THE TIME-SERIES PROPERTIES OF BOOK AND TAXABLE INCOME*

George A. Plesko and *David P. Weber,* University of Connecticut

INTRODUCTION

USING CONFIDENTIAL TAX RETURN DATA, WE provide a descriptive analysis of the time series properties of various measures of "book" (i.e., financial statement) and taxable income to determine the reasonableness of common assumptions about the pattern of earnings under each reporting system. Because tax return information is typically unavailable to academic researchers, financial statement information is often used to infer firms' tax positions. Therefore, understanding the specific elements of the financial statements that help to infer a firm's tax position is essential to using publicly available information in tax research. For example, Hanlon and Shevlin (2002), in their analysis of the accounting for stock options, argue that researchers who do not properly address the related accounting issues may significantly mismeasure taxable income. Indeed, even determining the sign of a firm's taxable income can be difficult given the effects of accounting rules, yet measures of tax rates depend on the researcher being able to do so. Plesko (2003, 2007), for example, shows that common estimates of taxable income drawn from financial statements can misclassify the sign of taxable income in approximately 12 percent of the cases. Hanlon (2003) and McGill and Outslay (2002, 2004) provide general discussions of the difficulties in inferring tax positions from financial statement information.

In addition to this potential for error in measures of current taxable income, projections of expected future taxable income are frequently used to estimate firms' marginal tax rates for studies of the effect of taxes on various corporate decisions (e.g., capital structure), and it is commonly assumed that the time series properties of book income apply equally to taxable income. For example, Shevlin (1990), Graham (1996), and Graham and Mills (2008) assume that both book and taxable income follow a random walk pattern.[1] However, evidence of increasing aggregate book-tax differences over time (e.g., Manzon and Plesko, 2002) suggests the importance of examining the empirical descriptiveness of this assumption.

MODEL

We estimate the following simple univariate model to examine the time-series properties of various measures of income (Y):

$$\textbf{(1)} \quad Y_t = \alpha + \beta Y_{t-1} + \varepsilon.$$

Similar models have been used to estimate the persistence of book measures of income (e.g., Sloan, 1996; Hanlon, 2005). Unlike the random walk model, equation (1) does not assume that income is perfectly persistent, and thus it allows for mean reversion and the existence of transitory items. We estimate this model for each of our income measures, which allows for comparisons of β across measures. Based on accounting literature suggesting that losses are expected to be less persistent than positive income (e.g., Hayn, 1995; Joos and Plesko, 2005), we also estimate the model separately for subsamples conditional on the sign of Y_{t-1}.

MEASURES OF INCOME

We employ two measures of pretax book income. The first, $PTBI^{ALL}$, includes income from all sources worldwide and facilitates comparisons to prior literature on the properties of book income. The second, $PTBI^{DOM}$, includes only income from domestic sources and is thus more comparable to our tax return information. The information for worldwide and domestic pretax income is taken from Compustat (data items *pi* and *pidom*, respectively).[2]

In addition, we construct two financial statement-based estimates of U.S. taxable income. The first, TI^{EST1}, is U.S. current tax expense (Compustat item *txfed*) grossed up at the top statutory tax rate, similar to (Shevlin, 2002). The second, TI^{EST2}, adjusts TI^{EST1} for the use of net operating loss (NOL) carryforwards, similar to Hanlon et al. (2005). Specifically, $TI^{EST2} = TI^{EST1} - \Delta NOL$, where *NOL* is Compustat item *tlcf*.[3]

*The views expressed in this paper are those of the authors, and should not be attributed to the Internal Revenue Service or the Department of the Treasury. All errors or omissions are our own.

Finally, we include a measure of income based on actual tax return information. We measure actual taxable income, TI^{ACT}, as Form 1120 U.S. Corporate Income Tax Return line 28 ("Taxable income before net operating loss deduction and special deductions"). We use line 28 rather than line 30 ("Taxable income"), because line 30 is affected by the use of NOL carryforwards originating in previous years, as well as special deductions, and is bounded below by zero. Thus, line 28 better reflects the firm's current year performance as measured by tax rules. We scale each of our measures of income by average total assets.

SAMPLE AND DESCRIPTIVE STATISTICS

Our financial statement information is drawn from firms' Form 10-K filings, as collected by Compustat, and matched to the corresponding tax returns by Employer Identification Number (EIN) for the years 1986 to 2007. The primary sample consists of all firm-years that can be matched from the two data sources and that have the necessary information available to construct each of our five measures of income. To reduce the effect of outlying observations, we trim the sample at the top and bottom 2 percent of the distribution of each income measure. This produces 48,181 usable firm-year observations for our empirical analyses.

