Income Misattribution under Formula Apportionment

James R. Hines Jr.
University of Michigan and NBER

May 2008

I thank the University of Michigan for financial support of this research, and Owen Kearney and Fan Fei for outstanding research assistance.

Income misattribution under formula apportionment

ABSTRACT

Alternatives to the current system of separate tax accounting would apportion a firm's worldwide profits using formulas based on the location of employment, capital or sales. This paper offers a new method of evaluating the accuracy of these apportionment rules and the ownership distortions they would create. Evidence from publicly traded American corporations indicates that apportionment formulas significantly misattribute income, since the factors they use explain little of the variation in income between firms. For example, the absolute value of prediction errors from a formula based equally on property, employment and sales exceeds predicted income more than 76% of the time, and exceeds three times predicted income more than 36% of the time. As a result, the use of formulas rewards or punishes international mergers and divestitures by reallocating taxable income between operations in jurisdictions with differing tax rates. The associated deadweight loss is minimized by choosing factor weights to minimize weighted squared prediction errors, for which, based on the U.S. evidence, labor inputs should play little if any role in allocation formulas.

JEL Classifications: H25, H87, F23, G34.

James R. Hines Jr.
Department of Economics
University of Michigan
343 Lorch Hall
611 Tappan Street
Ann Arbor, MI 48109-1220

jrhines@umich.edu

1. Introduction

Governments tax active business income earned within their borders, a practice that is fraught with difficulty under any circumstances, and particularly so when a firm earns income in more than one country. Since tax rates differ between countries, multinational firms usually have incentives to arrange their affairs in ways that reallocate taxable income between countries. An excessively transparent method of doing so would be to sell a paper clip from an affiliate in a low-tax location to an affiliate in a high-tax location, charging a price of $1 million. This transaction would create a tax deduction of $1 million in the high-tax buying country, and taxable income of $1 million in the low-tax selling country, thereby reducing total global tax obligations. Cognizant of these incentives, governments have adopted arm's length pricing rules dictating that for tax purposes the prices used for intercompany transactions must be the same as those that would have been chosen by unrelated parties transacting at arm's length. Clearly, the arm's length pricing standard takes care of the problem of $1 million paper clips, but there is widespread concern that the difficulty of applying the arm's length standard to many ordinary cases, to say nothing of complex transactions involving sophisticated financial instruments or intangible property such as patents and trademarks, leaves ample opportunity for tax avoidance.

In reaction to fears about actual or potential tax avoidance under the arm's length pricing standard, there have been numerous calls for stiffer enforcement of the pricing rules, with some expressing doubt that it is possible to craft intercompany pricing rules that can ever succeed. These advocates suggest abandoning altogether the current system of separately accounting for income earned in distinct jurisdictions, replacing it with a system that uses simple formulas to apportion the worldwide income of multinational firms among the jurisdictions in which they have operations. These formulas typically use some combination of employment, sales, and tangible property as implicit indicators of where firms actually earn their incomes. American states currently use simple formulas to apportion the incomes of multistate businesses within the United States, and, relying on that experience, some (e.g., Martens-Weiner, 2006) suggest that the system might work well internationally.

In order to adopt a system of formula apportionment it is necessary to specify the weights attached to different factors used to apportion income, and the difficulty of doing so in a sensible way makes vivid at least some of the costs associated with replacing separate accounting with a system of formula apportionment. In the United States, where state governments use formulas for their corporate income taxes, it has proven difficult for governments to settle on common formulas, with the result that a firm's income might ultimately be taxed more or less than 100% by differing state governments. It is far from clear what factors properly enter an apportionment formula, assuming that governments could coordinate on common formulas, if the goal is to allocate income roughly according to where it is earned. Furthermore, as noted by Gordon and Wilson (1986), the use of formulas to apportion taxable income effectively converts an income tax into a multiple rate tax on the use of the productive factors that enter the formula, with associated deadweight loss from this haphazard diversity of tax rates. These problems have a common source, which is that profits are not scalar functions of local employment, sales or tangible property, but instead the product of many factors omitted from formulas that governments use and that have been proposed. Put simply, the formulas do not apportion income accurately among the jurisdictions in which it is earned.

There is something distasteful and very possibly inequitable about misattributing income for tax purposes, but the associated problems do not stop there, as international income misattribution creates incentives for firms to structure their affairs new ways. Since the formulas apply to affiliates within consolidated groups, it follows that the use of allocation formulas creates incentives to modify the ownership of companies or operations in order to reduce associated tax burdens. Consider, for example, a profitable German company with income taxed by Germany at a high rate. If European companies were to allocate their profits among affiliates using formulas that rely heavily on the location of employment, then the German company would have a strong incentive to acquire an unprofitable Irish company with a large labor force. In joining the German and Irish operations under common ownership, many of the German profits would be attributed to Ireland for tax purposes, where they would be subject to the much lower Irish corporate tax rate. Conversely, if the Irish company had profits and the German company did not, then the use of formula apportionment might discourage a merger of the two firms even if the merger would otherwise make sense for business reasons.

Formulary methods are typically defended as pragmatic compromises, representing imperfect alternatives to the current, arguably flawed, system of separate accounting. The purpose of this paper is to analyze the nature and magnitude of ownership distortions created by allocation formulas, the extent to which formulas misattribute income when firms merge or divest their operations. Evidence of the likely magnitude of the associated change in the allocation of taxable income can be obtained by considering the consequences of hypothetical mergers between firms that are currently independent and therefore report their incomes, employment, sales and property separately. Using data from a large sample of American publicly traded corporations, it is clear that mergers among them, treating two firms for this purpose as though located in different countries, would result in significant reallocations of taxable income, even in the absence of any effect of the mergers on actual operations or profitability.

The formulas used to attribute income between countries can be thought of as forecasts of what fraction of total firm income affiliates with given shares of employment, sales, and property are likely to earn. The analysis in this paper formalizes this notion, identifying conditions under which the formula that minimizes the efficiency cost of ownership distortions is the same as the formula that minimizes the weighted sum of squared residuals in a regression explaining total pretax income. This framework implies that regressions can be used to compare formulas that assign different weights to employment, sales, and property, as well as to construct alternative formulas.

The American evidence implies that existing formulas fare poorly from a prediction standpoint: the absolute value of prediction errors from a formula based equally on employment, sales and property exceeds predicted income more than 76% of the time, and exceeds three times income more than 36% of the time. In unconstrained regressions employment seldom gets a significant coefficient, suggesting that employment might be well omitted from formulas in practice. This is hardly surprising given that employment expenditures are costs and therefore subtractions in calculating pretax income, but employment nonetheless persists in playing significant roles in most contemplated and actual allocation formulas.

Section 2 describes current systems of separate entity accounting, evidence of taxpayer responses, and proposed formulary alternatives and their consequences. Section 3 analyzes the distortions associated with misattribution of income using formulary methods, and presents a framework that can be used to estimate the consequences of formulary alternatives. Section 4 describes the available data on large publicly-traded American corporations, and section 5 presents the empirical estimates of the magnitudes of prediction errors due to the use of formulas. Section 6 is the conclusion.

2. Separate Accounting and Formulary Methods

The current international practice of using separate accounting to determine taxable income has come under considerable fire from critics who point to the difficulties of enforcing the arm's length pricing standard against the determined behavior of taxpayers. There is considerable direct and indirect evidence that firms currently arrange their affairs in ways that relocate taxable income from high-tax countries to low-tax countries,[1] which is likely to be inefficient, and which may ultimately influence other aspects of national tax policies, such as tax rates. There are several possible solutions to this problem, including stiffer enforcement together with minor modifications of existing rules (Gresik and Osmundsen, 2006), though radical reform is always an option. Whereas separate accounting is generally acknowledged to offer a theoretically satisfying method of measuring income for tax purposes, concern over the practical ability of governments to operate separate accounting underlies much of the appeal of alternative methods of determining the location of business income.

Formulary alternatives to separate accounting can take different forms, relying to differing degrees on sales, property, and employment factors to apportion income among related parties. In the equal-weighted three-factor formula once commonly used by American states, the fraction of a firm's national income taxed by an individual state equals the average state share of the firm's sales, tangible property, and employee compensation, with each of these three factors weighted equally. States are not obliged, however, to use the same three-factor formulas, and it has never been the case that all states did so; as a result, an individual company might find that the systems used by the states in which it does business wind up taxes more or less than its total national income.

[1] See, for example, Clausing (2001, 2003), Desai et al. (2003), Hines and Rice (1994), Huizinga and Laeven (2007), and Mintz and Smart (2004); Gresik (2001) reviews some of the earlier empirical evidence.

There is considerable recent interest in possible adoption of a form of formulary apportionment within Europe, the Common Consolidated Corporate Tax Base (Commission of the European Communities, 2001). This and other formulary alternatives to current tax practice have both positive and negative attributes. Gordon and Wilson (1986) identify some of the distortions introduced by making tax obligations functions of sales, property, and employment rather than the production of income. Given these distortions, it is perhaps not surprising that American states have modified their apportionment formulas over time to emphasize the less distortionary sales factors at the expense of property and employment,[2] even though this evolution may or may not be in the interest of states as a group. One of the open questions about formulary methods in the international context is the extent to which the use of separate accounting by the rest of the world increases or reduces the desirability of using formula apportionment within a small federation of countries.[3] The adoption of formulary methods would occasion significant redistributions through changes in the tax obligations of individual firms (Shackelford and Slemrod, 1998) and the tax revenues of individual countries (Devereux and Loretz, 2008; Fuest et al., 2007). Formula apportionment has the virtue of requiring less information on location-specific profitability, though this feature may make it more difficult for governments to tailor their tax systems to extract rents from taxpayers in the most efficient possible manner (Gresik, 2008). Even the possible administrative cost and compliance benefits of formulary methods (Mintz, 2004) may be suspect; Roin (2007) notes that many of the methods taxpayers have honed in avoiding income taxes under separate accounting can be deployed to avoid taxes determined by formulary methods, in some cases at greater social cost and to greater effect than currently.

One of the costs associated with using formulary apportionment is that these systems create incentives for firms to change their operations through mergers or divestitures. Even in the absence of formula apportionment it is very common for mergers and divestitures to have significant tax effects by triggering the realization of capital gains, changing the ability to offset profits from one operation against losses from another, influencing a firm's ability to claim

[2] See Goolsbee and Maydew (2000), Anand and Sansing (2000), Edmiston (2002), Wellisch (2004), and, for an alternate viewpoint, Runkel and Schjelderup (2007).
[3] See, for example, Mintz and Weiner (2003) and Riedel and Runkel (2007).

foreign tax credits, changing asset bases for depreciation purposes, and other tax consequences.[4] There is considerable evidence that tax attributes influence the likelihood and structure of mergers and divestitures, as well as the accompanying transaction prices.[5] Consequently it is reasonable to expect that the ownership incentives created by the adoption of formula apportionment might significantly influence patterns of mergers and divestitures.

3. Distortions

This section analyzes the ownership distortions associated with the use of formulary methods, and in particular, the method by which it is possible to estimate the magnitude of these distortions using data from a cross section of independent firms.

