


APA Training:
Test Periods, Averaging, Ranges,
and Testing Taxpayers' Results

October 18, 2001

Robert Weissler (these slides)

Kris Kim

Steps in Analysis

- (TPM)
 - (Comparables selection)
 - **Test period for comparables' results**
 - (Adjusting comparables' results)
 - **Combining comps' results over time**
 - **Deriving an arm's length point or range**
 - **Testing/adjusting tested party's results**
- 

Differences From Audit Context

- Mismatch between comparables test period and APA period
 - Some provisions in regulations assume match between comparables' test period and years under audit
- Taxpayer knows range in advance

Steps in Analysis

- (TPM)
- (Comparables selection)
- **Test period for comparables' results**
- (Adjusting comparables' results)
- Combining comps' results over time
- Deriving an arm's length point or range
- Testing/adjusting tested party's results

Comparables' Test Period

- Audit context
- APA context

Audit Context (-1(f)(2)(iii))

- Normally use comparable data from same year as being audited
- Circumstances warranting use of multiple year data:
 - Data availability
 - Business cycles
 - Life cycles of product or intangible
- Normally use multiple year data for CPM (normally not for CUP)

Audit Context (-1(f)(2)(iii))

- Multiple year comparable data includes one or more years before or after year under review
- Tested party's data for same period as comparables "ordinarily must be considered"
- If "such data is not available," can use reliable data from other years, adjusted if necessary [what does this mean?]

Audit Context (-5(b)(4))

- For CPM, comparables test period generally should include at least year under review plus preceding two years

APA Context

- Criteria for using multiple year data still make sense
- Hard to match comparable years with years under review
 - Refreshing comparable data after agreement means uncertain process and result
- Can update data during negotiations to get as late a test period as possible

How Many Comparable Years?

- Default of 3?
- Enough to cover cycles/fluctuations
- Avoid far past if circumstances changed
- Avoid many years unless needed (extra effort required)

Match Between Comparables' and Tested Parties' Years for Adjustments (e.g., Asset Intensity)

- Tested party's year ends Oct., 1999. What comparable years to pick? Some choices:
 - 6-month overlap (e.g., FYE April, 1999 – March, 2000)
 - FYE in Calendar 1999
 - FYE in Compustat's "fiscal 1999" (June, 1999 – May, 2000)
 - Select comparables by FYE

Steps in Analysis

- (TPM)
- (Comparables selection)
- Test period for comparables' results
- (Adjusting comparables' results)
- **Combining comps' results over time**
- Deriving an arm's length point or range
- Testing/adjusting tested party's results

Statistics Primer

- Two definitions of the “middle” of a set of results:
 - Mean (“arithmetic mean,” “average”)
 - Median

Statistics Primer

- Mean (“arithmetic mean,” “average”):
Add up the results and divide by the number of the results
- Median:
Take the middle result (if odd number of results) or the mean of the two middle results (if an even number of results)

$$\text{Mean} = (10+9+8+2+1)/5 = 30/5 = 6$$

10
9
8
2
1

Median = 8

Median = $(8+2)/2 = 5$

10
9
8
2
1

10
9
8
2
1
0

	10	100
	9	9
	8	8
	2	2
	1	1
Median	8	8
Mean	6	24

Combining Comparables' Results from Different Years

- Regulations speak of comparing tested party's "average result over a multiple year period with the average results of uncontrolled comparables over the same period" (-1(f)(iii)(D))
- APA Program has preference for averaging each comparable's results over time (even if tested party is tested yearly)

Ways to Average

- Simple average
- Average weighted by denominator of PLI (e.g., sales weighted average with operating margin PLI). This is equivalent to taking totals over the period (e.g., total operating profit divided by total sales).

Pooling

- Treat each comparable's result for each year as a separate observation
- May be appropriate in some cases when data for some comparables is missing for some years
- But APA policy normally favors averaging even if some data is missing

Co.	Yr 1	Yr 2	Yr 3	Yr 4	Yr 5	Ave
A			3	4	5	4
B	2	3	4	5	5	3.8
C	4	5	4	3	2	3.6
D	3	4				3.5
E	5	4	3	2	1	3

Pooling: 4.5 top, 4 median, 3 bottom
5, 5, 5, 5, **5**, **4**, 4, 4, 4, **4**, **4**, 3, 3, 3, **3**, **3**, 2, 2, 2, 1

Debate

- Average, even with missing data
- Pool, especially with missing data
- Require complete data
- Other approach?

