

## The Accuracy of Multi-Year Estimates vs. Annual Estimates for R&D Tax Credit

Wendy Rotz and Jinhee Yang

Ernst and Young, LLP, 1225 Connecticut Ave. NW, Washington DC 20036

**Keywords:** cumulative estimates, tax, R&D

**Abstract:** Sampling is used to estimate corporate tax credit for qualifying research and development (R&D) expenditures. Research and development studies encompass multiple year periods, but tax is reported on an annual basis and qualification for the credit requires that annual qualifying expenditures exceed a base amount. When sampling multiple years, some taxpayers estimate overall qualifying percentages and apply these to each year's expenditures to obtain annual qualifying amounts. However, an IRS concern is that a year with qualifying expenditures below the base amount may be allowed an R&D credit because that year's low expenditures are masked by other highly qualifying years. The effect of a low year is unclear because a low year would also reduce the overall qualifying percentage for the time frame, reducing the credit in highly qualifying years, which may balance credit erroneously taken in a low year. This paper studies estimation of the R&D tax credit in multiple year studies with variable annual qualifying percentages. We explore whether an overall estimated credit for a multi-year time period could be reasonably estimated from an overall qualifying percentage, or whether separate annual estimates are necessary.

**I. Introduction.** Corporate taxpayers sample their yearly expenditures,  $E_y$ , to estimate their qualifying research expenditures,  $Q_y$ , or QREs, which is the amount of their expenditures that qualify for the federal research and development (R&D) tax credit under Section 174 of the internal revenue code. The tax credit itself is a complex function. Each year, it is calculated by:

$$\hat{T}_y = K \begin{cases} \frac{\hat{Q}_y}{2} \dots\dots\dots \text{if } \dots \frac{\hat{Q}_y}{2} \leq \hat{Q}_y - \hat{B}_y \dots \text{and } \dots \hat{Q}_y > \hat{B}_y \\ \hat{Q}_y - \hat{B}_y \dots\dots\dots \text{if } \dots \frac{\hat{Q}_y}{2} \geq \hat{Q}_y - \hat{B}_y \dots \text{and } \dots \hat{Q}_y > \hat{B}_y \\ 0 \dots\dots\dots \text{if } \dots \hat{Q}_y \leq \hat{B}_y \end{cases}$$

where,  $\hat{T}_y$  is the estimated annual tax credit,  $y$  is the year,  $K$  is a constant that depends on the tax setting and is 20% for all analyses in this study,  $\hat{Q}_y$  is the estimated annual qualifying amount,  $\hat{B}_y = \sum_i^m k_{i,y} \times \hat{Q}_{y-i}$  is the estimated base amount,  $k_{i,y}$  are statistical constants for a year, and  $m=3$  or  $5$  depending on the tax year. Note that  $\hat{B}_y$  is a linear combination of the prior years' estimated qualifying amounts. The current year qualifying amount,  $\hat{Q}_y$ , must exceed this base to claim credit for the year.

With unbiased and independent estimates of  $Q_y$ , the components of the tax credit estimate,  $\hat{Q}_y/2$ ,  $\hat{B}_y$ , and  $\hat{Q}_y - \hat{B}_y$  are each unbiased. However, conditioning

on the relation between  $\hat{Q}_y$  and  $\hat{B}_y$ , and imposing two points of truncation ( $\hat{Q}_y - \hat{B}_y$  and 0), the behavior of  $\hat{T}_y$  is not easily discerned.

Samples to estimate  $Q_y$  involve extensive interviews and great disruption at the taxpayer's place of business. These are extremely labor-intensive studies for the taxpayer and the Internal Revenue Service (IRS). Achieving reasonable sample sizes is critical, since large studies are cost prohibitive and may lead taxpayers to use less rigorous non-statistical methods to estimate R&D credit. The resource problem is compounded when, as often is the case, estimates are required for multiple years. This may occur when a corporation first makes a profit after years of development and operating at a net loss.