3.1 The extent of income misattribution from using formulas.

It is helpful to consider the tax consequences of a merger of two firms, designated firm 1 and firm 2. Firm 1 operates entirely in country 1, where it faces a profit tax rate of τ_1; firm 2 operates entirely in country 2, where its profits are subject to tax at rate τ_2. Countries 1 and 2 are assumed to use the same formula to distribute the taxable incomes of multijurisdictional firms: the formula applies weights α_i to each of $i = 1,...,n$ factors such as employment, fixed capital, and sales, with $\sum_{i=1}^{n} \alpha_i = 1$. Firm 1's profits are represented by π_1, and its factors denoted $x_{1i}, \forall i = 1,...,n$; similarly, firm 2's profits are denoted π_2, and its factors denoted x_{2i}. The aggregate tax obligation for the two firms in the absence of a merger equals:

$$(\tau_1 \pi_1 + \tau_2 \pi_2). \tag{1}$$

If the two firms merge in such a way that their factor use and profits are unchanged, then the only difference produced by the merger is that their tax obligations to countries 1 and 2 will be determined by formula, and the total is:

[4] Scholes et al. (2005) offer a nontechnical review of some of the tax considerations in mergers and divestitures.
[5] See, for example, Auerbach and Reishus (1988), Dhaliwal et al. (2004), Erickson (1998), Erickson and Wang (1999a, 1999b, 2000, 2007), Hayn (1989), Kaplan (1989), Maydew et al. (1999), Schipper and Smith (1991), and Weaver (2000).

(2) $$\left(\pi_1+\pi_2\right)\left\{\tau_1\left[\sum_{i=1}^{n}\alpha_i \frac{x_{1i}}{\left(x_{1i}+x_{21}\right)}\right]+\tau_2\left[\sum_{i=1}^{n}\alpha_i \frac{x_{2i}}{\left(x_{1i}+x_{2i}\right)}\right]\right\}.$$

The difference between (1) and (2) is $\left(\tau_1-\tau_2\right)\psi$, for which:

(3) $$\psi=\sum_{i=1}^{n}\alpha_i \frac{\pi_2 x_{1i}-\pi_1 x_{2i}}{\left(x_{1i}+x_{2i}\right)}.$$

The difference in total tax obligation induced by formula apportionment is the product of tax rate differences and differences in ratios of profitability to factor use. Clearly, there is no difference if $\tau_1=\tau_2$ or if $\frac{\pi_2}{\pi_1}=\frac{x_{2i}}{x_{1i}}\forall i$. In the first of these cases the equality of tax rates implies that the taxpayer's total obligation is the same regardless of the jurisdiction to which income is assigned; in the second case, the formulas assign income exactly as it is earned. For most cases, however, neither of these conditions will hold. Introducing a new variable $\left(\frac{\bar{\pi}}{\bar{x}_i}\right)$, equal to the average value of the ratio of firm profits to inputs of factor *i*, equation (3) can be rewritten as:

(4) $$\psi=\sum_{i=1}^{n}\alpha_i\left[\pi_2-x_{2i}\left(\frac{\bar{\pi}}{\bar{x}_i}\right)\right]\frac{x_{1i}}{\left(x_{1i}+x_{2i}\right)}-\sum_{i=1}^{n}\alpha_i\left[\pi_1-x_{1i}\left(\frac{\bar{\pi}}{\bar{x}_i}\right)\right]\frac{x_{2i}}{\left(x_{1i}+x_{2i}\right)}.$$

In analyzing the implications of (4) it is helpful to define $s_{1i}\equiv\frac{x_{1i}}{\left(x_{1i}+x_{2i}\right)}$ to be firm 1's share of factor *i*, and $\bar{s}_1\equiv\sum_{i=1}^{n}\alpha_i\frac{x_{1i}}{\left(x_{1i}+x_{2i}\right)}=\sum_{i=1}^{n}\alpha_i s_{1i}$ to be firm one's average share of all factors, weighted as in the common allocation formula. Variables for firm two are similarly defined, so that $s_{2i}=1-s_{1i}$ and $\bar{s}_2=1-\bar{s}_1$. Then (4) can be rewritten as:

(5) $$\psi=\sum_{i=1}^{n}\alpha_i\left[\pi_2-x_{2i}\left(\frac{\bar{\pi}}{\bar{x}_i}\right)\right]\left[\left(s_{1i}-\bar{s}_1\right)+\bar{s}_1\right]-\sum_{i=1}^{n}\alpha_i\left[\pi_1-x_{1i}\left(\frac{\bar{\pi}}{\bar{x}_i}\right)\right]\left[\left(s_{2i}-\bar{s}_2\right)+\bar{s}_2\right].$$

Since $\left(s_{2i}-\bar{s}_2\right)=-\left(s_{1i}-\bar{s}_1\right)$, it follows that (5) implies:

$$6)\,\psi = \left[\pi_2 - \sum_{i=1}^{n} \alpha_i x_{2i} \left(\frac{\overline{\pi}}{\overline{x}_i}\right)\right]\overline{s}_1 - \left[\pi_1 - \sum_{i=1}^{n} \alpha_i x_{1i} \left(\frac{\overline{\pi}}{\overline{x}_i}\right)\right]\overline{s}_2 + \sum_{i=1}^{n} \alpha_i \left[(\pi_1 + \pi_2) - (x_{1i} + x_{2i})\left(\frac{\overline{\pi}}{\overline{x}_i}\right)\right](s_{1i} - \overline{s}_1)$$

The summation that is the third term on the right side of equation (6) can be simplified by noting that the definition of $\overline{s}_1$ implies that:

$$\sum_{i=1}^{n} \alpha_i \left[(\pi_1 + \pi_2)\right](s_{1i} - \overline{s}_1) = 0 .$$

Consequently, (6) implies:

$$(7) \qquad \psi = \left[\pi_2 - \sum_{i=1}^{n} \alpha_i x_{2i} \left(\frac{\overline{\pi}}{\overline{x}_i}\right)\right]\overline{s}_1 - \left[\pi_1 - \sum_{i=1}^{n} \alpha_i x_{1i} \left(\frac{\overline{\pi}}{\overline{x}_i}\right)\right]\overline{s}_2 - \sum_{i=1}^{n} \alpha_i \left[(x_{1i} + x_{2i})\left(\frac{\overline{\pi}}{\overline{x}_i}\right)\right](s_{1i} - \overline{s}_1),$$

which can be rewritten as:

$$(8) \qquad \psi = \left[\pi_2 - \sum_{i=1}^{n} \alpha_i x_{2i} \left(\frac{\overline{\pi}}{\overline{x}_i}\right)\right]\overline{s}_1 - \left[\pi_1 - \sum_{i=1}^{n} \alpha_i x_{1i} \left(\frac{\overline{\pi}}{\overline{x}_i}\right)\right]\overline{s}_2 - \sum_{i=1}^{n} \alpha_i \left(\frac{\overline{\pi}}{\overline{x}_i}\right)\left[x_{1i} - \overline{s}_1 (x_{1i} + x_{2i})\right].$$

The right side of equation (8) consists of three terms, the first of which is $\overline{s}_1$ times the difference between π_2 and the weighted average of $x_{2i}\left(\frac{\overline{\pi}}{\overline{x}_i}\right)$. One can think of this difference as being the residual from a regression equation in which the α_i s are regression coefficients used to predict π_2 based on the values of the x_{2i} s (interacted with the sample average values of the π/x_i s). The mean prediction error of this regression is clearly zero, since the α_i s sum to one, but the sum of squared residuals from this regression depends on the values of the α_i s. It is convenient to denote the value of the residual in this regression, the first bracketed term on the right side of equation (8), as r_1. Likewise the second term on the right side of equation (8) is the product of $\overline{s}_2$ and the analogous residual from the equation predicting π_1 based on the values of the x_{1i} s; hence this product can be represented as $r_1 \overline{s}_2$.

The third term on the right side of (8) is the weighted sum of the difference between x_{1i} and $\bar{s}_1(x_{1i} + x_{2i})$, with weights equal to $\alpha_i\left(\frac{\bar{\pi}}{\bar{x}_i}\right)$. This sum is a function not of the ability of factors to predict profitability, but instead of differences in the relative factor intensities of firms 1 and 2, since x_{1i} differs from $\bar{s}_1(x_{1i} + x_{2i})$ because $\bar{s}_1$ is an average taken across all input factors, not merely factor *i*. The value of this sum is likely to be of second order importance relative to the prediction errors that constitute the first two terms on the right side of equation (8), in particular equaling zero for the case in which a single factor receives unit weight in the allocation formula and all other factors have zero weight. Dismissing this term as largely inconsequential, equation (8) can be written as:

(9) $$\psi = r_2\bar{s}_1 - r_1\bar{s}_2 .$$

Hence the merger of two firms misallocates their income for tax purposes by an amount equal to the weighted difference of the residuals in the equations explaining their incomes.

3.2 The economic cost of income misattribution.

From the standpoint of efficient resource allocation, the cost of misattributing economic income by using an allocation formula includes the possibility that this misattribution may influence whether firms merge or divest their operations. A merger between firms 1 and 2 is tax favored if $(\tau_1 - \tau_2)\psi$ is positive, and tax disfavored if $(\tau_1 - \tau_2)\psi$ is negative; the magnitude of $(\tau_1 - \tau_2)\psi$ determines the extent of tax incentive. A conglomerate consisting of two affiliates equivalent to firms 1 and 2 similarly faces incentives to divest one affiliate if $(\tau_1 - \tau_2)\psi$ is negative, and to avoid divestiture if $(\tau_1 - \tau_2)\psi$ is positive. These tax incentives say nothing about the business merits of these mergers or divestitures, though as a general matter the existence of such tax incentives can be expected to reduce efficiency by introducing considerations other than pretax profits into ownership decisions.

In order to estimate the magnitude of the ownership distortion due to formula apportionment, it is necessary to understand the extent to which firms are likely to change their ownership of affiliates or other firms in response to these tax incentives. There is apt to be considerable variation, since influencing the decision requires that a firm be sufficiently close to the margin that tax considerations become decisive. Firm size is sure to be correlated with this proclivity: Exxon Mobil is unlikely to attempt to acquire British Petroleum in order to save $100 million in taxes, even though two medium sized firms would probably find a tax saving of that size an irresistible inducement to merge. Exxon Mobil might, on the other hand, consider acquiring a small oil company in return for a modest tax saving, since relative to the size of the acquisition the tax saving could loom large.