Summary statistics for the sample are presented in Table 1. Consistent with previous research on aggregate book-tax differences, taxable income, as measured by TI^{ACT}, tends to be lower than the corresponding book measure, $PTBI^{DOM}$.

Correlations across the variables are reported in Table 2 (Pearson correlations below the diagonal, Spearman correlations above). As expected, the correlations between TI^{ACT} and the other income

Table 1
Descriptive Statistics

Variable	*Mean*	*Std Dev*	*1%*	*25%*	*Median*	*75%*	*99%*
$PTBI^{ALL}$	0.025	0.156	-0.59	-0.01	0.05	0.11	0.27
$PTBI^{LDOM}$	0.018	0.151	-0.59	-0.02	0.05	0.10	0.26
TI^{EST1}	0.045	0.062	-0.05	0.00	0.02	0.08	0.24
TI^{EST2}	0.028	0.115	-0.46	0.00	0.03	0.08	0.25
TI^{ACT}	0.016	0.129	-0.48	-0.02	0.03	0.09	0.25

Notes: Number of observations: 48,181.
To ensure confidentiality, the percentile values are the rounded values of the average of observations around the cut-off.

Table 2
Pearson and Spearman Correlation Coefficients

	$PTBI^{ALL}$	$PTBI^{LDOM}$	TI^{EST1}	TI^{EST2}	TI^{ACT}
$PTBI^{ALL}$		0.963	0.787	0.766	0.788
$PTBI^{LDOM}$	0.983		0.802	0.782	0.787
TI^{EST1}	0.577	0.587		0.879	0.767
TI^{EST2}	0.628	0.633	0.652		0.754
TI^{ACT}	0.793	0.795	0.585	0.617	

Notes: Pearson coefficients are presented in the bottom left triangle; Spearman coefficients in the upper right. All correlations are significant at 0.001 or less.
Number of observations = 48,181.

Table 3
Persistence Regressions

Panel A: Pretax Book Income

	$Y = PTBI^{ALL}$			$Y = PTBI^{DOM}$		
	Y_{t-1}	*Adj* R^2	*n*	Y_{t-1}	*Adj* R^2	*n*
Pooled	0.711	0.511	48,181	0.707	0.508	48,181
Pooled, $Y_{t-1} > 0$ only	0.794	0.316	35,395	0.779	0.309	34,572
Pooled, $Y_{t-1} \leq 0$ only	0.610	0.308	12,786	0.619	0.323	13,609
Annual Regressions						
1986	0.726	0.400	1,376	0.722	0.400	1,376
1987	0.660	0.471	1,696	0.643	0.464	1,696
1988	0.738	0.500	1,711	0.732	0.487	1,711
1989	0.725	0.517	1,759	0.716	0.509	1,759
1990	0.779	0.479	1,759	0.768	0.471	1,759
1991	0.677	0.466	1,820	0.675	0.466	1,820
1992	0.749	0.483	1,895	0.748	0.478	1,895
1993	0.692	0.418	2,004	0.698	0.425	2,004
1994	0.658	0.458	2,071	0.658	0.452	2,071
1995	0.698	0.462	2,210	0.687	0.455	2,210
1996	0.697	0.461	2,370	0.694	0.461	2,370
1997	0.699	0.439	2,443	0.690	0.434	2,443
1998	0.697	0.439	2,530	0.698	0.444	2,530
1999	0.665	0.450	2,559	0.656	0.445	2,559
2000	0.718	0.486	2,841	0.710	0.475	2,841
2001	0.697	0.476	2,734	0.699	0.484	2,734
2002	0.662	0.470	2,614	0.666	0.469	2,614
2003	0.680	0.573	2,486	0.676	0.568	2,486
2004	0.738	0.580	2,528	0.733	0.578	2,528
2005	0.760	0.611	2,345	0.758	0.614	2,345
2006	0.791	0.601	2,275	0.786	0.604	2,275
2007	0.803	0.635	2,155	0.796	0.624	2,155
Mean of Annual	0.714	0.494	2,190	0.709	0.491	2,190