Taxpayers reduce sample size requirements by estimating the total qualifying amount,  $Q_T$ , over a multi-year period then apportioning the estimate,  $\hat{Q}_T$ , annually based on the known annual expenditures,  $E_y$ . That is,  $\hat{Q}'_y = \hat{P}_t \cdot E_y$ , where  $\hat{P}_t = \hat{Q}_t / E_t$  and  $E_t = \sum_y E_y$ . An advantage of this method is it requires a substantially smaller sample. The method assumes annual qualifying percentages are constant over the years, that is, annual percentages  $P_y$  equal  $P_t$ . While  $\hat{Q}'_y$  may not be as precise,  $\hat{Q}_T$  can be made reasonably accurate using a considerably smaller sample than would be required for individual annual estimates of  $Q_y$ .

A disadvantage of this approach is it raises IRS concerns with the estimated tax credit. To claim credit,  $\hat{Q}_y$  must exceed  $\hat{B}_y$ . If in reality  $P_y$  is substantially smaller than  $P_t$  some year, this low qualifying year is masked by the higher qualifying years leading to  $\hat{Q}_y$  overestimating  $Q_y$  that year and perhaps even claiming credit when it should not be allowed. However, this low qualifying year also lowers the estimated qualifying percentage,  $\hat{P}_t$ , resulting in underestimating the  $\hat{Q}_y$  in the remaining years. The question is:

Do the under and over estimations balance out so that the right amount of credit is claimed for the multi-year period?

Another complicating factor is the IRS use of downwardly biased estimates of  $Q_y$  based on using the lower bound of 90% confidence intervals. The bias is extreme when the lower bound is applied each year to  $Q_y$  rather than to the total qualifying amount,  $Q_t$ . Some in the IRS believe the annual approach is necessary due to the criteria of evaluating whether  $Q_y$  exceeds  $B_y$ .

Estimation must not only be resource efficient, but should also be reasonably accurate. This paper explores eight different multi-year estimation approaches for the R&D tax credit and evaluates the accuracy of the methods. In this study we are focusing on estimation approaches, rather than the sample designs.

Simulations based on random draws from hypothetical populations were used to evaluate the different estimation approaches. The hypothetical populations each cover three years and have varying

annual qualifying percents,  $P_y$ . The distribution of the estimated values, average error, and mean squared errors (MSEs) were used to compare the estimation techniques to discern which are robust when the assumption of constant  $P_y$  is violated.

## II. Simulation Methodology

To evaluate the different approaches, we created a hypothetical population data set spanning a three-year period from 1999 to 2001. The population file is based on the typically highly skewed financial data found in practice.

Four scenarios were developed for the estimation variable,  $Q_y$ . In each scenario, the overall three-year qualifying percentage,  $P_t$ , is the same, roughly 20%. However, there are different annual qualifying percentages in each scenario. In the first scenario, there is a constant qualifying percent. That is,  $P_y = 20\%$  for all  $y$ . In scenarios 2 through 4,  $P_{1999}$  is gradually reduced while  $P_{2000}$  and  $P_{2001}$  are increased to keep the overall percentage roughly 20%. See Figure 1 for the actual population parameters. Scenarios 1 & 2 have qualifying amounts above the base in all years. In Scenario 3, the 1999 qualifying amount is just above the base amount, but in Scenario 4, it is well below and therefore no tax credit should be allowed in that year.

A single stratified sample was designed for the three-year period. The expenditures were stratified by dollars and year so that annual estimates could be examined (although not as precisely as cumulative estimates). Three different sets of sample sizes were considered. The sample design is provided below in Figure 2.

**Figure 1. Actual Population Parameters**

Year	Qualifying Percent, ( $P_{vi}$ )				Base Amount ( $B_y$ ) (\$millions)	Qualifying Amounts ( $Q_{vi}$ ) (\$millions)				Tax Credit ( $T_{vi}$ ) (\$millions)			
	Scenario					Scenario				Scenario			
	1	2	3	4		1	2	3	4	1	2	3	4
1999	20%	15%	12%	5%	\$13	\$26	\$19	\$14	\$6	\$2.5	\$1.1	\$0.2	-
2000	20%	22%	25%	28%	\$14	\$27	\$29	\$34	\$37	\$2.5	\$2.9	\$3.3	\$3.6
2001	20%	23%	25%	28%	\$15	\$24	\$28	\$30	\$32	\$1.8	\$2.5	\$2.9	\$3.2
<b>Total</b>	<b>20%</b>	<b>20%</b>	<b>20%</b>	<b>20%</b>	-	<b>\$77</b>	<b>\$76</b>	<b>\$78</b>	<b>\$75</b>	<b>\$6.8</b>	<b>\$6.5</b>	<b>\$6.5</b>	<b>\$6.8</b>

Figure 1 shows the approximate actual values in the population studied for the qualifying percentage, base amount, qualifying amount and tax credit. All monetary values are in millions of dollars. Amounts may not sum due to rounding. In Scenario 4, since the 1999 qualifying amount is below the base amount, there should be no tax credit. In the other years the tax credit is a constant times either half the qualifying amount or the qualifying amount in excess of base, whichever is smaller.