One way to formalize these notions is to specify that the probability that firms 1 and 2 merge is given by:

$$k + \gamma \frac{(\tau_1 - \tau_2)\psi}{(\tilde{x}_1 + \tilde{x}_2)\tilde{s}_1^2 \tilde{s}_2^2}, \tag{10}$$

in which *k* is a function of various attributes of firms 1 and 2, γ is a constant, $\tilde{x}_1$ is the size of firm 1's assets, and $\tilde{s}_1^2$ is the squared value of firm 1's share of the total assets of firms 1 and 2, $\tilde{s}_1^2 = \frac{\tilde{x}_1^2}{(\tilde{x}_1 + \tilde{x}_2)^2}$. The numerator of the ratio in (10) is the tax consequence of the merger, and the denominator captures that doubling the size of both the target and the acquirer also doubles the required tax saving to have the same effect on merger probabilities. This size effect is not symmetric: whereas increasing the size of the smaller of firms 1 and 2 indeed increases the tax benefit necessary to encourage a merger, increasing the size of the larger of the two firms may reduce the needed tax benefit. The derivative of the denominator of (10) with respect to $\tilde{x}_1$ is:

$$\frac{\tilde{x}_1 \tilde{x}_2^2 (2\tilde{x}_2 - \tilde{x}_1)}{(\tilde{x}_1 + \tilde{x}_2)^4}, \tag{11}$$

which is positive if $\tilde{x}_1$ is sufficiently small, but becomes negative if $\tilde{x}_1 > 2\tilde{x}_2$. This function has the sensible feature that required tax savings decline with the size of the larger of the two firms as that firm's size increases. Thinking of the much larger firm as the acquirer, as the acquirer grows in size it has greater access to financial and managerial resources that would come at higher cost to a small acquirer.

With a probability of merger given by (10), the standard Harberger triangle approximation (Hines, 1999) to the expected deadweight loss associated with tax incentives can be denoted Δ and is given by:

(12) $$\Delta = \frac{1}{2}\gamma \frac{(\tau_1 - \tau_2)^2 \psi^2}{(\tilde{x}_1 + \tilde{x}_2)\tilde{s}_1^2 \tilde{s}_2^2}.$$

The value of ψ^2 is given by:

(13) $$\psi^2 = r_2^2 \bar{s}_1^2 + r_1^2 \bar{s}_2^2 - 2 r_1 r_2 \bar{s}_1 \bar{s}_2 .$$

Taking firms 1 and 2 to be randomly matched, it follows from the fact that r_1 and r_2 both have mean zero that the expected value of ψ^2 is:

(14) $$E(\psi^2) = \bar{s}_1^2 E(r_2^2) + \bar{s}_2^2 E(r_1^2).$$

In evaluating (12) it simplifies matters greatly to set $\bar{s}_1^2 = \hat{s}_1^2$, reflecting that the ratio of two firms' assets roughly tracks the ratios of their factors used in the allocation formulas. Furthermore, the expected value of the squared prediction error is itself a function of firm size, so:

(15) $$\frac{E(r_1^2)}{E(r_2^2)} = \frac{\hat{s}_1^2}{\hat{s}_2^2}.$$

Imposing (14) and (15), and the approximation that $\bar{s}_1^2 = \hat{s}_1^2$, (12) becomes:

(16) $$\Delta = \gamma \frac{(\tau_1 - \tau_2)^2 E(r_1^2)}{(\tilde{x}_1 + \tilde{x}_2)\tilde{s}_1^2} = \gamma \frac{(\tau_1 - \tau_2)^2 E(r_1^2)(\tilde{x}_1 + \tilde{x}_2)}{\tilde{x}_1^2},$$

which in turn implies that:

(17) $$\Delta = \gamma(\tau_1 - \tau_2)^2 \left[\frac{E(r_1^2)}{\tilde{x}_1} + \frac{E(r_2^2)}{\tilde{x}_2}\right].$$

Equation (17) indicates that the deadweight loss from income misattribution by formulary methods is proportional to the product of the square of tax rate differences and sum of squared estimation errors, normalized by asset sizes. Clearly, the formulary system that minimizes the deadweight loss associated with income misattribution is one that minimizes the weighted sum of squares of prediction errors, with weights equal to firm assets.

3.3. Interpretation.

Equation (17) implies that the expected economic cost of ownership distortions introduced by misattributing income is proportional to the product of the square of tax rate differences and the expected squared prediction error from using a formula. This comes from the Harberger triangle representation of deadweight loss, which takes a second-order approximation that the distribution of tax benefits is roughly uniform across the affected population, so for those firms whose merger or divestment decisions are influenced by the use of formula allocation, the average economic cost of this distortion equals half of the tax incentive produced by the formula. This, together with the equation determining the likelihood of tax-motivated ownership changes, implies that small tax incentives produce very small expected deadweight losses, whereas large tax benefits for some firms may create significant deadweight losses.

The analysis in section 3.2 considers a potential merger of two firms chosen at random, and it is the randomness of this matching that makes it possible to ignore the third term on the right side equation (13), the expectation of which is zero, and thereby replace equation (13) with equation (14). Clearly, potential merger candidates are not in fact randomly matched, though whether their matching is random relative to potential tax benefits introduced by the use of formula apportionment is another matter. The analysis considers individual firm matches, but once a system of formula apportionment is in place the distribution of firm attributes will change as assets are reallocated in response to tax incentives. As Gordon and Wilson (1986) note, in the absence of merger costs every firm will face the same average tax rate in equilibrium, since any tax differences will be eliminated through the process of tax-motivated mergers and divestments. It is, however, unrealistic to think that asset reallocation can proceed so costlessly or easily. The framework underlying the empirical analysis assumes that potential merger partners are brought together for non-tax business reasons, at which point tax considerations have the potential to influence the outcome by affecting the potential net benefits of a merger. Furthermore, tax-motivated mergers are either sufficiently limited in number that they do not significantly influence the distribution of firm attributes throughout the economy, so these attributes can be taken to be exogenous from the standpoint of the analysis, or else they are quite frequent, in which case it is clear that they are the source of large economic distortions.

Both tax and prediction error terms enter the formula for efficiency cost in Equation (17). Assuming that tax rates are determined by considerations that do not include the details of formula apportionment, and taking firm attributes to be distributed independently of tax rates, it follows that the formula that minimizes the expected squared prediction error also minimizes deadweight loss. This is the ordinary least squares estimate of the formula components: OLS is the minimum variance unbiased linear estimator, and by construction in this case the mean estimation error is zero and the estimates are linear in the components. Hence in order to implement equation (17) to find the distortion-minimizing formulary apportionment scheme, it is simply necessary to run regressions explaining profitability on the basis of observed factors.

4. Data

In order to evaluate the magnitudes of tax-induced ownership distortions it is necessary to estimate the extent to which apportionment factors explain the variation in firm profitability. For this purpose, and in the interest of data comparability and data quality, it is helpful to consider evidence from firms located in a single country. The Compustat database culls information from annual reports and 10-K filings for large firms that are publicly listed in the United States; the extensive research use and refinement of these data, as well as restricting the sample to relatively large American companies, make the entries perhaps more reliable than some alternatives. This study uses data from firms listed in the primary Compustat file for 2004, which starts with roughly 7,200 firms, and from this sample uses data for active corporations for which there are data reported on operating income, income taxes, market capitalization at yearend, sales, value of property, plant and equipment, assets, and numbers of employees. All figures are consolidated for groups that include controlled subsidiaries, and include information for worldwide operations.

The Compustat information includes several variables that can be used to infer a firm's taxable income. Perhaps the most obvious is a firm's income taxes, since the U.S. system of taxing worldwide incomes while providing foreign tax credits implies that a large American corporation will owe roughly 35% of its worldwide income in taxes. Certainly there is plenty of potential slippage, since income for tax purposes is not the same thing as income for financial reporting purposes; there are tax credits, income that is exempt from taxation, special deductions, loss carryforwards and carrybacks, and other considerations, but as a general matter the ratio of a firm's tax liability to 0.35 offers a good guide to its annual taxable income. Another widely used measure of annual income is operating income after depreciation. Finally, annual income itself is a rather noisy measure of what firms earn, since it is subject to well-known sources of intertemporal fluctuations. Market capitalization offers a more permanent, and for some purposes perhaps more reliable, measure of a firm's long-term profit potential, though seemingly random fluctuations in this variable are widely documented.

Of the active firms in the 2004 Compustat sample, 6,992 reported information on operating income, sales, property, plant and equipment, total assets, and employment. Of these, 6,986 reported information on income taxes, and 6,384 reported information on yearend market capitalization. Allocation formulas that include labor-based measures typically use labor compensation rather than numbers of employees, but labor compensation is much less commonly reported in annual reports and 10-K filings. Only 1,575 firms in the 2004 Compustat sample reported information on operating income, sales, property, plant and equipment, total assets, and labor compensation, with 1,574 of these reporting tax payments and 1,461 reporting market capitalization. Hence including the labor compensation fields reduces the sample size by more than 75%. Consequently, the results that follow are reported for two different samples: the smaller sample of firms that report labor compensation, and the larger sample of firms reporting numbers of employees.

Table 1 presents descriptive statistics for these variables;[6] the top panel reports the figures for the sample reporting labor compensation, and the bottom panel for the much larger sample reporting total employment. The skewed distribution of firm sizes is evident from this

---

[6] Appendix Table 1 presents descriptive statistics for versions of these variables weighted by the square root of firm assets; these are the variables used in the regressions that follow.

table, in that mean entries for every variable greatly exceed sample medians. In the sample of firms reporting employment compensation, median operating income for 2004 was $22.1 million, median taxable income (inferred from income taxes) was $10.6 million, and median market capitalization was $257 million. These firms had median sales of $103.2 million, median property, plant and equipment holdings of $20.4 million, and median labor compensation of $22.6 million. The larger sample of firms reporting total employment had smaller median operating incomes ($12.0 million) and taxable incomes ($4.0 million), and somewhat smaller median market capitalization ($220.2 million), though slightly larger sales and property, plant and equipment holdings, and a median of 550 employees per firm. Hence firms reporting labor compensation tend to be more profitable relative to observable factors than the group of firms reporting numbers of employees.

The empirical work in section 5 uses these data to estimate the extent to which observable factors explain income differences. Since the Compustat data are financial accounting entries, they need not correspond to taxable incomes, and are potentially subject to their own sources of bias and noise, as executives may manage entries to meet earning targets and otherwise present their firms to financial markets as favorably as possible. Whether these financial data are more or less reliably reported than information presented to tax authorities is an interesting question, but in any case tax data are confidential and therefore unavailable for this analysis, and it is revealing to use the financial data to compare the accuracy of alternative methods of predicting a firm's income.

The available information covers the operations and incomes of independent publicly listed companies. While this offers many advantages from the standpoint of data availability and independence, the analysis in section 3 concerns separate operations within the same multinational firms, rather than operations of disparate companies. In using data from separate companies to estimate the consequences of allocating income using formulary methods, there is an implicit assumption that the income generating process in a cross section of firms resembles the income generating process among affiliates of the same firm. Certainly this assumption looks sound if mergers are tax-driven marriages of convenience; and even in the arguably more common case that ownership of divisions or affiliates is determined pretax profitability, with production, sales, and cost spillovers among components of the firm, it is nonetheless reasonable to expect that to a first order the contribution of an affiliate to total firm profitability is closely related to its profitability as a stand-alone entity.