Table 3 (continued)
Persistence Regressions

Panel B: Measures of Taxable Income

	$Y = TI^{EST1}$			$Y = TI^{EST2}$			$Y = TI^{ACT}$		
	Y_{t-1}	$Adj\ R^2$	n	Y_{t-1}	$Adj\ R^2$	n	Y_{t-1}	$Adj\ R^2$	n
Pooled	0.722	0.573	48,181	0.541	0.313	48,181	0.719	0.527	48,181
Pooled, $Y_{t-1} > 0$ only	0.724	0.510	33,159	0.614	0.236	32,992	0.665	0.279	32,409
Pooled, $Y_{t-1} \leq 0$ only	-0.110	0.009	15,022	0.424	0.161	15,189	0.701	0.387	15,772
Annual Regressions									
1986	0.754	0.611	1,376	0.611	0.315	1,376	0.707	0.434	1,376
1987	0.810	0.568	1,696	0.553	0.247	1,696	0.628	0.419	1,696
1988	0.687	0.528	1,711	0.563	0.319	1,711	0.718	0.493	1,711
1989	0.708	0.582	1,759	0.586	0.413	1,759	0.711	0.501	1,759
1990	0.719	0.597	1,759	0.625	0.334	1,759	0.707	0.453	1,759
1991	0.715	0.621	1,820	0.570	0.318	1,820	0.673	0.495	1,820
1992	0.735	0.578	1,895	0.604	0.360	1,895	0.736	0.528	1,895
1993	0.737	0.608	2,004	0.519	0.246	2,004	0.755	0.539	2,004
1994	0.807	0.640	2,071	0.508	0.301	2,071	0.790	0.573	2,071
1995	0.746	0.594	2,210	0.635	0.398	2,210	0.726	0.518	2,210
1996	0.728	0.573	2,370	0.600	0.356	2,370	0.728	0.495	2,370
1997	0.715	0.557	2,443	0.580	0.338	2,443	0.734	0.527	2,443
1998	0.687	0.542	2,530	0.594	0.373	2,530	0.768	0.540	2,530
1999	0.686	0.499	2,559	0.544	0.303	2,559	0.746	0.522	2,559
2000	0.702	0.521	2,841	0.581	0.321	2,841	0.705	0.475	2,841
2001	0.583	0.465	2,734	0.487	0.255	2,734	0.586	0.428	2,734
2002	0.638	0.478	2,614	0.480	0.271	2,614	0.747	0.516	2,614
2003	0.711	0.556	2,486	0.559	0.302	2,486	0.760	0.596	2,486
2004	0.793	0.580	2,528	0.454	0.224	2,528	0.741	0.534	2,528
2005	0.805	0.576	2,345	0.479	0.241	2,345	0.752	0.544	2,345
2006	0.803	0.628	2,275	0.511	0.279	2,275	0.680	0.552	2,275
2007	0.763	0.622	2,155	0.541	0.305	2,155	0.766	0.574	2,155
Mean of Annual	0.729	0.569	2,190	0.554	0.310	2,190	0.721	0.512	2,190

Notes: This table reports regression coefficients from estimating the following cross-sectional model: $Y_t = \alpha + \beta Y_{t-1} + \varepsilon$. Pooled regressions include year fixed effects.

measures are all reasonably high, but far from perfect. Interestingly, the Pearson correlations between TI^{ACT} and the financial statement-based estimates of taxable income (TI^{EST1} and TI^{EST2}) are both lower than those between TI^{ACT} and the book measures of income ($PTBI^{ALL}$ and $PTBI^{DOM}$).

RESULTS

We begin by estimating equation (1) for each measure of income using the entire pooled sample. Results for the measures of book income are reported in Panel A of Table 3, and the measures of taxable income are reported in Panel B. Across all income measures except TI^{EST2}, the coefficient on lagged income (Y_{t-1}) is similar, with estimates near 0.700, which is a level consistent with prior research on pre-tax book income, such as Hanlon (2005). This level is also substantially below unity, which supports recent research suggesting that modeling income as a random walk may not be appropriate (Blouin, Core, and Guay, Forthcoming; Graham and Kim, 2009). The estimated persistence coefficient for TI^{EST2} is lower than the others at 0.541, likely reflecting the data quality issues surrounding NOLs mentioned earlier.

To check the robustness of our pooled sample results as well as identify any time trends, we also estimate equation (1) using separate annual cross-sectional regressions. These results are included in both panels of Table 3. Across time, the estimated coefficients are fairly stable for each measure, though there does appear to be a minor dip in magnitude around 2001 for the measures of taxable income. The means from the annual regressions are all very close to the pooled results. Overall, it appears that the pooled results are reasonably representative and are not overly influenced by any particular years.

