Metrics for International Business Research

Randall Morck and Bernard Yeung

The Oxford Handbook of International Business (2 ed.)
Edited by Alan M. Rugman

Print Publication Date: Jan 2009
Subject: Business and Management, International Business, Research Methods
Online Publication Date: Sep 2009 DOI: 10.1093/oxfordhb/9780199234257.003.0028

Abstract and Keywords

This article focuses on performance measures. It starts with the usual focus on profitability and market value as performance measures, followed by alternative performance measures based on firm productivity, growth and survival. It then discusses issues on survey data, and the problem of *ex post* reasoning in empirical work. There are several important themes in this discussion. First, it discusses the definitional and variable construction issues. Second, it addresses the analytical contents in these variables as performance measures. Third, it highlights the statistical and empirical difficulties in using these performance measures, especially in an international cross-country context.

Keywords: performance measures, profitability, market value, productivity, international business research, cross-country comparisons

28.1 Introduction

INTERNATIONAL Business is still a young discipline. Like all branches of science, the field's growth depends not just on its relevance, but on its ability to produce knowledge in a cumulative manner. As an illustration, the field of finance is one of the most advanced branches in business and economic studies that has experienced very rapid growth in the past thirty years. This cannot be done but for that its theories can be rigorously contested empirically. The rigorous empirical tests filter out unsupported ideas and push the development of relevant ideas that have an empirical component. Curious empirical patterns also solicit rigorous theoretical development. The result is cumulative and mutually reinforcing development in theory and empirical methods. International business is beginning to develop in the same manner. Empirical work is particularly the driving force in international business because it is an applied field. In this context, the issue of ‘metrics’ in international business deserves special attention.

Generically, most IB empirical research focuses on the firm level, linking international business behaviour to firm performance. While this is a gross simplification, (p. 798) it allows us to conduct a focused and manageable discussion. We do hope that this gross simplification does not create the impression that many other important related research styles are under appreciated. Even in this simplified framing, it is still close to impossible to characterize generic 'international business' firm behaviour. The most broad-based metering attempt in international business is perhaps the measurement of 'internationalization'. One should consult Rugman and Oh (2007) on the various issues and the latest thoughts. The article illustrates the high level of intrigue even on just the measurement of multinationality.

In this chapter we, therefore, focus only on performance measures. Our discussion starts with the usual focus on profitability and market value as performance measures, followed by alternative performance measures based on firm productivity, growth and survival. We then discuss issues on survey data, and the problem of *ex post* reasoning in empirical work. There are several important themes in our discussion. First, we discuss the definitional and variable construction issues. Second, we address the analytical contents in these variables as performance measures. Third, we highlight the statistical and empirical difficulties in using these performance measures, especially in an international cross-country context. Fourth, we take this opportunity to make useful suggestions on basic ingredients for solid empirical work. We make a pitch for recognizing the complementarities between contemporary cross-sectional data studies and historical studies as well.

28.2 Financial Bottom-Line Measures

28.2.1 RoA and q

28.2.1.1 What is Being Measured, and How?

We start our discussion with 'bottom-line' performance measures. The typical ones are return on assets and listed firms' stock value, e.g. Tobin's *q*, defined as a firm's market value (stock and debt included) over the value of its tangible assets (see e.g. Lindenberg and Ross 1981; Morck and Yeung 1991). Researchers use changes in firm value as a metric too, leading to the so called 'event studies', (see e.g. Fama et al. 1969; Morck and Yeung 1992; MacKinley 1997). Using these bottom-line performance measures researchers ask the question how a firm's strategy is related to a firm's capital return and/or value. For presentation convenience, we discuss the first two measures first and then the event studies afterwards.

Conceptually, without worrying about data integrity, returns on asset capture a firm's revenues net of labour and material costs; that is, the portion of earnings available to capital owners. The earnings of course are often greater for a firm with more assets; hence researchers scale the earnings by total assets, e.g. total book value (p. 799) of assets, and thus the term *ROA*, which is a rate of return measure. Because the denominator includes all assets, the numerator should be earnings available to both shareholders and creditors.