5. Determinants of Corporate Income

In evaluating the predictive quality of different allocation formulas, it is helpful to start with benchmark formulas whose coefficients are unconstrained by prior choices but instead determined by the data. These benchmark formulas come simply from regressions of income on factors that enter the formulas. These regressions have the potential to reveal the ability of different factors to account for income production, and to help evaluate the extent to which simple formulary rules depart from weights that are indicated by the data.

Table 2 presents regressions of each of the three income variables on measures of sales, tangible property, and employment. The regressions reported in columns 1, 3, and 5 use labor compensation as measures of labor input, whereas the regressions reported in columns 2, 4, and 6 use numbers of employees; as a result, the sample sizes are much larger in the even-numbered columns. In all of the regressions variables are weighted by the square root of firm assets. Observable measures do a creditable job of predicting operating income, in that the R-squareds slightly exceed 0.5, though this largely reflects differences between small firms and large firms. The point estimates in column one imply that, controlling for property and employment, $100 of additional sales (interacted with the ratio of mean income to mean sales) is associated with $7.55 of additional operating income. Given the values reported in Appendix Table 1, this implies that $100 of additional sales, conditional on property and employment, correlates with $1.02 of additional income. Similarly, controlling for other factors, the regression implies that $100 of additional property (interacted with the ratio of mean income to mean property) is associated with $6.57 of additional operating income – or that $100 of additional property, conditional on sales and employment, correlates with $1.96 of additional income. The estimated coefficients on sales and tangible property are both significant, though it is notable that the t-statistic on the sales coefficient exceeds that on the property coefficient. The 0.0064 point estimate on employment implies that $100 of additional labor compensation (interacted with the ratio of mean income to mean labor compensation) is associated with less than $1 of additional income, and this magnitude does not differ statistically from zero. The regression reported in column two using the much larger sample produces similar estimates for sales and property effects on income, but this time the estimated employment coefficient is negative (albeit statistically indistinguishable from zero).

Columns 3 and 4 report coefficient estimates obtained by repeating these regressions with the measure of income implied by income tax payments as the dependent variable. Sales and property again significantly affect measured income, whereas the estimated coefficients on the employment variables are negative and insignificant. Columns 5 and 6 report coefficients from equations explaining market capitalization, for which the estimated effects of sales and tangible property are very similar to the estimates appearing in columns 1 and 2. The employment variables, however, seem to perform significantly better in these regressions, with a positive and statistically significant coefficient on labor compensation in the regression reported in column five. It is noteworthy that the estimated coefficients on labor compensation, tangible property, and sales in column five are all roughly similar in magnitude (they range from 0.5801 to 0.6699), suggesting that an equally weighted three-factor formula may be a reasonable predictor of market capitalization. In the larger sample, however, the coefficient on numbers of employees is not statistically significant, as reported in column six.

Figure 1 plots predicted and actual values of taxable income for the smaller sample that includes data on labor compensation. It is clear from the figure that while predicted values capture the central tendencies of the data, there is just an enormous amount of idiosyncratic variation that no three-factor formula can hope to reflect. Figure 2 performs the same exercise for the equation predicting market capitalization, a dependent variable that is truncated at zero, but that nevertheless exhibits considerable unexplained variation. Operating income exhibits even more variation than does taxable income and market capitalization, which again is impossible to forecast with just three variables. Of course, constraining the formulas by omitting one or two of the factors, or imposing equality among the coefficients on all three, only reduces the predictive power of the equations.

Tables 3-5 present estimated coefficients from regressions that include two variables at a time, corresponding to unconstrained two-factor formulas. Table 3 presents regressions that include sales and tangible property. The coefficients in this table resemble those presented in

Table 2 for the three-factor formulas, suggesting the unimportance of the labor variables to the equations. The only notable difference between the coefficients in Table 3 and those in Table 2 is that the sales coefficient in the market capitalization equation presented in column 5 of Table 3 is somewhat larger than the coefficient in the corresponding equation in column 5 of Table 2, reflecting that the sales variable here picks up the employment effects that appeared in the Table 2 equation.

Table 4 presents regressions that include sales and employment factors, in which the results indicate that the sales variable does almost all of the work in explaining measured income. The employment coefficients are insignificant (and frequently negative) in all of the equations except that reported in column five explaining market capitalization, where the 0.6439 coefficient on labor compensation differs significantly from zero, though its t-statistic of 2.74 is quite a bit smaller than the corresponding t-statistic of 9.65 on the sales coefficient. Table 5 presents estimated coefficients that repeat this analysis using property and employment as the two factors explaining income. The R-squareds of these regressions are quite a bit lower than are the comparable R-squareds reported in Tables 3 and 4, reflecting the loss of the significant explanatory power of the sales variable. The value of tangible property has a significant positive effect on measured income in all of the regressions, as does labor compensation, though again with t-statistics that are quite a bit smaller (e.g., 2.54 in column one) than the corresponding t-statistics for the coefficients on tangible property (10.23 in column one). The estimated coefficients on numbers of employees in columns 2, 4, and 6 never differ significantly from zero, but, for the first time, all of the point estimates are positive.

In practice, formulary methods as used by U.S. states and Canadian provinces, and advocated for use internationally, do not correspond to the unconstrained regressions presented in Tables 2-5. Instead these formulas follow fixed rules, such as equal weights on each of the sales, property and employment factors, or double weighting sales in a three-factor formula, or two factor formulas with equal weights on property and employment. Tables 6 and 7 present summary statistics from regressions based on these and other formulary alternatives. Table 6 presents evidence from the sample of firms reporting labor compensation, while Table 7 presents evidence from the larger sample of firms reporting total employment. The tables report mean squared prediction errors from the underlying regressions in each row. Mean squared errors are sample sums of squareds adjusted for degrees of freedom, which is appropriate in considering the predictive power of alternative formulas with differing numbers of regressors. The first row of each table presents data for the unconstrained three-factor regression, and all of the following rows report the extent to which their mean sequared errors exceed that of the unconstrained regression.

The figures in Table 6 illustrate the extent to which constraining the regressions reduces their predictive power. The equal weighted three-factor formula produces a mean squared error that is 19.4% higher than that for the unconstrained formula predicting operating income, 30.5% higher in predicting taxable income, and 8.5% higher in predicting market capitalization. Use of a three-factor formula with double weight on the sales factor significantly reduces these mean squared errors, and better still are the two-factor regressions in which labor compensation does not appear. Even with sales and property constrained each to have 50% weight, these two-factor formulas come within 4.2% of the mean squared error of the unconstrained three-factor formula predicting operating income, 8.7% of the formula predicting taxable income, and 7.7% of the formula predicting market capitalization.

The two-factor formulas that include labor compensation and are constrained to use equal weights perform very poorly from a mean squared error standpoint, the closest fit coming in the equation in which property and wages predict market capitalization with a mean squared error only 24% higher than that for the unconstrained three-factor formula. The culprit is clearly the labor compensation variable, as revealed by the one-factor formulas reported in the bottom three rows of Table 6. Whereas the use of the sales factor alone produces means squared errors that exceed by never more than 7.05% the errors from the unconstrained three-factor formula, the closest that the labor compensation factor comes to the three-factor mean squared error is 62.6% in the market capitalization equation. The property factor produces results between these two extremes, though closer to the sales factor. Similar results appear in Table 7 for regressions using the much larger sample and employment in place of labor compensation.

One indicator of the performance of a formula is the distribution of its predictions relative to actual values. Table 8 presents the distribution of the ratio of forecast errors to forecasted incomes for the unconstrained regressions. In the smaller samples that include labor compensation, the absolute value of the prediction error exceeds the predicted value of operating income 68.6% of the time, and the magnitude of the prediction error exceeds three times predicted operating income 31.8% of the time. Similarly, the absolute value of the prediction error exceeds the predicted level of taxable income 47.0% of the time and exceeds three times predicted taxable income 12.6% of the time; and the prediction error exceeds market capitalization 61.5% of the time, and exceeds three times market capitalization 22.8% of the time.

The constrained regressions perform considerably worse on average than do the unconstrained three-factor formulas. Table 9 presents the distribution of forecast errors for the three-factor formulas in which sales, property and employment get equal weights, and Table 10 presents forecast errors for the three-factor formulas that double weight sales. Table 11 presents forecast errors from the two-factor formula constrained to weight property and labor equally. The top panels of tables 9-11 present the distributions of prediction errors, and the bottom panels compare these distributions to the distribution for the unconstrained three-factor formula. The absolute value of the prediction error from the equal-weighted three-factor formula exceeds predicted operating income 70.6% of the time, and exceeds three times predicted operating income 36.6% of the time. These prediction errors do not differ greatly from those reported for the unconstrained regressions in Table 8, and the differences between the prediction errors of the unconstrained and constrained three-factor regressions are similarly modest for equations predicting taxable income and market capitalization. The double-sales-weighted constrained three-factor regression even boasts smaller percentage prediction errors at these points in the distribution than does the unconstrained three-factor regression, but matters are again very much different for the two-factor equations that omit sales and include labor, as Table 11 indicates. In these equations, the absolute value of the prediction error exceeds five times predicted operating income 26.1% of the time, exceeds five times predicted taxable income 13.5% of the time, and exceeds five times predicted market capitalization 19.5% of the time.

6. Conclusion

Formulary alternatives to separate entity accounting hold the undeniable appeal of reducing certain opportunities for tax-motivated international income reallocation. This comes at a serious cost, which is that the factors that enter the formulas do not accurately correspond to the determinants of business incomes. As a result, the formulas misattribute income, so their use in an international setting would misallocate tax revenue among countries and create incentives for new forms of tax avoidance through mergers and divestitures.

Evidence from large American corporations indicates that commonly proposed formulas predict only about half of the observed variation in corporate incomes. In particular, the labor cost factors do a very poor job of predicting income, which may not be surprising given that labor expenses are deductible in calculating taxable income. The analysis in section 3 of the paper shows that the ownership distortions associated with the use of formulary methods are proportional to the mean squared prediction error in a regression explaining firm income, with weights equal to the inverse square root of firm size. Since these regressions are very unkind to the labor factors, they suggest that governments that must use formulas maximize the accuracy of income attribution and minimize the deadweight loss of ownership distortions by ignoring or significantly downplaying the labor input factors.

Is it sensible to consider formulary alternatives to separate accounting for tax purposes, given the inaccuracy of formulary methods and the incentives they create? Evaluating this question requires a careful comparative examination of all of the unappetizing tax choices that governments face. Hard experience makes problems more evident in the tax systems that governments use than in the alternatives, but it does not follow that tax reform improves matters, since it generally replaces one set of problems with another. It is clear from the evidence that formulas attribute income very imperfectly, so whether the associated costs are acceptable depends on how dire one considers the international tax regime today.

References

Anand, Bharat N. and Richard Sansing, The weighting game: Formula apportionment as an instrument of public policy, *National Tax Journal*, June 2000, 53 (2), 183-200.

Auerbach, Alan J. and David Reishus, The effects of taxation on the merger decision, in Alan J. Auerbach ed., *Corporate takeovers: Causes and consequences* (Chicago: University of Chicago Press, 1988), 157-183.