Figure 2. Sample Design

Stratum	Stratum Boundaries	Expenditures (\$Millions)			Population Count			Sample Sizes Tested		
		$E_{1999}$	$E_{2000}$	$E_{2001}$	$N_{1999}$	$N_{2000}$	$N_{2001}$	large	medium	small
1	\$0 to \$48k	\$56	\$56	\$42	1,366	1,365	1,036	32	14	5
2	\$48k to \$59	\$40	\$39	\$38	757	741	722	32	14	5
3	\$59k to \$122k	\$27	\$33	\$34	394	482	494	32	14	5
4	\$122k and over	\$1	\$2	\$2	7	11	10	all	all	all
<b>Total</b>		<b>123</b>	<b>130</b>	<b>117</b>	<b>2,524</b>	<b>2,599</b>	<b>2,262</b>	<b>316*</b>	<b>154*</b>	<b>73*</b>

Figure 2 shows the sample design used in the simulations. The same stratum boundaries based on expenditures are used each year. The expenditures (in millions) and population counts per stratum are shown for each year. Three levels of sample sizes were tested. With the exception of the certainty stratum, the same sample sizes per stratum were used each year. \*The total sample sizes shown are the 3-year totals.

One thousand simulated samples were drawn for each sample size. The eight different estimators, based on a combined regression<sup>1</sup> calculation, were used to estimate the qualifying amount,  $\hat{Q}_y$ , and tax credit,  $\hat{T}_y$ , for each of the four scenarios. These values were compared to the true values in the simulation.

### III. Estimators

Eight different estimators are considered and compared in this study. These include techniques used by the IRS, taxpayer methods, and newly proposed techniques designed to compensate for possible annual over/under estimating in the presence of low credit years. Both annual and cumulative estimates are considered as well as point estimates and estimates based on a lower bound. Below is a description of each estimator and its formula for its computation of the annual qualifying amount, from which the tax credit is derived.

#### LBA - Lower Bound Annual

$$\hat{Q}_{LBA,y} = \hat{Q}_y - M_{90}(\hat{Q}_y),$$

where  $M_{90}(\hat{Q}_y)$  is the margin of error of the annual estimated qualifying amount based on 90% two-sided confidence.

This is the approach advocated by the IRS. LBA is downwardly biased and statistically flawed. Over a three-year period, the triangle inequality applied to annual variances can be used to demonstrate that the overall margin of error is less than the sum of the annual margins of error. That is,

<sup>1</sup> A zero intercept is expected since no credit is expected from zero expenditure. Regression estimators were used in this study to be consistent with the IRS directive that requires taxpayers to calculate the mean per unit, difference, combined ratio and combined regression estimates then use the estimate with the smallest sampling error. This is typically the regression estimator.

$$M_{90}(\hat{Q}_t) \leq \sum_y M_{90}(\hat{Q}_y).$$

LBA also defies logic. It is possible for the estimated total qualifying amount to be \$20 million, yet the IRS would only allow \$10 million (based on the sum of annual lower limits) even though a 90 percent confidence interval may have a lower bound of \$18 million for the three-year period. If a downwardly biased method must be employed to incorporate a margin of error, it should at least be based on the more accurate margin of error of  $\hat{Q}_t$ . This brings us to the LBC approaches below.