Because losses are generally expected to have different patterns of persistence than positive income (Joos and Plesko, 2005), we also estimate the pooled model separately for subsamples based on whether lagged income is positive ($Y_{t-1} > 0$) or non-positive ($Y_{t-1} \leq 0$). These results are also reported in Table 3. As expected, the persistence of the book income measures in Panel A is lower for the loss firm subsamples. Turning to the measures of taxable income in Panel B, the results for TI^{EST2} are consistent with the book income results in that losses are again substantially less persistent. For TI^{EST1}, the estimated coefficient for loss firms is actually negative, which likely reflects accounting conventions where a current tax benefit (i.e., negative current tax expense) is generally only reported to the extent that losses can be carried back to offset taxable income from previous years. Such amounts are not likely to persist. Finally, for actual taxable income, TI^{ACT}, unlike the book and book-based measures, there is little difference in persistence between income and loss firms. A possible explanation is that conservatism in book accounting may contribute to book losses being recorded in a more "lumpy" (and thus less persistent) manner than tax losses.

Taken as a whole, three basic themes emerge from Table 3. First, the persistence of book and taxable income measures are fairly similar for the full sample. Second, both of the two most common financial statement-based estimates of

Table 4
Distribution of Firm-Specific Regression Coefficients

Measure	*Mean*	*Std Dev*	*25%*	*Median*	*75%*	*Unique Firms*
$PTBI^{ALL}$	0.468	0.342	0.259	0.487	0.681	1,696
$PTBI^{DOM}$	0.471	0.343	0.265	0.494	0.688	1,696
TI^{EST1}	0.445	0.498	0.212	0.473	0.672	1,709
TI^{EST2}	0.369	0.359	0.120	0.397	0.636	1,703
TI^{ACT}	0.406	0.345	0.195	0.435	0.635	1,696

Notes: This table reports the distribution of regression coefficients from estimating the following model as a series of individual, firm-specific regressions: $Y_t = \alpha + \beta Y_{t-1} + \varepsilon$, where Y represents a particular measure of income. A minimum of 10 usable observations per firm was required to estimate the regressions.

taxable income differ in their persistence properties from actual taxable income. Third, persistence is substantially lower for loss firms, based on book measures of income, but not for actual taxable income.

As a final analysis, we also estimate equation (1) as a series of firm-specific time-series regressions. Estimating separate firm-specific regressions allows us to observe a distribution of persistence parameters across firms. Summary statistics from the coefficient estimates in these regressions are reported in Table 4. While the distributions are fairly similar across income measures (with the exception of TI^{EST2}, which again tends to be lower), it is noteworthy how much variation exists across firms within each measure. For example, the median estimated coefficient for TI^{ACT} (0.435) is more than double the 25th percentile (0.195). Likewise the 75th percentile for TI^{ACT} (0.635) is much larger than the median. These results highlight the potential importance of using firm-specific information in forecasting measures of future income (e.g., Graham and Kim, 2009).

CONCLUSION

The results of this paper suggest three related, and important, inferences about the time-series behavior of book and taxable income, and can inform those who need to estimate future taxable income. First, with the exception of TI^{EST2} (which accounts for changes on NOLs in estimating taxable income), book and taxable income have similar time-series properties, with persistence coefficients differing by less than 0.015. While this result does not address whether the underlying estimate of taxable income is accurate, it does suggest that changes in income follow similar patterns, which is important for forecasting future amounts of income.

Second, loss firms appear to have a very different pattern of earnings persistence than profitable firms, which raises an important caveat about the first conclusion. For profitable firms, taxable income is estimated to be similar, but slightly less persistent than book income. Examining firms with non-positive earnings, however, shows both book and taxable income estimated from financial statements to be substantially less persistent for these firms. By contrast, the persistence of actual taxable income is similar across firms with positive and non-positive income.

Third, the aggregate estimation results disguise the existence of substantial firm-level heterogeneity. The existence of such heterogeneity strongly suggests that firm-specific information is important for estimating future book and/or taxable income.

Acknowledgments

We appreciate helpful comments from Ed Outslay, Ken Szeflinski, and session participants at the 2009 National Tax Association meetings as well as the assistance of Erin Henry. Plesko acknowledges support from the Internal Revenue Service's Tax Statistics Research Program.

Notes

[1] Blouin, Core, and Guay (Forthcoming) question the appropriateness of assuming random walks, and our evidence bears on this assumption as well.