Clausing, Kimberly A., The impact of transfer pricing on intrafirm trade, in James R. Hines Jr., ed. *International taxation and multinational activity* (Chicago: University of Chicago Press, 2001), 173-194.

Clausing, Kimberly A., Tax-motivated transfer pricing and U.S. intrafirm trade policies, *Journal of Public Economics*, September 2003, 87 (9-10), 2207-2223.

Commission of the European Communities, Towards an internal market without tax obstacles: A strategy for providing companies with a consolidated corporate tax base for their EU-wide activities, 2001, COM 582 final.

Desai, Mihir A., C. Fritz Foley and James R. Hines Jr., Chains of ownership, tax competition, and the location decisions of multinational firms, in Heinz Herrmann and Robert Lipsey, eds. *Foreign direct investment in the real and financial sector of industrial countries* (Berlin: Springer-Verlag, 2003), 61-98.

Devereux, Michael P. and Simon Loretz, The effects of EU formula apportionment on corporate tax revenues, *Fiscal Studies*, 2008, 29 (1), 1-33.

Dhaliwal, Dan S., Merle Erickson and Shane Heitzman, The effect of seller income taxes on acquisition price: Evidence from purchases of taxable and tax-exempt hospitals, *Journal of the American Taxation Association*, 2004, 26 (2), 1-21.

Fuest, Clemens, Thomas Hemmelgarn and Fred Ramb, How would the introduction of an EU-wide formula apportionment affect the distribution and size of the corporate tax base? An analysis based on German multinationals, *International Tax and Public Finance*, 2007, 14 (5), 605-626.

Edmiston, Kelly D., Strategic apportionment of the state corporate income tax: An applied general equilibrium analysis, *National Tax Journal*, June 2002, 55 (2), 239-262.

Erickson, Merle, The effects of taxes on the structure of corporate acquisitions, *Journal of Accounting Research*, Autumn 1998, 36 (2), 279-298.

Erickson, Merle and Shiing-wu Wang, Exploiting and sharing tax benefits: Seagram and DuPont, *Journal of the American Taxation Association*, Fall 1999a, 21 (2), 35-54.

Erickson, Merle and Shiing-wu Wang, Earnings management by acquiring firms in stock for stock mergers, *Journal of Accounting and Economics*, 1999b, 27, 149-176.

Erickson, Merle and Shiing-wu Wang, The effect of transaction structure on price: Evidence from subsidiary sales, *Journal of Accounting and Economics*, 2000, 30, 59-97.

Erickson, Merle and Shiing-wu Wang, Tax benefits as a source of merger premiums in acquisitions of private corporations, *The Accounting Review*, 2007, 82 (2), 359-387.

Goolsbee, Austan and Edward Maydew, Coveting thy neighbor's manufacturing: The dilemma of state income apportionment, *Journal of Public Economics*, 2000, 75 (1), 125-143.

Gordon, Roger H. and John D. Wilson, An examination of multijurisdictional corporate income taxation under formula apportionment, *Econometrica*, November 1986, 54 (6), 1357-1373.

Gresik, Thomas A., The taxing task of taxing transnationals, *Journal of Economic Literature*, September 2001, 39 (3), 800-838.

Gresik, Thomas A., Optimal separate accounting vs. optimal formula apportionment, Working Paper, University of Notre Dame, 2008.

Gresik, Thomas A. and Petter Osmundsen, Transfer pricing in vertically integrated industries, Working Paper, University of Notre Dame, 2006.

Hayn, Carla, Tax attributes as determinants of shareholder gains in corporate acquisitions, *Journal of Financial Economics*, 1989, 23, 121-153.

Hines, James R., Jr., Three sides of Harberger triangles, *Journal of Economic Perspectives*, Spring 1999, 13 (2), 167-188.

Hines, James R., Jr. and Eric M. Rice, Fiscal paradise: Foreign tax havens and American business, *Quarterly Journal of Economics*, February 1994, 109 (1), 149-182.

Huizinga, Harry and Luc Laeven, International profit shifting within multinationals: A multi-country perspective, CEPR Discussion Paper 6048, 2007.

Kaplan, Steven, The effects of management buyouts on operating performance and value, *Journal of Financial Economics*, 1989, 24, 217-254.

Martens-Weiner, Joann *Company tax reform in the European Union: Guidance from the United States and Canada on implementing formulary apportionment in the EU* (New York: Springer, 2006).

Maydew, Edward L., Katherine Schipper and Linda Vincent, The impact of taxes on the choice of divestiture method, *Journal of Accounting and Economics*, 1999, 28, 117-150.

Mintz, Jack, Corporate tax harmonization in Europe: It's all about compliance, *International Tax and Public Finance*, 2004, 11, 221-234 .

Mintz, Jack and Michael Smart, Income shifting, investment, and tax competition: Evidence from provincial taxation in Canada, *Journal of Public Economics*, 2004, 88, 1149-1168.

Mintz, Jack and Joann M. Weiner, Exploring formula allocation for the European Union, *International Tax and Public Finance*, November 2003, 10 (6), 695-711.

Riedel, Nadine and Marco Runkel, Company tax reform with a water's edge, *Journal of Public Economics*, 2007, 91, 1533-1554.

Roin, Julie, Can the income tax be saved? The promise and pitfalls of unitary formulary apportionment, University of Chicago Law School Public Law Working Paper No. 170, 2007.

Runkel, Marco and Guttorm Schjelderup, The choice of apportionment factors under formula apportionment, CESifo Working Paper No. 2072, 2007.

Schipper, Katherine and Abbie Smith, Effects of management buyouts on corporate interest and depreciation tax deductions, *Journal of Law and Economics*, October 1991, 34 (2), 295-341.

Scholes, Myron S., Mark A. Wolfson, Merle Erickson, Edward L. Maydew, and Terry Shevlin, *Taxes and business strategy: A planning approach* (Upper Saddle River, NJ: Prentice Hall, 2005).

Shackelford, Douglas and Joel Slemrod, The revenue consequences of using formula apportionment to calculate U.S. and foreign-source income: A firm-level analysis, *International Tax and Public Finance*, 1998, 5, 41-59.

Weaver, Connie D., Divestiture structure and tax attributes: Evidence from the Omnibus Budget Reconciliation Act of 1993, *Journal of the American Taxation Association*, 2000, 22 (Supplement), 54-71.

Wellisch, Dietmar, Taxation under formula apportionment: Tax competition, tax incidence, and the choice of apportionment factors, *FinanzArchiv*, 2004, 60 (1), 24-41.