#### LBC90, LBC80, and LBC60 - Lower Bound Cumulative Approaches

$$\hat{Q}_{LBC_{90,y}} = \left( \hat{Q}_t - M_{90}(\hat{Q}_t) \right) \frac{E_y}{E_t}$$

Setting aside the statistical issue for the moment as to whether the tax credit should be based on a lower bound estimate of the qualifying amount, if a lower bound estimate is used, it makes sense to use a more accurate cumulative figure. A lower bound based on the overall margin of error still underestimates the total qualifying amount,  $Q_t$ , but the underestimation is not as severe as in the LBA method.<sup>2</sup>

Confidence intervals are used to evaluate the accuracy of point estimates. However, the IRS is using confidence limits as estimates themselves. Is 90% confidence too much for that purpose? We also investigated LBC80 and LBC60. Similar to LBC90, these are based on 80% and 60% two-sided confidence intervals respectively. These will also

<sup>2</sup>  $\hat{Q}_{LBC_{90,y}} = \hat{Q}_y - M_{90}(\hat{Q}_t) \frac{\hat{Q}_y}{\hat{Q}_t}$  is the another LBC

approach and there are many other variations of this method. We did not have time to consider these alternatives in our analyses.

underestimate the total qualifying amount, but not as severely as LBC90.

**PEA – Point Estimate Annual**

$$\hat{Q}_{PEA,y} = \hat{Q}_y$$

The annual point estimate, PEA, is an obvious logical statistical choice. It estimates  $\hat{Q}_y$  on an annual basis from an annual sample. This straightforward approach is founded in statistical theory and has negligible bias with sufficient sample sizes using combined regression estimation. Unfortunately, larger sample sizes are required for reasonable annual relative precision. However, the approach could be used with less accurate annual numbers if reasonable cumulative 3-year accuracy can be achieved.

**PEC – Point Estimate Cumulative**

$$\hat{Q}_{PEC,y} = \hat{Q}_t \times \frac{E_y}{E_t} = \hat{P}_t \times E_y$$

Using a cumulative point estimate is a possible solution to obtain more stable estimates with smaller samples. Some taxpayers prefer this method because it requires smaller sample sizes than PEA. When combined over the three year time period,  $\sum_y \hat{Q}_{PEC,y} = \hat{Q}_t$ , therefore, the total qualifying amount is correct.

The tax credit is another consideration. It can be shown that the total 3-year tax credit,  $\hat{T}_t = \sum_y \hat{T}_y$ , is not affected by a PEC approach if each year the same method,  $\hat{Q}_y \div 2$  or  $\hat{Q}_y - \hat{B}_y$ , is used to determine the tax credit. However, even if different methods are used to determine the credit, the results can still be robust for estimating the total tax credit.

This method also assumes  $P_t$  is constant. The IRS is concerned this may allow credit in a year when  $Q_y < B_y$ . However, our research shows that this does not necessary mean over-estimating  $T_t$ . Even when credit is erroneously claimed in one year, the subsequent years are underestimated and the net result can be over or under-estimating the credit for the 3-year period.

**PECH – Point Estimate Cumulative Hybrid**

$$\hat{Q}_{PECH,y} = \begin{cases} \hat{Q}_{PEC,y} & \text{if } \hat{Q}_{PEA,y} > \hat{B}_y \\ 0 & \text{otherwise} \end{cases}$$

In an effort to obtain a correct tax estimate in a year below base amount, a hybrid of the PEA and PEC approaches was developed. If the annual point estimates (calculated using only sample data from the year in question) shows an estimated QRE is below base, then take no credit that year. Otherwise, use the PEC estimate for the year.

An advantage of this method is it allows taxpayers (and the IRS) to use a less labor-intensive multi-year study. Yet it substantially reduces the chance of claiming credit in a year when it should not be allowed. The approach will even work when there is scant data per stratum on an annual basis. Rather than attempting to estimate an accurate annual value from a PEA estimator, we are only using the annual point estimate to determine whether the QREs are above base.

An obvious flaw is that the taxpayer’s qualifying amount can be substantially underestimated over the three-year period. The reduced overall qualifying percentage, gives the taxpayer too little credit in the remaining years without offsetting underestimated years with an overestimated year. However, the sample size cost savings could outweigh the reduced tax credit estimate.

A mathematical correction for the flaw in PECH can be made when a zero tax credit year is found. After dropping that year’s data from the sample and the population, we can recalculate the qualifying percentage based only on the qualifying years. Then we can apply that percentage only to the years that qualify.

