteroskedasticity caused by the variation in $n(\theta)$.

D   Details for Quantitative Exercises

D.1   Calibrated Error Variances

I calibrate σ_θ^2 and σ_ϵ^2 from Table B-2 of Schochet and Chiang (2012) normalizing the total variance to one. To most closely match a policy where an administrator would like to rank teachers across a school district, I calibrate $\sigma_\theta^2 = 0.046$ by summing the average of school- and teacher-level variances in random effects. To most closely approximate an environment where both student and aggregate-level shocks may affect student test scores, I calibrate $\sigma_\epsilon^2 = 0.953$ by summing the average of class- and student-level variances in random effects. Note that, due to the much greater student-level error variance, the approximate sizes of σ_θ^2 and σ_ϵ^2 are approximately the same if school-level variances are excluded from σ_θ^2 or class-level variances are excluded from σ_ϵ^2, lending robustness to the quantitative findings.

D.2   Heteroskedasticity Correction for Relationship Between Class Size and Teacher Quality

The advantage of the indirect inference approach is that it can be implemented using a vector of auxiliary moments which do not necessarily correspond to structural econometric parameters. This is useful in the current context, where the micro-data to directly correct for heteroskedasticity are not available.[10]

Indirect Inference Algorithm   The following is done separately for Reading and Math.

0. Estimate the relationship between class size (n_i) and teacher i's estimated quality in the subject ($\hat{\theta}_i$) by running the regression $n_i = \beta_0^{data} + \beta_1^{data}\hat{\theta}_i + \beta_2^{data}(\hat{\theta}_i)^2 + e_i$. The regression coefficients $\left(\hat{\beta}_0^{data}, \hat{\beta}_1^{data}, \hat{\beta}_2^{data}\right)$ and residual standard error $\hat{\sigma}_e^{data}$ form the first set auxiliary parameters to fit. Compute the 25th, 50th, and 75th percentiles of the empirical distribution of class sizes, $(n_{p25}^{data}, n_{p50}^{data}, n_{p75}^{data})$. These are the remaining auxiliary parameters. The target vector of auxiliary parameters is then $\left(\hat{\beta}_0^{data}, \hat{\beta}_1^{data}, \hat{\beta}_2^{data}, \hat{\sigma}_e^{\,data}, n_{p25}^{data}, n_{p50}^{data}, n_{p75}^{data}\right)$.

1. Given σ_θ^2, simulate teacher quality θ_i^{sim} once for each teacher in the sample. (Recall the population mean has been normalized to 0).

2. Simulate the random component of class sizes $n_{i,i.i.d}^{sim}$, which is distributed normal with mean zero and standard deviation $\sigma_{n_{sim}}$. As described below, this algorithm chooses

---

[10]If micro-data had been available, then one could in principle use an approach like the one in Lockwood and McCaffrey (2014) to account for the nonlinearities produced by heteroskedastic errors.

the parameter $\sigma_{n_{sim}}$. Note these are independent from teacher quality to get an idea of the role heteroskedasticity plays.

3. Assign *incremental class sizes* according to $n^{inc}(\theta_i^{sim}) = a_0 + a_1 \theta_i^{sim} + a_2 (\theta_i^{sim})^2$. As described below, this algorithm chooses the parameters (a_0, a_1, a_2). The final simulated class size for teacher i is then $n_i^{sim} = \text{round}\{n_{i,i.i.d}^{sim} + n^{inc}(\theta_i^{sim})\}$, i.e., class sizes are integer-valued.

4. Given σ_ϵ^2 and n_i^{sim} simulate an average shock for each teacher, $\bar{\epsilon}_i^{sim}$; form simulated estimated teacher quality according to $\hat{\theta}_i^{sim} = \theta_i^{sim} + \bar{\epsilon}_i^{sim}$.

5. Regress $n_i^{sim} = \beta_0^{sim} + \beta_1^{sim} \hat{\theta}_i^{sim} + \beta_2^{sim} (\hat{\theta}_i^{sim})^2$, estimating the auxiliary coefficients $(\hat{\beta}_0^{sim}, \hat{\beta}_1^{sim}, \hat{\beta}_2^{sim})$ and auxiliary residual standard error $\hat{\sigma}_e^{sim}$. Compute the 25th, 50th, and 75th percentiles of the simulated distribution of class sizes, $(n_{p25}^{sim}, n_{p50}^{sim}, n_{p75}^{sim})$. The simulated vector of auxiliary parameters is then $\left(\hat{\beta}_0^{sim}, \hat{\beta}_1^{sim}, \hat{\beta}_2^{sim}, \hat{\sigma}_e^{sim}, n_{p25}^{sim}, n_{p50}^{sim}, n_{p75}^{sim} \right)$.