Figure 1

Errors in Unconstrained Equations Predicting Taxable Income

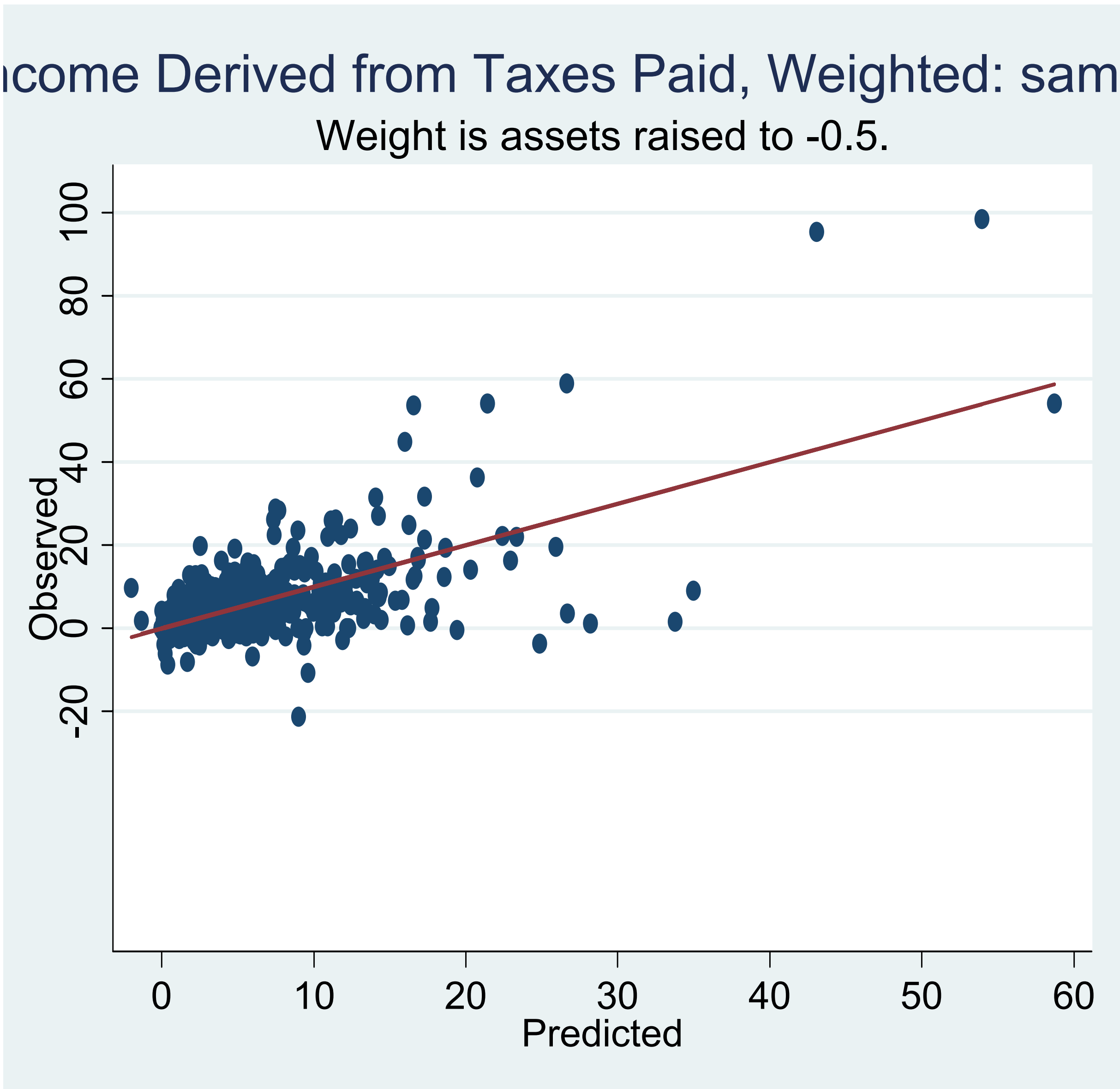

Figure 2

Errors in Unconstrained Equations Predicting Market Capitalization

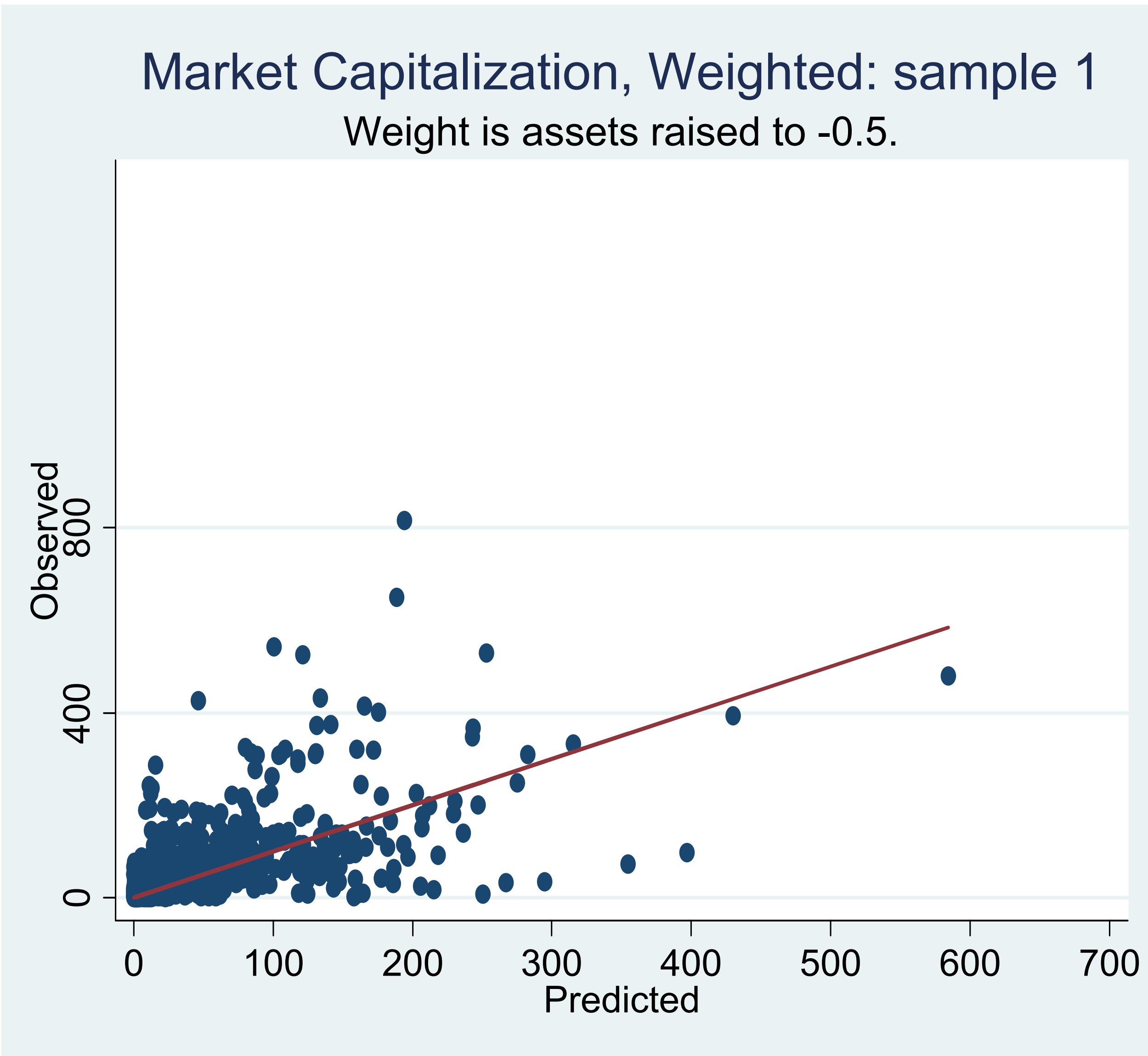

Table 1
Descriptive Statistics - Unweighted

Notes: all variables in 1000s.

Labor Compensation Samples	Mean	Median	Standard Deviation	No. Obs.
Operating Income	728,875	22,131	2,731,153	1575
Taxable Income as Reflected in Taxes Paid	457,937	10,631	2,052,976	1574
Market Capitalization	5,659,048	257,046	18,300,000	1461
Sales	4,316,875	103,182	16,900,000	1574
Property, Plant and Equipment	1,767,868	20,372	6,870,991	1574
Labor Compensation	824,823	22,557	3,998,434	1574

Employment Samples	Mean	Median	Standard Deviation	No. Obs.
Operating Income	325,538	12,025	1,614,135	6992
Taxable Income as Reflected in Taxes Paid	235,164	3,971	1,388,442	6986
Market Capitalization	3,170,152	220,243	13,500,000	6384
Sales	2,466,091	141,511	11,500,000	6986
Property, Plant and Equipment	1,041,252	22,609	4,833,397	6986
Employment	8.414	0.550	46.249	6986

Table 2
Determinants of Profits, Three-Factor, WLS Results

Notes: The Table presents estimated coefficients from regressions explaining 2004 profits as functions of contemporaneous sales, capital, and labor inputs. All equations are estimated using weighted least squares, with weights equal to one divided by the square root of contemporaneous firm assets. Robust standard errors are presented in parentheses.

Dependent Variable:	Operating Income		Taxable Income as Reflected in Taxes Paid		Market Capitalization	
	(1)	(2)	(3)	(4)	(5)	(6)
Sales	0.0755	0.0580	0.0765	0.0561	0.6699	0.7292
	(0.0108)	(0.0038)	(0.0153)	(0.0049)	(0.1157)	(0.0534)
Property, Plant and Equipment	0.0657	0.0896	0.0403	0.0638	0.5801	0.6960
	(0.0167)	(0.0065)	(0.0160)	(0.0078)	(0.1501)	(0.0854)
Labor Compensation	0.0064		-0.0194		0.6386	
	(0.0141)		(0.0231)		(0.2215)	
Employment		-57.88		-181.05		1060.14
		(103.35)		(150.47)		(1023.32)
No. of Obs.	1,575	6,992	1,574	6,986	1,461	6,384
R-Squared	0.5425	0.5427	0.5715	0.5004	0.5281	0.4049

Table 3
Determinants of Profits, Two-Factor, WLS Results

Notes: The Table presents estimated coefficients from regressions explaining 2004 profits as functions of contemporaneous sales and capital. All equations are estimated using weighted least squares, with weights equal to one divided by the square root of contemporaneous firm assets. Robust standard errors are presented in parentheses.

Dependent Variable:	Operating Income		Taxable Income as Reflected in Taxes Paid		Market Capitalization	
	(1)	(2)	(3)	(4)	(5)	(6)
Sales	0.0766	0.0578	0.0735	0.0554	0.7769	0.7333
	(0.0099)	(0.0037)	(0.0132)	(0.0047)	(0.1174)	(0.0521)
Property, Plant and Equipment	0.0658	0.0896	0.0400	0.0639	0.5835	0.6957
	(0.0167)	(0.0065)	(0.0160)	(0.0077)	(0.1531)	(0.0856)
No. of Obs.	1,575	6,992	1,574	6,986	1,461	6,384
R-Squared	0.5424	0.5426	0.5698	0.4999	0.5142	0.4048

Table 4
Determinants of Profits, Two-Factor, WLS Results

Notes: The Table presents estimated coefficients from regressions explaining 2004 profits as functions of contemporaneous sales and labor inputs. All equations are estimated using weighted least squares, with weights equal to one divided by the square root of contemporaneous firm assets. Robust standard errors are presented in parentheses.

Dependent Variable:	Operating Income		Taxable Income as Reflected in Taxes Paid		Market Capitalization	
	(1)	(2)	(3)	(4)	(5)	(6)
Sales	0.1003	0.0838	0.0917	0.0745	0.8908	0.9232
	(0.0090)	(0.0039)	(0.0125)	(0.0051)	(0.0923)	(0.0441)
Labor Compensation	0.0078		-0.0185		0.6439	
	(0.0151)		(0.0229)		(0.2350)	
Employment		-79.71		-196.60		973.14
		(128.22)		(165.59)		(1120.19)
No. of Obs.	1,575	6,992	1,574	6,986	1,461	6,384
R-Squared	0.5174	0.4803	0.5581	0.4615	0.5083	0.3835

Table 5
Determinants of Profits, Two-Factor, WLS Results

Notes: The Table presents estimated coefficients from regressions explaining 2004 profits as functions of contemporaneous capital and labor inputs. All equations are estimated using weighted least squares, with weights equal to one divided by the square root of contemporaneous firm assets. Robust standard errors are presented in parentheses.

Dependent Variable:	Operating Income		Taxable Income as Reflected in Taxes Paid		Market Capitalization	
	(1)	(2)	(3)	(4)	(5)	(6)
Property, Plant and Equipment	0.1586	0.1714	0.1344	0.1431	1.3922	1.8011
	(0.0155)	(0.0073)	(0.0205)	(0.0105)	(0.1398)	(0.0892)
Labor Compensation	0.0776		0.0527		1.2573	
	(0.0306)		(0.0228)		(0.3943)	
Employment		599.40		455.29		8629.72
		(383.88)		(326.03)		(5923.59)
No. of Obs.	1,575	6,992	1,574	6,986	1,461	6,384
R-Squared	0.4342	0.4148	0.4145	0.3531	0.4439	0.2770

Table 6
Income Prediction Accuracy among Formulas with Labor Compensation

Notes: The Table presents mean square errors from regressions explaining 2004 profits as functions of contemporaneous sales, capital, and labor costs. All equations are estimated using weighted least squares, with weights equal to one divided by the square root of contemporaneous firm assets.

Dependent Variable:	Operating Income			Taxable Income Reflected in Taxes Paid			Market Capitalization		
	Obs.	MSE	% higher MSE	Obs.	MSE	% higher MSE	Obs.	MSE	% higher MSE
3-factor, unconstrained	1575	28,068,291	0.00%	1574	18,631,321	0.00%	1461	2.609E+09	0.00%
3-factor, equal weights	1575	33,499,441	19.35%	1574	24,315,983	30.51%	1461	2.830E+09	8.47%
3-factor, double sales weights	1575	31,173,686	11.06%	1574	22,162,656	18.95%	1461	2.688E+09	3.03%
2-factor (sales and PPE), unconstrained	1575	28,058,319	-0.04%	1574	18,692,620	0.33%	1461	2.684E+09	2.87%
2-factor (sales and PPE), constrained	1575	29,226,697	4.13%	1574	20,250,335	8.69%	1461	2.810E+09	7.70%
2-factor (sales and wages), unconstrained	1575	29,590,430	5.42%	1574	19,198,054	3.04%	1461	2.717E+09	4.14%
2-factor (sales and wages), constrained	1575	39,674,233	41.35%	1574	28,568,962	53.34%	1461	3.258E+09	24.88%
2-factor (PPE and wages), unconstrained	1575	34,695,905	23.61%	1574	25,438,003	36.53%	1461	3.073E+09	17.78%
2-factor (PPE and wages), constrained	1575	39,027,225	39.04%	1574	28,922,762	55.24%	1461	3.234E+09	23.96%
1-factor (sales)	1575	29,583,480	5.40%	1574	19,252,446	3.33%	1461	2.793E+09	7.05%
1-factor (PPE)	1575	36,048,545	28.43%	1574	26,057,458	39.86%	1461	3.422E+09	31.16%
1-factor (wages)	1575	51,415,773	83.18%	1574	37,441,279	100.96%	1461	4.243E+09	62.63%

Table 7
Income Prediction Accuracy among Formulas with Total Employment

Notes: The Table presents mean square errors from regressions explaining 2004 profits as functions of contemporaneous sales, capital, and number of employees. All equations are estimated using weighted least squares, with weights equal to one divided by the square root of contemporaneous firm assets.