Normally, we want to exclude extra-ordinary income so that ‘earnings’ is ‘operating income after taxes’ plus interest, rental, depreciation, amortization and depletion expenses.[1] Some may scale the earnings by networth; thus the term *ROI*. If so, the numerator should be only earnings available to shareholders, not creditors; that is, interest expenses should be excluded. In both *ROA* and *ROI*, tax expenses should be excluded which is why ‘operating income *after* taxes’ is used. However, actual tax expenses are difficult to obtain and ‘operating income after taxes’ may have errors.[2] Therefore, some simply use ‘operating income before taxes’ as a substitute counting the government as an asset holder. Finally, some may directly estimate *ROI* based on the *rate of return* to a firm's publicly traded shares, if it is available.

Return data have quite a bit of volatility. Some therefore use the average over a few years' worth of data if the research question permits. Still, a fundamental issue is that there is a trade-off between expected returns and risks: higher returns are associated with more risk-taking. Hence, *ROA* or *ROI* data can be misleading unless we properly adjust for risks. Note that industry fixed factors cannot absorb away the risks, e.g. John, Litov, and Yeung (2008) show that there is quite a bit of risk variations even after controlling for industry fixed effects.

The *Tobin's q* offers an alternative. The variable is defined as the value of equity and debt over the replacement costs of tangible assets. Typically, the numerator includes the sum of the market value of common stocks, estimated value of preferred stocks, estimated value of long-term debts, current liabilities net of current assets and the denominator includes the sum of the estimated replacement costs of tangible assets and inventories. The numerator captures a firm's total value; that is, the total capitalized value of the expected earnings available to various categories of investors and creditors. This valuation by definition has adjusted for risks. A firm's Tobin's *q* is higher the more value it can generate using the same among of tangible assets. We should note, however, this method is applicable only to publicly traded companies.

The estimation of *q* can be quite laborious and data intensive.[3] In recent years, researchers have adopted some simpler versions, such as replacing the estimated market value of long term debts and preferred stocks by their book value and the replacement cost of tangible assets by the sum of their book value. The simplification seems not to matter much, as Christophe (1997) points out. Note, however, that the experience is based on US data only and in a rather low inflation regime. Sometimes, researchers use ‘*market to book*’ value, which is the market value of common stock over the book value of net worth. Sometimes, researchers use *market value over sales*. We are not aware of results documenting misleading discrepancies among these measures: Tobin's *q*, market to book, market value over sales, etc. Again, note that most of our experiences are based on US and Western European data. In a broader international context, there might well be non-trivial problems.

(p. 800) 28.2.1.2 Data Integrity and Information Content

No data are perfect. An integral part of good empirical work is to pay attention to data integrity.

28.2.1.2.1 Data Errors

The first question is—what are actually put into my variables? All good empiricists carefully examine the question. We should never blindly trust our data, especially if they are from secondary sources, which is often the case when we use accounting or financial data. Two issues are most common: (1) a variable is not what we think it is; and (2) data entry errors.

On the first one, let us illustrate. For years, researchers used patent citations as an indication of knowledge spillovers assuming that author A cites author B in a patent because A receives knowledge spillovers from B. Alcacer and Gittelman (2006), going to the sources on how data are composed, find that a large number of citations are imposed by examiners; that does not represent spillovers. Their work clearly alerts us of possible misuse of patent citations.

Let us illustrate the concern about data entry error too. No matter how careful the supplying company is, errors can happen; after all data entry jobs are not high paying jobs. For example, we often find confusing results from relating firm value to institutional ownership. Gompers and Metrick (2001) spend a huge amount of time examining every observation in their institutional investor equity blocks data, and find that simple data entry errors are so common in the standard dataset that the inconsistent results of previous studies are likely entirely due to data problems. Hornstein (2006) examines firm-level international presence data reported in the Directory of Corporate Affiliation; she finds that a non-trivial number of records in the Directory do not agree with publicly filed company records. Our suggestion is that we should always first verify that the definition of a variable coincides with our desire. Then, we validate our data by cross-comparing our data from multiple sources, e.g. company financial statements and data banks. If a complete check is too costly or impossible, even a partial check could be valuable.

28.2.1.2.2 Managerial Manipulations

Even if data definitions fit our desire and data entries are not problematic, we still have to worry about the actual information content in our variables. There is a rather long list of potential problems.