**PECH2 – Point Estimate Cumulative Hybrid # 2**

$$\hat{Q}_{PECH,y} = \begin{cases} \hat{Q}_{PEC,y} & \text{if } \hat{Q}_{PEA,y} > \hat{B}_y \forall y \text{ sampled} \\ \hat{Q}'_{PEC,y} & \text{if } \hat{Q}_{PEA,y} > \hat{B}_y \\ 0 & \text{otherwise} \end{cases}$$

Where  $\hat{Q}'_{PEC,y}$  is calculated from  $\hat{P}'_t$ , the qualifying percentage over years for which  $\hat{Q}_{PEA,y} > \hat{B}_y$ .

The advantage of this method is that it allows the sample size reduction of a cumulative estimation approach without the overestimation of non-qualifying years as in PEC, nor the under estimation of qualifying years as in PECH. For this reason it is expected to be a better estimator than PEC or PECH.

IV. Results

For each scenario and sample size, the eight estimation approaches were used to estimate the tax credit in each of the 1,000 simulations. The distribution of the estimated tax credit was examined using box plots, shown in Figure 3.

Average errors are shown in Figure 4 and MSEs are in Figure 5. Only results for the midsize of 154 are founded in this paper. The remaining results are similar. All plots illustrate well the bias of any of the lower bound approaches (LBA, LBC90, LBC80, and LBC60). These are the first four estimators plotted in each scenario.