[2] When *pidom* is missing in Compustat, we assume the firm has only domestic source income and set $PTBI^{DOM} = PTBI^{ALL}$ if pretax foreign income (*pifo*), current foreign tax expense (*txfo*), and deferred foreign tax expense (*txdfo*) are all coded as either zero or missing. Because the Schedule M-3 is unavailable prior to 2004, we are unable to utilize the book income data on the form in our time series tests. See Boynton et al. (2006) for an early analysis of Schedule M-3 data.

[3] The difference between TI^{EST1} and TI^{EST2} highlights a tradeoff. TI^{EST1} ignores NOLs, which can be problematic (e.g., reported current tax expense is bounded below by zero when firms have current year tax losses that they are unable to carryback to previous years). TI^{EST2} considers changes in NOL carryforwards, but the Compustat data on carryforwards can include other, non-U.S. jurisdictions and is not always reliable (e.g., Mills, Newberry, and Novack, 2003). Because each measure has relative strengths and weaknesses, we present both.

References

Blouin, Jennifer, John E. Core, and Wayne Guay. Have the Tax Benefits of Debt Been Overstated? *Journal of Financial Economics* (forthcoming).

Boynton, Charles E., Portia DeFilippes, Petro Lisowsky, and Ellen Legel. A First Look at 2004 Schedule M-3 Reporting by Large Corporations. *Tax Notes* 112 (September 2006) 943-981.

Graham, John R. Proxies for the Corporate Marginal Tax Rate. *Journal of Financial Economics* 42 (October 1996): 187-221.

Graham, John R. and Hyunseob Kim. Simulating Corporate Marginal Income Tax Rates and Implications for Corporate Debt Policy. Durham, NC: Duke University, 2009. Working Paper.

Graham, John R. and Lillian F. Mills. Using Tax Return Data to Simulate Corporate Marginal Tax Rates. *Journal of Accounting and Economics* 46 (December 2008): 366-388.

Hanlon, Michelle.

What Can We Infer about a Firm's Taxable Income from Its Financial Statements? *National Tax Journal* 56 (December 2003): 831-863.

The Persistence and Pricing of Earnings, Accruals, and Cash Flows When Firms Have Large Book-Tax Differences. *Accounting Review* 80 (January 2005): 137-166.

Hanlon, Michelle, Stacie Kelley Laplante, and Terry Shevlin. Evidence for the Possible Information Loss of Conforming Book Income and Taxable Income. *Journal of Law and Economics* 48 (October 2005): 407-442.

Hanlon Michelle and Terry Shevlin. Accounting for Tax Benefits of Employee Stock Options and Implications for Research. *Accounting Horizons* 16 (March 2002): 1-16.

Hayn, Carla. The Information Content of Losses. *Journal of Accounting and Economics* 20 (September 1995): 125-153.

Joos, Peter and George A. Plesko. Valuing Loss Firms. *Accounting Review* 80 (July 2005): 847-870.

Manzon, Gil and George A. Plesko. The Relation Between Financial and Tax Reporting Measures of Income. *Tax Law Review* 55 (Winter 2002): 175-214.

McGill, Gary A., and Edmund Outslay.

Did Enron Pay Taxes? Using Accounting Information to Decipher Tax Status. *Tax Notes* 96 (August 2002): 1125-1136.

Lost in Translation: Detecting Tax Shelter Activity in Financial Statements. *National Tax Journal* 57 (September 2004): 739-756.

Mills, Lillian, Kaye Newberry, and Garth Novack. How Well Do Compustat NOL Data Identify Firms with US Tax Return Loss Carryovers? *Journal of the American Taxation Association* 25 (Fall 2003): 1-17.

Plesko, George A.

An Evaluation of Alternative Measures of Corporate Tax Rates. *Journal of Accounting and Economics* 35 (April 2003): 201-226.

Estimates of the Magnitude of Financial and Tax Reporting Conflicts. Cambridge, MA: National Bureau of Economics Research, 2007. NBER Working Paper W13295.

Shevlin, Terry.

Estimating Corporate Marginal Tax Rates with Asymmetric Tax Treatment of Gains and Losses. *Journal of the American Taxation Association* 11 (Spring 1990): 51-67.

Commentary: Corporate Tax Shelters and Book-Tax Differences. *Tax Law Review* 55 (Spring 2002): 427-443.

Sloan, Richard G. Do Stock Prices Fully Reflect Information in Accruals and Cash Flows about Future Earnings? *Accounting Review* 71 (July 1996): 289-315.