Accounting data are subject to managerial manipulation, sometimes illegally and sometimes not. The accounting literature calls this earnings management, see, e.g. Healy and Wahlen (1999). The blatant example in the US is the Enron case, where revenues were overbooked and expenses were hidden. The basic issue is that accounting data may deviate from what they are supposed to represent and the deviation may have a systematic relationship with the firm strategy we are examining. For example, a manager who advocates and actually pursues an international expansion strategy may be tempted to inflate his or her company's earnings to garner support for his or her strategy. We expect this problem to be more prevalent in locations where accounting (p. 801) disclosure require-

ments are less rigid, auditing is less rigorous, and investors' rights are less well protected.

There are more subtle cases stemming from transaction manipulations that lead to misleading data. In the US and UK, almost all firms are stand alone; outside of the US and UK, many firms are often a part of a big business group. A group often has an ultimate controlling interest which controls group member firms via pyramidal and cross holdings and the rights to hire and fire executives and board members. Some groups may not have an ultimate control interest, but, they are controlled by a small group of elite executives running the groups' flagship enterprises. While the ultimate controlling interest directly or indirectly controls a vast among of member firms, it has varying degree of equity investment in them. (See e.g. Morck, Wolfenzon, and Yeung 2005 for a comprehensive survey.) The controlling interest has the incentive to, can, and will use within group transactions to transfer resources from one firm to another (see e.g. Johnson et al. 2000; and Morck, Stangeland, and Yeung 2000). The result is that accounting earnings data from one firm may not really reflect a strategy's actual impact. For example, if firm XX controlled by Ms. Zhang in China establishes a wholly owned subsidiary in Canada and Ms. Zhang uses the subsidiary to tunnel resources to an apparently separated firm in the US that is actually controlled by her, XX's bottom line could be negatively affected by the establishment of the Canadian subsidiary and so may the whole business group. The actual economic effect of the strategy, however, is not clear. To find out, a researcher has to examine the total impact of the establishment of the subsidiary on the empire controlled by Zhang.

International business researchers should be familiar with tunneling; we call that income shifting or transfer pricing (see e.g. Harris et al. 1993). The implication of international income shifting is that one subsidiary's performance is not an adequate representation of the overall effect of an international strategy. Indeed, this point is valid even if there is no income shifting. For example, if a subsidiary is established for cost centre purposes, the subsidiary's low earnings could mean success.

28.2.1.2.3 Cross-Holdings and the Historical Cost Principle

Cross-holdings and business group arrangements is the source of another problem, and this one applies specifically to Tobin's *q* or to *market to book* index. Consider the case of Japan. A firm with lots of cross holdings is required to enter the investments in its balance sheet in historical costs. As the market value of the holdings rises, which one would expect to be normally the case, the firm's market value rises far and above the book value (see Morck and Nakamura (1999) for more detailed discussion). This effect, when not taken out, can distort empirical results based on the *q* and *market to book* indices.

Along this line, one needs to be sensitive to the use of historical costs in book entries. As we point out above, careful and detailed estimation of *q* is very laborious and data intensive. Sometime, we use the book value to replace tangible assets' replacement costs and also long term debts' market value. The same short cut is also adopted in calculating *RoA*. This practice is not problematic when we are using data in a low inflationary regime. In a

high inflation regime, we need to be concerned (p. 802) because the discrepancies between replacement costs and book entries are large and grow rapidly overtime. For example, Hall (1993) compares return data from the 1970s and 1980s and she needs to worry about inflation distortions. Later studies using later US data do not need to do this because historical costs are better approximations to actual asset and debt. Similarly, in linking insider ownership to *q*, Morck, Shleifer and Vishny (1988) have to worry about debts and assets carried at historical cost and adjust these for inflation because they use data at the end of a decade of high inflation in the US.

The implication is that we need to exercise caution when we compare variables like *q*, *market to book*, *RoA*, etc., across countries and over time. Perhaps, that raises the appeal of variables like *market value over sales*. Still, in these variables, we have to control for country factors. Moreover, these variables are not free of the following problem.

28.2.1.2.4 The Information Content of Capital Market Data

The use of capital-market-based data to measure performance is based on the premise that capital markets incorporate firm specific information in the evaluation of assets and they do the job reasonably well. (The following discussion also applies to Section 28.2.2.) Consider the value of an instrument, e.g. a company's common stock. Investors dig up firm-level information, evaluate the company's situation, make a projection on the stream of future earnings the share entitles them to, and then calculate the present value of the earnings based on an appropriate risk adjusted discount rate. They trade shares based on the difference between their private valuation and the market price—a typical buy-low-sell-high strategy. The consequence of this process is that the latest market price of a share represents the marginal trader's evaluation of the share's value. This is the informed risk arbitrage process explained in Grossman and Stiglitz (1980) and Shleifer and Vishny (1997).

