Journal of Economic Perspectives—Volume 30, Number 2—Spring 2016—Pages 245–248

Correspondence

To be considered for publication in the Correspondence section, letters should be relatively short—generally less than 1,000 words—and should be sent to the journal offices at jep@jepjournal.org. The editors will choose which letters will be published. All published letters will be subject to editing for style and length.

Scoring Social Security Proposals

Response from Kashin, King, and Soneji

We are grateful to Peter Diamond for his interest in our article, which offered the first systematic evaluation, by anyone in or out of government, of the Social Security Office of the Chief Actuary's demographic and financial forecasts and policy scores. We demonstrated that these forecasts depend on nontransparent, unreplicable, and antiquated methods and, as a result, are systematically biased and overconfident.

To clarify what is at issue here, the Office of the Chief Actuary makes baseline forecasts for the future of Social Security, and also estimates the effects of proposed policy changes. It does not offer any uncertainty estimates. Our paper makes claims about severe bias in the baseline estimates, and further claims that similar or greater bias exists in estimates concerning proposed policy changes. We also offer estimates for the extent of uncertainty implied by these biases. Diamond's letter offers no objection to our claims about bias in the baseline estimates or policy proposal, or about our arguments concerning uncertainty surrounding the baseline estimates. Diamond's criticism focuses on the two paragraphs in our article that seek to provide the first uncertainty estimates ever for the gap between the policy counterfactual C and the baseline estimate B.

Diamond offers a thought experiment about estimating the uncertainty in the 75-year forecast around a policy change involving a $1 million payment in the present. His analysis is a special case built on three underlying assumptions, two of which are incorrect in the present setting and a third which depends on an arbitrary choice of a theory of inference. We appreciate the opportunity to clarify how uncertainty depends on the magnitude of the policy shock and covariance, which of course we did not ignore.

First, we switch from Diamond's hypothetical small policy to the more realistic actual massive proposals evaluated by the Office of the Chief Actuary. These include (at the median over the last 15 years) five major provisions, 28 complicated interactions, and an estimated change in the actuarial balance of 100 percent. The uncertainty over time in the costs of these counterfactual proposals C equals i) the uncertainty in the baseline estimates B, plus ii) the uncertainty due to assumptions about each proposal's provisions, interactions, and never-before-observed effects. As a result, the standard deviation of C is much larger than B, that is, $a \equiv \sqrt{V(C)}/\sqrt{V(B)} \gg 1$. An estimate for this ratio, using all proposals evaluated by the Office since 2000, is $a = 3.5$, or $a = 4.3$ after adjusting for characteristics of policies and proposers. Yet, even if a is as small as 2, $V(B)$ is a lower bound of $V(C - B)$, as we claimed. The bound is obtained by rewriting $V(C - B) = V(C) + V(B) - 2\,\mathrm{Cov}(B,C) = [1 + a(a - 2r)]\,V(B)$.

Second, as Diamond writes, his analysis "assume[s] that the two baseline forecasts of the cost rate are equal to the expected values of the baseline cost rates." This unbiasedness claim has been false for 15 years, as documented in our article. Biases in estimates of baseline forecasts B are large and increasing (even though the Office of the Chief Actuary had the luxury of basing its forecasts on a long observed historical record). To claim that forecasts of counterfactual proposals C (based on no observed history) are somehow less biased than B, or to claim that we know that these biases in some way cancel each other out, requires believing in implausible and unobservable coincidences.

Third, consider the correlation r between errors in B and C across reruns of policy changes across a range of plausible worlds with implementation at year 0 and measurement 75 years later. How one thinks about this correlation actually depends on one's chosen theory of inference. Under frequentist theory, the true potential outcomes are fixed (and so cannot contribute to the variance) and the forecasts are random but almost identical. In this setting, the kind of hypothetical small policy described by Diamond may have r close to 1. In effect, this theory results in recognizing error in observable quantities, while implausibly assuming perfect foresight and no uncertainty for unobserved quantities—that is, in effect assuming that $V(C - B) \approx 0$. In contrast, under our preferred Bayesian theory, the true outcome has a random distribution and the forecasts are fixed thus, if a correlation between the errors of B and C is induced at year 0, that correlation can degrade over time. If the correlation degrades after 75 years to $r = 0.5$, our claim holds even in the unlikely case that $a = 1$, or more generally if $a/r \geq 2$. The forecasts of the Office of the Chief Actuary implicitly take the Bayesian view: across all proposals evaluated since 2000, the empirical correlation between their forecasts B and C is only $r = 0.51$, or $r = 0.36$ after we adjust for characteristics of policies and proposers. At one point, Diamond's letter also

seems to express support for this Bayesian view, when he writes that "the variances in forecast errors equal the variances in *outcomes*" (emphasis added).

We hope future researchers will improve our uncertainty estimates. Ignoring uncertainty, or assuming it away, does a disservice to science, public policy, and millions of current and future retirees. If the Social Security Administration would follow scientific standards, the replication movement in academia, and recent Executive Orders requiring openness and transparency, more proposals and more science could become part of the political debate.

Konstantin Kashin, Harvard University, Cambridge, Massachusetts

Gary King, Harvard University, Cambridge, Massachusetts

Samir Soneji, Dartmouth College, Hanover, New Hampshire