Data Patterns

- Comparables far apart, but each comparable varies little from year to year
- Comparables not far apart, but each comparable varies randomly a great deal from year to year
- Substantial business cycles

One Practitioner Argument

- Pool if tested party will be tested yearly
- Average if tested party will be tested on average basis

- It is unfair to test the tested party annually against a range determined by averaging comparables' results over time

APA Program Response

- Cure any unfairness by testing the tested party on an average basis (so pooling not needed)

Steps in Analysis

- (TPM)
- (Comparables selection)
- Test period for comparables' results
- (Adjusting comparables' results)
- Combining comps' results over time
- **Deriving an arm's length point or range**
- Testing/adjusting tested party's results

Point or Range? (-1(e)(1))

- “In some cases, application of a pricing method will produce a single result that is the most reliable measure of an arm’s length result. In other cases, application of a method may produce a number of results from which a range of reliable results may be derived.”

How Determine Range? (-1(e)(2))

- Normally determined by applying single TPM to two or more uncontrolled transactions
 - Might also be determined by using two or more TPMs (reference to -1(c)(2)(iii))
- Select top tier of comparables (exclude ones with “significantly lower level of comparability and reliability”)

Use the Full Range?

- Yes, if “the information on the controlled transaction and the uncontrolled comparables is sufficiently complete that it is likely that all material differences have been identified, each such difference has a definite and reasonably ascertainable effect on price or profit, and an adjustment is made to eliminate the effect of each such difference.” (-1(e)(2)(iii)(A))

If Cannot Use Full Range

- Three step logic (-1(e)(2)(iii)(B):
 - Increase reliability of result using “valid statistical method”
 - In general, such a statistical method is based on 75 percent probabilities
 - Interquartile range generally deemed a sufficient approximation to the 75 percent probabilities, though other method could be used

Logical Step 1: Increase Reliability

- “The reliability of the analysis must be increased, where it is possible to do so, by adjusting the range through application of a valid statistical method to the results of all of the uncontrolled comparables so selected.”

Logical Step 2: 75% Probabilities

- “The reliability of the analysis is increased when statistical methods are used to establish a range of results . . . such that there is a 75 percent probability of a result falling above the lower end of the range and a 75 percent probability of a result falling below the upper end of the range.”

25% of results above range

50% of results within range

25% of results below range

Logical Step 3: Interquartile Range (IQR)

- “The interquartile range ordinarily provides an acceptable measure of this range; however a different statistical method may be applied if it provides a more reliable measure.”

Interquartile Range (-1(e)(2)(iii)(C))

- Bottom (25th percentile) is “lowest result such that at least 25 percent of the results are at or below the value of that result”
- But “if exactly 25 percent of the results are at or below a result,” bottom is “the average of that result and the next higher result”

Interquartile Range Top “Determined Analogously”

- Top (75th percentile) is lowest result such that at least 75 percent of the results are at or below the value of that result
- But if exactly 75 percent of the results are at or below a result, top is the average of that result and the next higher result

Top: Average 4 & 3

4	100%
3	75%
2	50%
1	25%

Top: 4

5	100%
4	80%
3	60%
2	40%
1	20%

Bottom: Average 1 & 2

Bottom: 2

Possible Statistical Alternative to IQR

- For large sets (at least 30 results) with a normal distribution, compute estimated mean and standard deviation; use band of 0.675 estimated standard deviations around estimated mean
- Not specifically described in regulations; has been used occasionally
- Comparables should be homogeneous

Practical Choices in APAs

- Point (especially royalties, services)
- Range
 - Full
 - Interquartile
 - Other
- Floor or ceiling (disfavored now)
- Special cases
 - Statistical confidence intervals (Fin. Products)
 - Residual profit splits

Practical Choices in APAs

- Debate: IQR vs. point or smaller range

Residual Profit Splits

- Structure of residual profit split:
 - Assign routine profit to each party
 - Split residual profit or loss based on intangible contribution
- Usually split residual in definite ratio (point, not range)
- Use points or ranges for routine profits?