This process reliably produces equity prices that reflect firm specific information. But the result depends on that (1) investors are confident that their investors' rights are protected, otherwise they will not be interested to even look for information; and that (2) most investors are rational. The two conditions are related. Given the first condition, mis-pricing due to irrational investors will be mitigated and the proportion of irrational traders will shrink. (For a discussion of noise traders, see e.g. De Long et al. 1990.)

The question is whether the first condition holds. In many countries, investors’ rights are poorly protected; perhaps the government is corrupt and prone to expropriate private assets, or, perhaps, the location does not have well enforced rules and regulations that protect outside investors from insiders' appropriation of corporate resources.[4] In such situations, investors do not find it worthwhile to dig up information because the benefits are so uncertain. Morck, Yeung, and Yu (2000) show that in countries where property rights are weak equity prices have little firm specific information content and are possibly much affected by ‘noise’ trading. The lesson here is that a basic data assumption is questionable in many countries, particularly in developing economies.

(p. 803) 28.2.1.3 Old Fashion Data Scanning

Besides conducting checks on data integrity and information content, a good empiricist typical would do some extra work before submitting data to statistical packages. For generations of good empiricists old fashion data scanning is a required empirical ritual. Identifying extreme value observations is often a useful step to identify irregularities and to understand data problems. Plotting of the dependent and independent variables to identify common and outlying patterns is a powerful tool to develop a 'feel' for the data. We highly recommend these practices.

28.2.1.4 What Benchmark?

Bottom-line measures are meant to be compared with benchmarks. Choosing the right benchmark is a critical issue. Let us use the empirical test of the internalization theory as an illustration. The test is similar to the large empirical literature on whether diversification (segment or geographic) improves or hurts firm performance. Indeed, Morck and Yeung (2003) regress Tobin's q on both firm level diversification measures.

The approach has at least two benchmarking problems. The first problem is on the firms we compare a diversified firm to—the key is to correct for endogeneity. Consider the hypothetic case that firms I and II are similar firms. They both face diminished profit opportunities at home so that their Tobin's q is lower than the general population of firms. Suppose firm A finds some profitable application of its intangible capability in another country while firm B does not. Thus, firm A's diversification enhances its value above B's. However, A's firm value could still be lower than other firms so that using the general population of firms' q as a benchmark leads to misleading inference (see e.g. Campa and Kedia 2002; and Villalonga 2004).

The second problem is due to an inaccurate construction of a benchmark comparison. Let a firm have γ proportion of its assets located in the US and $(1 - \gamma)$ in the rest of the world. Assume that the Tobin's q for a 'pure play' (purely domestic) US firm be q_{US} and for a 'pure play' firm in rest of the world be q_{ROW}. Let the synergy for international diversification be S, which can be positive or negative and is what we want to find out. The internationally diversified firm's q is '$\gamma q_{US} + (1 - \gamma)\, q_{ROW} + S$'. A regression that regresses the q's of US multinationals and US domestic firms on an indicator of geographic diversification is similar to comparing '$\gamma q_{US} + (1 - \gamma)\, q_{ROW} + S$' to q_{US} which is similar to examining whether '$(1 - \gamma)\,(q_{ROW} - q_{US}) + S$' is positive or negative. Note that this comparison does not identify the sign of S cleanly. '$(1 - \gamma)\,(q_{ROW} - q_{US}) + S$' can be positive even if S is negative as long as q_{ROW} is sufficiently greater than q_{US}. Also, '$(1 - \gamma)\,(q_{ROW} - q_{US}) + S$' can be negative even if S is positive as long as q_{ROW} is sufficiently smaller than q_{US}. Note that in the late 1970s and early 1980s q_{ROW} was likely greater than q_{US}. In the 1990s q_{ROW} was likely less than q_{US}. Hence, using US firms we may inadvertently find that a multinational structure raises firm value in the 1970s and 1980s and does the opposite in 1990s; yet we have no clear-cut information on 'S,' the real impact. To deal with the problem, the (p. 804) benchmark comparison should be the q of the geographically diversified firm against '$\alpha q_{US} + (1 - \alpha)\, q_{ROW}$.' More generally, the benchmark q for compari-

son, which is often called the *chop shop q,* should be $\Sigma_i \gamma_i q_i$ where i is a location subscript, γ_i represents the proportion of assets in location i and q_i represents the q for a non-diversified firm in location i. Unfortunately, to identify γ_i and q_i location by location is a very difficult exercise.