Figure 3. Box Plots of the 3-year Total Estimated Tax Credit, n=154

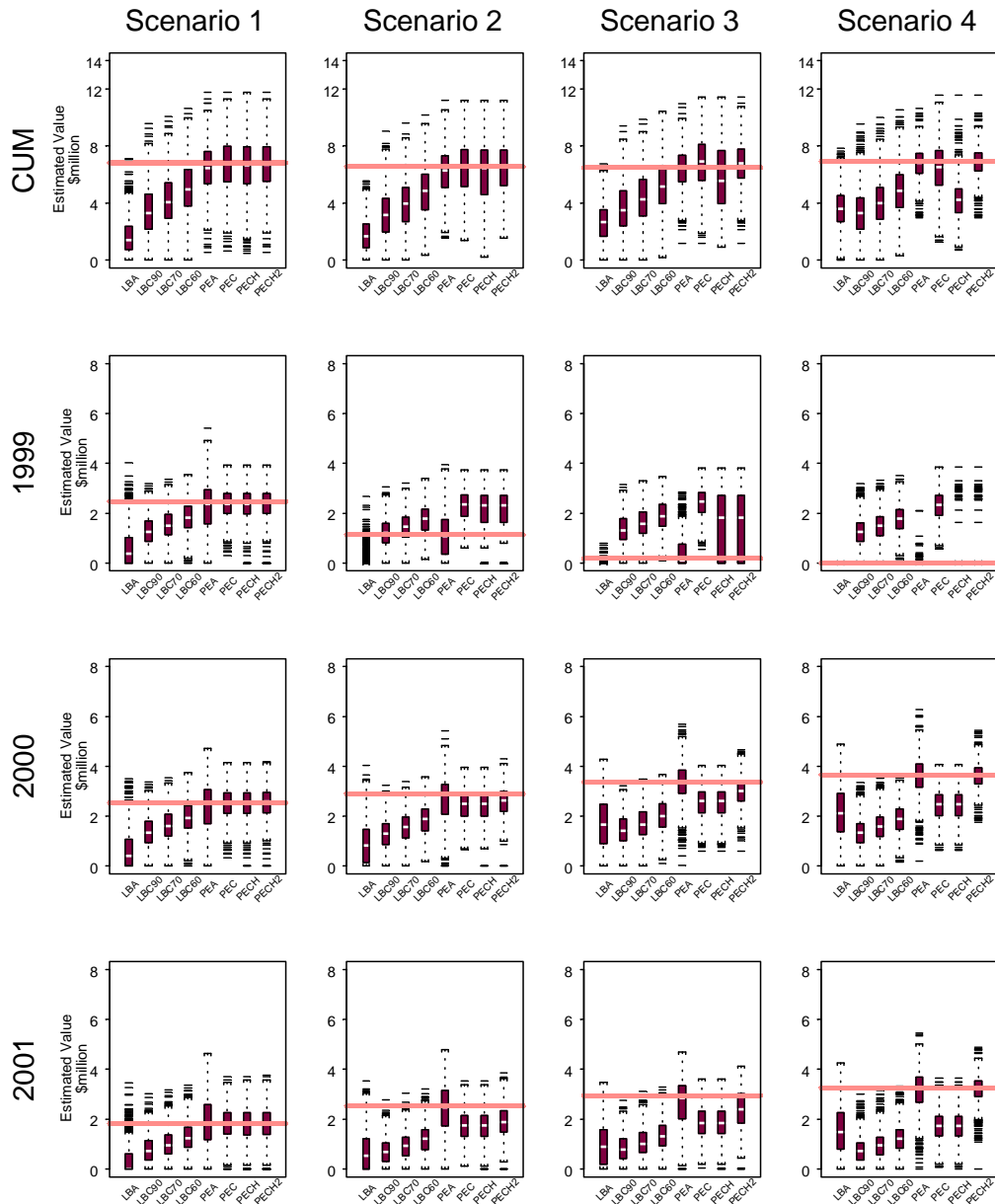


Figure 3 shows box plots for the distribution of the 1,000 simulated estimates of the total tax credit in each scenario for the middle size sample of 154 records. The rectangular box is drawn from the 1<sup>st</sup> to the 3<sup>rd</sup> quartile. The median is depicted by the line inside the rectangle. The dashes represent the outliers as defined by values that exceed 1.5 times the inner quartile range, which is marked by the brackets at the ends of the dotted lines. The horizontal line across the center of each plot shows the actual value.

**Figure 4. The Average Error of Tax Credit Estimate, n=154**

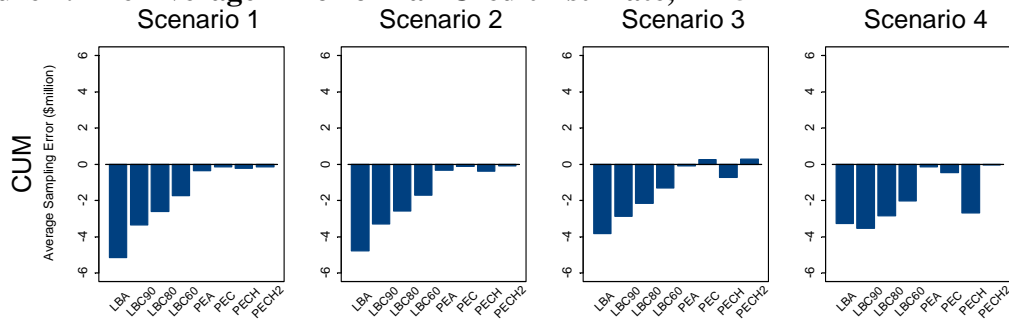


Figure 4 shows the average error from each estimation approach. Values of PEA and PECH2 are not missing in any of the graphs. The average error by these methods is too small to be discernible in some of the scenarios.

**Figure 5. Mean Squared Error of Tax Credit Estimate, n=154**

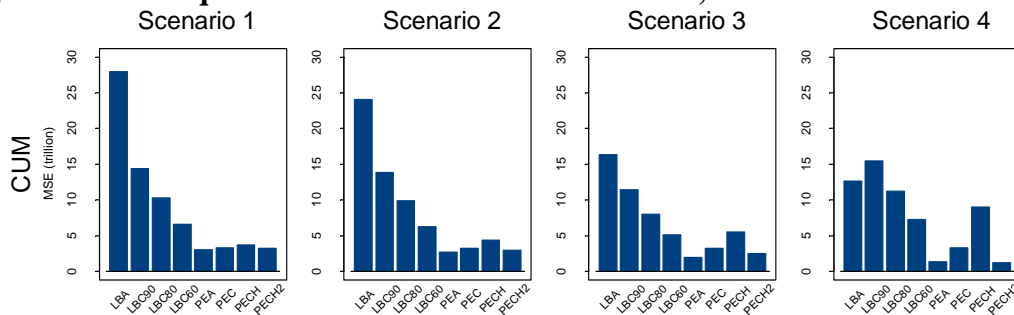


Figure 5 shows the mean squared error in the 1,000 simulations using the different estimation approaches in each scenario.

On a cumulative basis, Figure 3 shows that only outliers from the LBA method have a total tax credit anywhere near the actual value. Thus, the IRS penalty for sampling is severe. Even with LBC60, the 75<sup>th</sup> percentile is still below the actual value.