Residual Profit Splits

- Simplest approach is to use points for routine profits. Then get point result for total U.S. profit.
- If use ranges for routine profits, will get range of total U.S. profit.

- System profit 9
- Both routine profit IQRs are 1 to 4
- Split residual 50%/50%

System profit	9	9
US routine profit	4	1
Japanese routine profit	1	4
Residual profit	4	4
U.S. share of residual profit	2	2
Total U.S. profit	6	3

Should Taxpayer Have Freedom of IQR for Each Routine Return?

- Taxpayers say regulations give them IQR
- But regulations did not contemplate two sets of comparables
- Nulon example of residual profit split in reg. 1.482-6(c) gives point for routine return (even though only one set of comparables)

Should Taxpayer Have Freedom of IQR for Each Routine Return?

- Basic principle: rough justice: probabilities
- One set of comparables: allow result if at least 25% of the time it would be that high or low by chance
- At least 25% of time Japan's routine profit would be as high as top of IQR. At least 25% of time U.S. routine profit would be as low as bottom of IQR. But now both are happening at once!

Steps in Analysis

- (TPM)
- (Comparables selection)
- Test period for comparables' results
- (Adjusting comparables' results)
- Combining comps' results over time
- Deriving an arm's length point or range
- **Testing/adjusting tested party's results**

If Outside Range: Audit Context

- Service may adjust to any point within arm's length range (-1(e)(3))
 - If IQR used, normally adjust to median (definition in regulations of median amounts to same as normal statistical definition)
 - Otherwise, normally adjust to mean
- Taxpayer may self-adjust only on timely filed return “[i]f necessary to reflect an arm's length result” (-1(a)(3))

If Outside Range: APA Context

- An APA “may permit . . . a compensating adjustment” (Rev. Proc. 96-53, section 11.02); APAs generally do permit this
 - Some early APAs said taxpayer in violation if too far from range
- Often adjust to nearest edge of range
- Can adjust to median or similar point to discourage targeting the edge of the range

Compensating Adjustments (Rev. Proc. 96-53, section 11.02)

- Payment between related parties after tax year ends, by certain deadline
- If paid properly, no interest on intercompany account
- If paid properly, free ride for estimated tax
- Taxpayer still must pay applicable interest on any extra tax due

Time Period To Test Tested Party: Audit Context (-1(f)(2)(iii))

- Normally, test tested party on annual basis
- If use multiple-year averages for comparables, test tested party over (normally same) multiple year period

Audit Adjustments Using Multiple-year Testing

- If tested party's multiple-year results are outside range, adjust year being audited as usual (normally to median or mean) but “only to the extent that it would move the controlled taxpayer's multiple year average closer to the arm's length range for the multiple year period or to any point within such a range.”

Range 1 to 3; Can't adjust last year enough
to get three-year average in range

Yearly Result	2	1	3	5	-2	0	-1
3-Year Rolling Average			2	3	2	1	-1

Range 1 to 3; Don't adjust the last year because it would move three-year average in wrong direction

Yearly Result	3	3	-3	9	-3	-3	4
3-Year Rolling Average			1	3	1	1	-.67

How the APA Context Differs

- Can “review” multiple years at once (up to whole APA term)
- Mismatch between APA Term and comparables’ test period
- Taxpayer knows in advance the range and testing method

Single or Multiple Year Test?

- Presumption for single year test (avoid deferral)
- Can overcome presumption:
 - Cyclical nature of industry or products
 - Renewals
 - Treaty partner concerns
 - Prior conduct

Methods of Averaging

- Three-year rolling average
 - Traditionally common
 - Complicated scenarios
 - Two bites at cherry
 - Five-year pattern: low, low, high, low, low
 - Might agree to adjust current year beyond range if needed to get rolling average within range

Methods of Averaging

- Average over whole term (problem of deferral, disfavored)
- Cumulative average (e.g., true up from second year on) (not yet an established approach)
- Somewhat relaxed individual year ranges combined with stricter average range

End of Presentation