28.2.1.5 Common Statistical Issues

Finally, we would like to point out other common statistic issues, besides the endof-geneity issue we just raised above. The bottom-line measures—*RoA, q, market to book value,* etc.—are all subject to not just firm specific influences but are also influenced by the market environment. For example, the development of an international opportunity that stems from a common technological break through could impact on multiple firms, not just on the firm that is taking current international action. The implication is that the capital market data of a cluster of firms, perhaps all from the same region and in very similar industries, are affected. The result of the common latent shock is a common component in the error term for the cluster of firms— there could be a common constant, a common standard error, a common cross correlations and so on. All these are treatable problems. However, if ignored, they can cause exaggerated estimates for statistical significance and inaccurate inferences (see e.g. Petersen 2005; and Donald and Lang 2007).

We strongly advocate carefully theoretical and empirical exploration of the error structure of our empirical model. Failing to do so is the reason for a garden variety of common statistical problems in weak empirical papers.

In addition, we believe that, given the nature of performance measure data, we should never over look the importance of residual diagnostics. Our results could be driven by outliers. More often than not that is a blessing in disguise; outliers give us a chance to deepen our understanding of our empirical structure. Residual diagnostics checks save us from future embarrassment.

28.2.1.6 Summary

Using bottom-line performance measures to track the impact of international business behaviour is much easier said than done. In the above, we draw the readers' attention to a variety of issues which we believe apply to other sections. It would be useful to summarize by providing a generic check list of useful questions for careful empiricists:

1. How are your variables constructed? What are being measured?
2. How is the data integrity? Data entry errors? Any systematic deviation from what the data are supposed to measure? Are the assumptions that allow the data to measure what they are supposed to measure acceptable?
(p. 805)
3. Do you thoroughly understand the regularity and irregularity of your variables? Do they assume reasonable values? Do the simple relationships among variables behave as unexpected? If not, why?
4. Have you set up the right comparison benchmark in your empirical specification?

5. Have you investigated the error structure of your empirical specification? Any self-selection and endogeneity issues? Any time series and cross-section correlations in the error terms? Any possibility of misrepresenting the statistical significance of your results?
6. Have you conducted residual diagnostics checks? How statistically robust are your results?

28.2.2 Event Studies

28.2.2.1 The Logic

Capital market valuation data are forward looking. That hinders causal interpretation of any relationship between a firm's market value and its strategy. For example, a positive relationship between a company's current *q* and strategy could be driven by a common latent factor: the development of a new investment opportunity. Event studies are sometimes proposed as a solution. For an original introduction of the methodology, see Fama et al. (1969). The other advantage of an event study is that it saves researchers the trouble in developing a *chop shop q* or to identify the risk adjustment for *RoA*. The disadvantage is that we cannot use the method absent a functioning and efficient stock market.

The logic of an event study is that if an exogenous event, previously unknown to investors, changes investors' assessment of a firm's future earnings, the firm's market value will change accordingly. Given that the stock market efficiently impounds the information into share prices, there would be an immediate change in the firm's share price, lest there would be predictable profitable arbitrage opportunity. If the event has little impact, there should be no significant changes in stock prices. Hence, the change in the firm's stock price allows us to see how an event impacts on the firm's performance. Note that if the logic is correct, we should be able to identify commensurate actual changes in earnings data post an event. Therefore an event study is similar to an intervention analysis applied to earnings data (e.g. *RoA*).

It is worthwhile to emphasize the basic assumptions of an event study are that capital markets efficiently impound information into stock prices and the focal event hitherto is not in the information set of investors, at least not completely. *A researcher must carefully check these conditions before adopting the methodology*. For example, the methodology is unlikely to be effective if an event develops gradually and the information slips into the investors' information set over a rather long stretch of time. As well, a research must carefully check whether there are confounding events. Another possibility is that stock prices are noisy determined so that firm value and (p. 806) firm specific information do not have a strong relationship. In either case, not finding a significant impact of an event on stock value does not necessarily mean the event has no real impact.