Average errors and MSEs are the largest with the lower bound approaches. If these were 1,000 actual taxpayer claims, Scenario 1 in Figure 4 shows the IRS would have erroneously denied taxpayers an average of about \$5 million using the LBA approach (and still would have undervalued the tax credit amount an average of \$2 million using LBC60).

The last four estimators in each scenario show the point estimate approaches (PEA, PEC, PECH, and PECH2). As expected, the cumulative average errors and MSEs are generally much smaller with these approaches.

Although PEC, PECH and PECH2 can overestimate the low year (1999), Figure 3 demonstrates how they also underestimate the remaining years. On a cumulative basis, these estimators seem well

balanced in that cumulative medians are near their actual values.

An exception is PECH in scenarios 3 and 4. It is downwardly biased and has larger errors than the other point estimate approaches. Recall that in Scenario 3, the 1999 qualifying amount is just above base and in scenario 4, it is below.

Figure 3 shows PECH will underestimate the credit in 2000 and 2001 but it does not overestimate enough in 1999 to compensate. This results in a net downward bias similar to the lower bound approaches. Yet PECH is an improvement to the lower bound approaches due to its better performance in Scenarios 1 and 2.

PECH2 compensates for the shortcomings of PECH. It behaves similarly to PECH except in Scenarios 3 and 4. Its average error is near zero, behaving more like PEA. The smallest cumulative MSEs are from PEA and PECH2.

Overall, PEA and PECH2 were the best estimators from a statistical standpoint.

## V. Conclusions

Currently, under the IRS LBA approach, taxpayers cannot be expected to claim their full R&D credit using statistical sampling.

True, the downward bias in taxpayer-conducted samples can be reduced by larger samples. However, achieving highly accurate samples to significantly reduce the bias on an annual basis is cost prohibitive. For example, achieving 10% relative precision with 90% confidence each year in a multi-year study may require sample sizes exceeding 1,000 selections.

Use of the least advantageous confidence limit, the lower bound, when taxpayers conduct samples, is not founded in statistics. It is an IRS policy decision.

When the IRS conducts their own samples, their guidance suggests using the most advantageous limit for the taxpayer, which is the upper bound in the case of R&D. This leads to the reciprocal policy for taxpayer conducted samples. However, other IRS guidance allows the IRS to disregard the sampling error if it is too large. Taxpayers do not have that option. Still the IRS needs some criteria to insure reasonably accurate estimates are achieved.

This research has shown that cumulative approaches, even in the presence of low qualifying years, need not adversely affect the tax credit when totaled over the multi-year period.

Taxpayers and their statisticians can continue to work with the IRS to determine alternative sampling approaches that are efficient and provide reasonable estimates for the R&D setting.

**Acknowledgements** Many thanks to Ru Sun for our data displays.

## VI References

- [1] Cochran, W.G. *Sampling Techniques*, Third Edition, John Wiley and Sons, New York, 1977
- [2] Roberts, Donald M. *Statistical Auditing*, American Institute of Certified Public Accountants, New York 1978.
- [3] Guy, D. M.; Carmichael, D.R. and Whittington, O. R. *Audit Sampling An Introduction*, 4<sup>th</sup> ed. New York, NY: John Wiley & Sons, Inc. 1998.
- [4] Kish, L., *Survey Sampling*. New York, NY: John Wiley & Sons, Inc. 1965.
- [5] Lohr, S. L., *Sampling: Design and Analysis*. Pacific Grove, CA: Brooks/Cole Publishing Company 1999.
- [6] *Research Credit Audit Techniques Guide*. IRS <http://www.irs.gov/businesses/article/0,,id=101382,00.html>
- [7] *Field Directive on the Use of Statistical and Judgement Sampling in the Research Field Cases*. IRS [http://www.irs.gov/pub/irs-utl/field\\_directive\\_samp\\_method\\_research\\_credit\\_cases.pdf](http://www.irs.gov/pub/irs-utl/field_directive_samp_method_research_credit_cases.pdf)
- [8] *Field Directive-Use of Sampling Methodologies in the Research Field Cases*. IRS [http://www.irs.gov/pub/irs-utl/field\\_directive\\_stat\\_judge\\_research\\_credit\\_cases.pdf](http://www.irs.gov/pub/irs-utl/field_directive_stat_judge_research_credit_cases.pdf)