Dependent Variable:	Operating Income			Taxable Income Reflected in Taxes Paid			Market Capitalization		
	Obs.	MSE	% higher MSE	Obs.	MSE	% higher MSE	Obs.	MSE	% higher MSE
3-factor, unconstrained	6992	9,993,415	0.00%	6986	8,887,307	0.00%	6384	1.838E+09	0.00%
3-factor, equal weights	6992	17,932,267	79.44%	6986	15,020,281	69.01%	6384	2.641E+09	43.69%
3-factor, double sales weights	6992	16,363,335	63.74%	6986	13,811,066	55.40%	6384	2.465E+09	34.11%
2-factor (sales and PPE), unconstrained	6992	9,992,911	-0.01%	6986	8,895,087	0.09%	6384	1.838E+09	0.00%
2-factor (sales and PPE), constrained	6992	10,100,910	1.08%	6986	9,117,071	2.59%	6384	1.889E+09	2.77%
2-factor (sales and employees), unconstrained	6992	11,354,185	13.62%	6986	9,577,684	7.77%	6384	1.903E+09	3.54%
2-factor (sales and employees), constrained	6992	19,713,998	97.27%	6986	16,266,271	83.03%	6384	2.822E+09	53.54%
2-factor (PPE and employees), unconstrained	6992	12,787,180	27.96%	6986	11,505,961	29.47%	6384	2.232E+09	21.44%
2-factor (PPE and employees), constrained	6992	19,712,897	97.26%	6986	16,355,192	84.03%	6384	2.841E+09	54.57%
1-factor (sales)	6992	11,354,315	13.62%	6986	9,586,987	7.87%	6384	1.903E+09	3.54%
1-factor (PPE)	6992	12,888,984	28.97%	6986	11,564,105	30.12%	6384	2.253E+09	22.58%
1-factor (employees)	6992	21,202,372	112.16%	6986	17,367,467	95.42%	6384	2.999E+09	63.17%

Table 8: Unconstrained - Three Factor

	Operating Income		***Taxable Income Derived from Taxes***		***Market Capitalization***	
Sample[1]	***1***	***2***	***1***	***2***	***1***	***2***
N (Sample)[2]	1575	6992	1574	6986	1461	6384
N (0-value for hat(π))[3, 4]	3	24	3	24	2	21
N (mean computation)[5]	1572	6968	1571	6962	1459	6363
mean[|π-hat(π)| / hat(π)]	56.43	7822.66	2.28	-25.16	10.60	150.95
median[|π-hat(π)| / hat(π)]	2.08	1.11	1.00	1.00	1.51	0.98
Percent of firms for which:						
|π-hat(π)| > 0.05 * hat(π)	98.47%	96.96%	97.77%	97.67%	98.42%	97.16%
|π-hat(π)| > 0.1 * hat(π)	96.25%	94.00%	95.86%	95.65%	96.44%	94.06%
|π-hat(π)| > 0.2 * hat(π)	92.88%	88.09%	91.79%	91.12%	92.53%	88.70%
|π-hat(π)| > 0.25 * hat(π)	90.39%	85.40%	89.88%	88.75%	90.13%	85.62%
|π-hat(π)| > 0.5 * hat(π)	82.06%	71.48%	79.95%	77.82%	80.40%	72.31%
|π-hat(π)| > 0.75 * hat(π)	74.43%	60.07%	70.02%	66.83%	71.01%	59.42%
|π-hat(π)| > 1 * hat(π)	68.64%	53.07%	46.98%	32.28%	61.48%	49.17%
|π-hat(π)| > 2 * hat(π)	51.46%	37.17%	24.76%	15.66%	38.93%	35.90%
|π-hat(π)| > 3 * hat(π)	31.81%	29.16%	12.60%	9.41%	22.76%	27.27%
|π-hat(π)| > 4 * hat(π)	18.00%	23.13%	7.70%	5.90%	14.32%	21.91%
|π-hat(π)| > 5 * hat(π)	12.21%	19.25%	4.26%	4.45%	10.90%	18.76%
|π-hat(π)| > 7.5 * hat(π)	8.08%	14.85%	2.23%	2.47%	6.99%	14.25%
|π-hat(π)| > 10 * hat(π)	7.25%	12.94%	2.04%	1.94%	5.48%	11.99%

Notes:

1. Sample *1* is the collection of firms for which the income variable, sales, net PPE, total assets, and **labor costs** are all nonmissing in the 2004 Compustat data. Sample *2* is the collection of firms for which the income variable, sales, net PPE, total assets, and **number of employees** are all nonmissing in the 2004 Compustat data.
2. WLS estimation performed on this sample for following specification: income variable on sales, net PPE, and labor costs (Sample *1*) OR number of employees (Sample *2*). The weight is total assets raised to -0.5. The constant term is omitted.
3. The income variable prediction, hat(π), is the dot product of the vector of WLS estimates and the vector of covariate values.
4. The income variable prediction, hat(π), is exactly 0 for firms for which the covariates are all exactly 0 (recall that the constant term is omitted).
5. The mean of the proportional absolute prediction error, |π-hat(π)| / hat(π), is computed for this smaller sample because of division by 0 for firms with predicted income of 0.

Table 9: Constrained - Three Factor: (1/3 sales, 1/3 PPE, 1/3 labor measure

	Operating Income		***Taxable Income Derived from Taxes***		***Market Capitalization***	
Sample[1]	***1***	***2***	***1***	***2***	***1***	***2***
N (Sample)[2]	1575	6992	1574	6986	1461	6384
N (0-value for hat(π))[3, 4]	3	24	3	24	2	21
N (mean computation)[5]	1572	6968	1571	6962	1459	6363
mean[|π-hat(π)| / hat(π)]	38.66	29.74	1.86	6.44	11.56	56.64
median[|π-hat(π)| / hat(π)]	2.24	3.51	1.00	1.12	1.68	4.01
Percent of firms for which:						
|π-hat(π)| > 0.05 * hat(π)	98.22%	98.84%	97.64%	98.88%	98.29%	98.52%
|π-hat(π)| > 0.1 * hat(π)	95.99%	97.69%	96.12%	97.85%	96.71%	97.27%
|π-hat(π)| > 0.2 * hat(π)	93.07%	95.09%	92.11%	95.50%	92.67%	94.80%
|π-hat(π)| > 0.25 * hat(π)	91.16%	94.19%	90.52%	94.30%	90.68%	93.56%
|π-hat(π)| > 0.5 * hat(π)	83.84%	88.81%	81.54%	88.75%	81.15%	87.03%
|π-hat(π)| > 0.75 * hat(π)	76.59%	83.50%	72.37%	82.36%	71.90%	81.06%
|π-hat(π)| > 1 * hat(π)	70.61%	78.72%	50.03%	52.53%	63.47%	75.47%
|π-hat(π)| > 2 * hat(π)	54.45%	63.94%	29.34%	39.04%	44.21%	64.23%
|π-hat(π)| > 3 * hat(π)	36.64%	53.66%	16.49%	31.21%	27.48%	56.26%
|π-hat(π)| > 4 * hat(π)	22.52%	46.73%	9.23%	25.61%	17.82%	50.04%
|π-hat(π)| > 5 * hat(π)	16.35%	41.39%	6.37%	21.26%	12.41%	44.49%
|π-hat(π)| > 7.5 * hat(π)	9.67%	32.65%	2.55%	13.88%	7.61%	33.73%
|π-hat(π)| > 10 * hat(π)	7.44%	25.52%	1.21%	9.16%	5.41%	26.17%
vs. unconstrained model[6]	*(percentage point difference)*					
|π-hat(π)| > 0.05 * hat(π)	-0.25%	1.88%	-0.13%	1.21%	-0.14%	1.37%
|π-hat(π)| > 0.1 * hat(π)	-0.25%	3.69%	0.25%	2.20%	0.27%	3.21%
|π-hat(π)| > 0.2 * hat(π)	0.19%	7.00%	0.32%	4.38%	0.14%	6.10%
|π-hat(π)| > 0.25 * hat(π)	0.76%	8.78%	0.64%	5.54%	0.55%	7.94%
|π-hat(π)| > 0.5 * hat(π)	1.78%	17.32%	1.59%	10.93%	0.75%	14.73%
|π-hat(π)| > 0.75 * hat(π)	2.16%	23.42%	2.36%	15.53%	0.89%	21.64%
|π-hat(π)| > 1 * hat(π)	1.97%	25.65%	3.06%	20.25%	1.99%	26.29%
|π-hat(π)| > 2 * hat(π)	2.99%	26.77%	4.58%	23.38%	5.28%	28.34%
|π-hat(π)| > 3 * hat(π)	4.83%	24.50%	3.88%	21.80%	4.73%	29.00%
|π-hat(π)| > 4 * hat(π)	4.52%	23.59%	1.53%	19.71%	3.50%	28.13%
|π-hat(π)| > 5 * hat(π)	4.13%	22.14%	2.10%	16.81%	1.51%	25.73%
|π-hat(π)| > 7.5 * hat(π)	1.59%	17.80%	0.32%	11.40%	0.62%	19.47%
|π-hat(π)| > 10 * hat(π)	0.19%	12.57%	-0.83%	7.22%	-0.07%	14.18%

Notes:

1. Sample *1* is the collection of firms for which the income variable, sales, net PPE, total assets, and **labor costs** are all nonmissing in the 2004 Compustat data. Sample *2* is the collection of firms for which the income variable, sales, net PPE, total assets, and **number of employees** are all nonmissing in the 2004 Compustat data.

2. WLS estimation performed on this sample for following specification: income variable on the "index" variable. For firm *i*, the index equals Σ α_j (*i*'s value for factor *j* / sample mean value for factor *j*), where α_j is the weight on factor *j* specified by the relevant formula. This is the only RHS variable. The weight is total assets raised to -0.5. The constant term is omitted.

3. The income variable prediction, hat(π), is the product of the WLS estimate for the slope on the index variable and the index variable value.

4. The income variable prediction, hat(π), is exactly 0 for firms for which the component factors of the index variable (sales, PPE, etc.) are all exactly 0 (recall that the constant term is omitted).

5. The mean of the proportional absolute prediction error, |π-hat(π)| / hat(π), is computed for this smaller sample because of division by 0 for firms with predicted income of 0.

6. Each cell reveals the additional percent of the sample for which the income variable prediction error is outside of the specified interval, for the constrained vs. the unconstrained model.