28.2.2.2 The Abnormal Return Measure

A critical step in an event study is to capture the change in a share price that is not due to a change in market return—that is the so called *abnormal return*. Accumulating abnor-

mal returns over a stretch of a few days, e.g. three or five days, around an event day leads to the *cumulative abnormal return (CAR)*. To estimate the abnormal return, researchers first estimate an asset pricing model, e.g. *CAPM*, international *CAPM*, or the recent Fama and French multiple factor models (Fama and French 1992), using pre-event data. They then use the result to predict the portion of return that is driven by the variables in the right-hand side of the market model, e.g. the market return, on the event day. The abnormal return is the stock return on the event day minus the predicted returns. To take care of heteroskedasticity researchers scale the abnormal return by the standard deviation of the abnormal returns estimated using the window of data before the event day.

We typically find that event study results are not sensitive to the type of asset pricing models one uses. Suppose that two models give an annualized difference in predicted return to the extent of 5 per cent. In a three day window, that is miniscule because $(1 + 5\%)^{(3/365)} - 1$ (= 0.0004011) is almost zero. By the same token, we also often find that simply defining 'abnormal return' as the target firm's return minus the market return on the event day is an acceptable approach. Indeed, if an event is real, its *CAR*s should be large enough that all reasonable benchmarks would give qualitatively similar results. If they do not, the results are suspect.

For international studies a different issue exists—a simple comparison of abnormal returns across countries is illegitimate if the abnormal returns are in different currency units. We should first convert stock returns across countries in an identical currency unit and employ an acceptable global capital asset pricing to derive abnormal returns. Note that simple country fixed effects do not necessarily absorb away the influence of exchange rate changes because the influence is not time invariant.

28.2.2.3 Statistical Concerns

Concerns about statistical problems we raised in the previous subsection all apply here. In particular, researchers should be most concerned about common components in the abnormal returns. For example, the error terms in the abnormal returns for a cluster of firms could have a common constant, a common standard error, a common cross-correlation, etc. These issues, if not treated, lead to incorrect statistical inferences. There are many standard econometric techniques in dealing with the issues, e.g. fixed effects model, random effects model, clustered estimation of standard errors, etc. Readers should consult Campbell, Lo, and MacKinley (1996) and MacKinley (1997) to understand the event studies method and related econometrics.

(p. 807) 28.3 Productivity, Growth, and Survival

Researchers sometimes adopt performance measures that are not directly related to the financial bottomline. For example, many studies of the impact of foreign direct investment focus on the host country firms' and industries' *productivity* since Caves (1974), e.g. Blomstrüm (1986); Haddad and Harrison (1993); Kokko (1994). Aitken and Harrison (1999); and Chung, Mitchell, and Yeung (2003), to name a few. The commonly adopted

measure in recent years is total factor productivity. Others adopt performance measures like growth in sales, or growth in employment, or survival.

28.3.1 Total Factor Productivity

28.3.1.1 Meaning of TFP

One can illustrate the definition of total factor productivity using a simple Cobb- Douglas production function, (1) $$Y = AK^{\beta_1} L^{\beta_2} M^{\beta_3}$$

where *Y* is quantity output and *K, L,* and *M are,* respectively, capital, labour, and material inputs. Often we assume the β*s* sum to 1 (constant returns to scale), but we can do without the restriction in this presentation. Using the lower case to represent natural log, we can write (2) $$y = \alpha + \beta_1 k + \beta_2 l + \beta_2 m$$

Typically, people refer to a as the total factor productivity. Verbally, this means that assuming that firms 1 and 2 have identical β*s,firm 1 has higher total factor productivity than firm 2 because given the same inputs firm 1 can produce more than firm 2.* We also define *total factor productivity growth as the proportion of output growth that is not accounted for by factor employment growth.* That is, when defined using equation (2), total factor productivity growth is $\Delta y - (\hat{\beta}_1 \Delta k + \hat{\beta}_2 \Delta l + \hat{\beta}_3 \Delta m)$ where the β's represent the predicted value of β's.