6. Compute the Euclidean distance between target auxiliary parameters and simulated auxiliary parameters (e.g., $\hat{\beta}_0^{data}$ and $\hat{\beta}_0^{sim}$, respectively) as a function of the parameters governing class size, $d(a_0, a_1, a_2, \sigma_{n_{sim}})$.

Repeat steps 1-6 for the vector $(a_0, a_1, a_2, \sigma_{n_{sim}})$, until the distance between data and simulated auxiliary moments is minimized.

D.3 Details for Quantitative Illustration for Hidden Action Model

Output in the hidden action model depends on several parameters, including the variance of measurement error on output, σ_η^2. I adjust the error variance in several steps, using Reading test scores as the measure:

1. Simulate teacher quality, class sizes, and measurement errors using the parameters from Section 5.1, for 30,000 teachers. Each simulated teacher then has a simulated quality θ_i^s and a simulated fixed-effect estimate $\hat{\theta}_i^{s,FE}$.

2. Use the empirical Bayes weights $\lambda(\cdot)$ to generate simulated EB measures of teacher quality according to $\hat{\theta}_i^{s,EB} = \lambda(n(\theta_i^s))\hat{\theta}_i^{s,FE}$.

3. Standardize θ_i^s, $\hat{\theta}_i^{s,FE}$, and $\hat{\theta}_i^{s,EB}$ to have variances of 1, to make the residual variances comparable.

4. Finally, I estimate the residual variance from a regression of standardized $\hat{\theta}_i^{s,FE}$ on the standardized true (simulated) quality θ_i^s and the residual variance from a regression of standardized empirical Bayes measure $\hat{\theta}_i^{s,EB}$ on standardized true (simulated) quality. The ratio of residual variances, or amount unexplained in each regression, tells us how much more (or less) the fixed effects estimator would inform the administrator about teacher quality.

The regression results, shown in Table D.1, indicate that the fixed-effects estimator explains about 3.2% more variation in teacher quality than the empirical Bayes estimator $(1 - 0.6956^2/0.7070^2 = 0.032)$. That is, the fact that the EB estimator makes it more difficult to separate high- and low-performing teachers when the class size function is negative quadratic, as it is in the data, can be modeled as increasing the measurement error variance on teacher output, σ_η^2, by this amount.

Table D.1: Regressions of simulated teacher quality on FE and EB estimates

| | *Dependent variable:* | |
| | θ^s (standardized) | |
	(1)	(2)
$\hat{\theta}^{s,FE}$ (standardized)	0.718***	
	(0.004)	
$\hat{\theta}^{s,EB}$ (standardized)		0.707***
		(0.004)
Constant	0.002	-0.001
	(0.004)	(0.004)
Observations	30,000	30,000
R^2	0.516	0.500
Residual Std. Error (df = 29998)	0.696	0.707

Note: *p<0.1; **p<0.05; ***p<0.01

References

Boyd, D., H. Lankford, S. Loeb and J. Wyckoff, "Measuring Test Measurement Error: A General Approach," *Journal of Educational and Behavioral Statistics*, 38(6):629–663, 2013.

Bromiley, P., "Products and Convolutions of Gaussian Probability Density Functions," *Tina-Vision Memo*, 3(4), 2003.

Chechile, R. A., "Properties of Reverse Hazard Functions," *Journal of Mathematical Psychology*, 55(3):203–222, 2011.

Lockwood, J. and D. McCaffrey, "Should Nonlinear Functions of Test Scores be Used as Covariates in a Regression Model?" in R. Lissitz and H. Jiang, eds., "Value-added Modeling and Growth Modeling with Particular Application to Teacher and School Effectiveness," chap. 1, pp. 1–36, Information Age, Charlotte, NC, 2014.

Schochet, P. and H. Chiang, "What Are Error Rates for Classifying Teacher and School Performance Using Value-Added Models?" *Journal of Educational and Behavioral Statistics*, 2012.

Staiger, D. and J. Rockoff, "Searching for Effective Teachers with Imperfect Information," *Journal of Economic Perspectives*, 24(3):97–117, 2010.

Wiswall, M., "The Dynamics of Teacher Quality," *Journal of Public Economics*, 100:61–78, 2013.