Table 10: Constrained - Three Factor: (1/2 sales, 1/4 PPE, 1/4 labor measure)

	Operating Income		***Taxable Income Derived from Taxes***		***Market Capitalization***	
Sample[1]	***1***	***2***	***1***	***2***	***1***	***2***
N (Sample)[2]	1575	6992	1574	6986	1461	6384
N (0-value for hat(π))[3, 4]	3	24	3	24	2	21
N (mean computation)[5]	1572	6968	1571	6962	1459	6363
mean[|π-hat(π)| / hat(π)]	41.63	24.55	1.62	5.04	10.27	39.72
median[|π-hat(π)| / hat(π)]	2.02	2.35	1.00	1.00	1.55	2.59
Percent of firms for which:						
|π-hat(π)| > 0.05 * hat(π)	97.84%	98.09%	98.15%	98.43%	98.08%	98.15%
|π-hat(π)| > 0.1 * hat(π)	96.31%	96.66%	95.93%	96.94%	96.98%	96.54%
|π-hat(π)| > 0.2 * hat(π)	92.56%	93.50%	91.73%	93.94%	92.25%	93.59%
|π-hat(π)| > 0.25 * hat(π)	90.90%	91.73%	89.62%	92.63%	89.72%	91.83%
|π-hat(π)| > 0.5 * hat(π)	83.14%	83.74%	80.52%	84.96%	80.40%	82.51%
|π-hat(π)| > 0.75 * hat(π)	75.06%	76.98%	70.27%	77.54%	70.32%	73.90%
|π-hat(π)| > 1 * hat(π)	68.38%	70.62%	46.47%	45.73%	61.00%	66.59%
|π-hat(π)| > 2 * hat(π)	50.38%	54.06%	24.00%	30.88%	39.48%	55.18%
|π-hat(π)| > 3 * hat(π)	31.23%	44.35%	12.22%	23.14%	23.58%	46.31%
|π-hat(π)| > 4 * hat(π)	18.45%	37.74%	6.87%	17.44%	14.80%	39.78%
|π-hat(π)| > 5 * hat(π)	13.04%	32.82%	3.95%	13.29%	11.10%	33.96%
|π-hat(π)| > 7.5 * hat(π)	8.40%	23.48%	1.15%	7.07%	6.79%	23.59%
|π-hat(π)| > 10 * hat(π)	6.87%	18.10%	0.64%	3.96%	5.00%	18.39%
vs. unconstrained model[6]	*(percentage point difference)*					
|π-hat(π)| > 0.05 * hat(π)	-0.64%	1.13%	0.38%	0.76%	-0.34%	0.99%
|π-hat(π)| > 0.1 * hat(π)	0.06%	2.65%	0.06%	1.29%	0.55%	2.48%
|π-hat(π)| > 0.2 * hat(π)	-0.32%	5.41%	-0.06%	2.82%	-0.27%	4.89%
|π-hat(π)| > 0.25 * hat(π)	0.51%	6.33%	-0.25%	3.88%	-0.41%	6.21%
|π-hat(π)| > 0.5 * hat(π)	1.08%	12.26%	0.57%	7.14%	0.00%	10.20%
|π-hat(π)| > 0.75 * hat(π)	0.64%	16.91%	0.25%	10.70%	-0.69%	14.47%
|π-hat(π)| > 1 * hat(π)	-0.25%	17.55%	-0.51%	13.46%	-0.48%	17.41%
|π-hat(π)| > 2 * hat(π)	-1.08%	16.89%	-0.76%	15.23%	0.55%	19.28%
|π-hat(π)| > 3 * hat(π)	-0.57%	15.18%	-0.38%	13.73%	0.82%	19.05%
|π-hat(π)| > 4 * hat(π)	0.45%	14.61%	-0.83%	11.53%	0.48%	17.87%
|π-hat(π)| > 5 * hat(π)	0.83%	13.58%	-0.32%	8.83%	0.21%	15.20%
|π-hat(π)| > 7.5 * hat(π)	0.32%	8.63%	-1.08%	4.60%	-0.21%	9.34%
|π-hat(π)| > 10 * hat(π)	-0.38%	5.15%	-1.40%	2.03%	-0.48%	6.40%

Notes:

1. Sample *1* is the collection of firms for which the income variable, sales, net PPE, total assets, and **labor costs** are all nonmissing in the 2004 Compustat data. Sample *2* is the collection of firms for which the income variable, sales, net PPE, total assets, and **number of employees** are all nonmissing in the 2004 Compustat data.

2. WLS estimation performed on this sample for following specification: income variable on the "index" variable. For firm *i* , the index equals Σ α_j (*i* 's value for factor *j* / sample mean value for factor *j*), where α_j is the weight on factor *j* specified by the relevant formula. This is the only RHS variable. The weight is total assets raised to -0.5. The constant term is omitted.

3. The income variable prediction, hat(π), is the product of the WLS estimate for the slope on the index variable and the index variable value.

4. The income variable prediction, hat(π), is exactly 0 for firms for which the component factors of the index variable (sales, PPE, etc.) are all exactly 0 (recall that the constant term is omitted).

5. The mean of the proportional absolute prediction error, |π-hat(π)| / hat(π), is computed for this smaller sample because of division by 0 for firms with predicted income of 0.

6. Each cell reveals the additional percent of the sample for which the income variable prediction error is outside of the specified interval, for the constrained vs. the unconstrained model.

Table 11: Constrained - Two Factor: (1/2 PPE, 1/2 labor measure)

	Operating Income		***Taxable Income Derived from Taxes***		***Market Capitalization***	
Sample[1]	***1***	***2***	***1***	***2***	***1***	***2***
N (Sample)[2]	1575	6992	1574	6986	1461	6384
N (0-value for hat(π))[3, 4]	28	64	28	64	27	58
N (mean computation)[5]	1547	6928	1546	6922	1434	6326
mean[|π-hat(π)| / hat(π)]	27.72	49.51	2.92	14.34	6.98	53.26
median[|π-hat(π)| / hat(π)]	2.97	7.30	1.40	2.36	2.21	8.39
Percent of firms for which:						
|π-hat(π)| > 0.05 * hat(π)	98.38%	99.32%	98.58%	99.38%	98.54%	99.30%
|π-hat(π)| > 0.1 * hat(π)	97.29%	98.80%	97.15%	98.79%	97.00%	98.51%
|π-hat(π)| > 0.2 * hat(π)	93.92%	97.73%	93.98%	97.27%	94.35%	96.90%
|π-hat(π)| > 0.25 * hat(π)	92.11%	97.06%	92.56%	96.48%	93.03%	96.27%
|π-hat(π)| > 0.5 * hat(π)	85.46%	94.05%	84.48%	92.98%	82.91%	92.78%
|π-hat(π)| > 0.75 * hat(π)	78.67%	91.73%	78.27%	88.88%	74.90%	89.72%
|π-hat(π)| > 1 * hat(π)	74.60%	89.03%	57.96%	62.47%	67.99%	86.36%
|π-hat(π)| > 2 * hat(π)	62.06%	80.01%	40.69%	52.59%	53.77%	79.40%
|π-hat(π)| > 3 * hat(π)	49.58%	71.59%	28.98%	45.94%	39.12%	72.29%
|π-hat(π)| > 4 * hat(π)	37.10%	64.65%	19.40%	40.23%	28.31%	66.91%
|π-hat(π)| > 5 * hat(π)	26.12%	59.37%	13.52%	36.35%	19.53%	62.16%
|π-hat(π)| > 7.5 * hat(π)	14.74%	49.26%	6.53%	28.63%	9.97%	52.99%
|π-hat(π)| > 10 * hat(π)	9.83%	42.39%	3.49%	22.84%	7.04%	45.61%
vs. unconstrained model[6]	*(percentage point difference)*					
|π-hat(π)| > 0.05 * hat(π)	-0.09%	2.36%	0.80%	1.71%	0.11%	2.15%
|π-hat(π)| > 0.1 * hat(π)	1.04%	4.80%	1.29%	3.14%	0.57%	4.45%
|π-hat(π)| > 0.2 * hat(π)	1.05%	9.65%	2.20%	6.15%	1.82%	8.20%
|π-hat(π)| > 0.25 * hat(π)	1.72%	11.65%	2.68%	7.72%	2.90%	10.65%
|π-hat(π)| > 0.5 * hat(π)	3.39%	22.57%	4.53%	15.16%	2.52%	20.47%
|π-hat(π)| > 0.75 * hat(π)	4.24%	31.65%	8.25%	22.04%	3.89%	30.30%
|π-hat(π)| > 1 * hat(π)	5.96%	35.96%	10.98%	30.19%	6.51%	37.18%
|π-hat(π)| > 2 * hat(π)	10.59%	42.84%	15.92%	36.93%	14.83%	43.51%
|π-hat(π)| > 3 * hat(π)	17.77%	42.43%	16.37%	36.53%	16.37%	45.02%
|π-hat(π)| > 4 * hat(π)	19.10%	41.52%	11.70%	34.33%	13.99%	45.01%
|π-hat(π)| > 5 * hat(π)	13.90%	40.12%	9.25%	31.90%	8.63%	43.39%
|π-hat(π)| > 7.5 * hat(π)	6.66%	34.41%	4.31%	26.16%	2.98%	38.73%
|π-hat(π)| > 10 * hat(π)	2.57%	29.45%	1.46%	20.90%	1.56%	33.61%

Notes:

1. Sample *1* is the collection of firms for which the income variable, sales, net PPE, total assets, and **labor costs** are all nonmissing in the 2004 Compustat data. Sample *2* is the collection of firms for which the income variable, sales, net PPE, total assets, and **number of employees** are all nonmissing in the 2004 Compustat data.

2. WLS estimation performed on this sample for following specification: income variable on the "index" variable. For firm *i*, the index equals Σ α_j (*i*'s value for factor *j* / sample mean value for factor *j*), where α_j is the weight on factor *j* specified by the relevant formula. This is the only RHS variable. The weight is total assets raised to -0.5. The constant term is omitted.

3. The income variable prediction, hat(π), is the product of the WLS estimate for the slope on the index variable and the index variable value.

4. The income variable prediction, hat(π), is exactly 0 for firms for which the component factors of the index variable (sales and PPE) are both exactly 0 (recall that the constant term is omitted).

5. The mean of the proportional absolute prediction error, |π-hat(π)| / hat(π), is computed for this smaller sample because of division by 0 for firms with predicted income of 0.

6. Each cell reveals the additional percent of the sample for which the income variable prediction error is outside of the specified interval, for the constrained vs. the unconstrained model.

Appendix Table 1
Descriptive Statistics - Weighted

Notes: The weight is one divided by the square root of contemporaneous firm assets.

Labor Compensation Samples	Mean	Median	Standard Deviation	No. Obs.
Operating Income	100.81	25.26	225.66	1575
Taxable Income as Reflected in Taxes Paid	72.46	12.25	195.10	1574
Market Capitalization	1,125.56	326.00	2,158.37	1461
Sales	744.98	108.45	1,583.82	1574
Property, Plant and Equipment	337.74	23.23	828.14	1574
Labor Compensation	170.64	25.73	495.51	1574

Employment Samples	Mean	Median	Standard Deviation	No. Obs.
Operating Income	63.77	24.95	181.59	6992
Taxable Income as Reflected in Taxes Paid	57.71	9.05	163.85	6986
Market Capitalization	1160.62	553.60	2085.67	6384
Sales	767.60	305.71	1401.35	6986
Property, Plant and Equipment	287.64	46.15	645.29	6986
Employment	0.0041	0.0011	0.0220	6986