28.3.1.2 Estimation Difficulties

The estimation of total factor productivity is non-trivial. The estimation inevitably involves a specification of the production function, say, in the form of equation (2). First, note that some of the variables are not readily observable, e.g. *k. y* could have a quality dimension. Even *l* can be problematic because the natural log of labour may mean not just the natural log of the number of employees, but, the number of effective employees. In such case the quality of labour (e.g. proxied by workers' year of schooling) should be considered. While careful studies delve into these issues, many (p. 808) studies simply proxy for *y* by the natural log of revenues, *k* by the natural log of total property plants and equipment, *m* by the natural log of expenditures on purchases, on energy, and on land rental, and *l* by the natural log of number of employees. We should ascertain that *y* is the natural log of valued-added if only *k* and *l* are kept on the right hand side of equation (2) or that other factor inputs like purchases and energy, as represented by *m* in equation (2), are included in the right-hand side.

The estimation of a production function and thus the total factor productivity is by itself a difficult exercise. There is an endogeneity problem: factor inputs are correlated with productivity shocks so that the regression coefficient estimates for equation (2) is biased and inconsistent, e.g. see Olley and Pakes (1996). There is also a survival bias problem because firms with very bad negative productivity shock may have to exit. To simplify our discussion, we shall overlook the survival bias. Let us now illustrate by first include the

error term explicitly in equation (2):

(2') $$y_t = \alpha + \beta_1 k_t + \beta_2 l_t + \beta_3 m_t + \omega_t + \varepsilon_t$$

The error terms ω_t and ε_t represent, respectively, productivity shocks correlated and uncorrelated with inputs. Note that the endogeneity problem stems from that the inputs are correlated with ω_t. Olley and Pakes (1996) turns the problem into a solution strategy. Assuming an 'instrument', like investment, I_t (in natural log), is a monotonic function of the productivity shock ω_t , we may be able to invert the function and write the shock as a function of I_t. For example, if we can write:

(3) $$\beta_1 k_t + \omega_t = \varphi(I_t, k_t)$$

where *cp* is a polynomial in I and k, we can now express equation (2') as follows:

(4) $$y_t = \alpha + \beta_2 l_t + \beta_3 m_t + \varphi(I_t, K_t) + \varepsilon_t$$

If equation (4) is correct, we can estimate α, β_2 and β_3 consistently.

We now need to estimate β_1. Let us further assume that $k_t = (1 - d)\, k_{t-1} + I_{t-1}$ so that k*t* is predetermined and that $\omega_t = \omega_{t-1} + \eta_t$ (i.e. ω_t is a 'random walk'). We then can rearrange equation (4) as follows:

(5) $$\begin{aligned} y_t - \hat{\alpha} - \hat{\beta}_2 l_t - \hat{\beta}_3 m_t &= \beta_1 k_t + \omega_{t-1} + \eta_t + \varepsilon_t \\ &= \beta_1 k_t + g(\hat{\varphi}_{t-1} - \beta_1 k_{t-1}) + \eta_t + \varepsilon_t \end{aligned}$$

Equation (5) allows a consistent estimate of β_1. (We can estimate equation (5) using OLS because, by assumptions, k_t, k_{t-1}, and I_{t-1} are uncorrelated with η_t and ε_t).

The estimation procedure is non-trivial and involves many assumptions that some may find hard to accept. In particular, in many environments it may not be easy to accept the critical assumption that productivity shocks have a monotonic relationship with investment for all sample firms. (If investment is a function of productivity shocks and if the function is 'invertible', we can express productivity shocks as a function of investment and develop a solution strategy.) For example, in weak institutional environments, one could easily find counter examples in which investment capital is not allocated according to productivity and also in which agency problem is prevalent. Other solution strategies try to by pass this particular assumption (e.g. (p. 809) Levinsohn and Petrin 2003). Inevitably, a solution strategy involves finding instruments to solve the endogeneity problem. It typically raises data requirements. In lower income countries where reliable firm level data are particularly hard to obtain, these additional data requirements can be difficult to fulfil.

Given the difficulties in estimating total factor productivity, some may resort to simple average capital or labour productivity measures, *Y*/*K* or *Y*/*L*. We need to recognize that these measures are affected by other factor inputs, e.g. *Y*/*K* is inevitably greater when more labour is employed.

28.3.2 Growth and Survival

Using growth and survival as measures of performance is popular in international business research. Growth and survival do not necessarily reflect better performance or better capabilities. Self-serving managers are known to pursue excess growth and to prolong inefficient survival. Likewise, while shrinkage and non-survival may indicate business failures, they may not be either. A smart manager may pursue shrinkage for better efficiency and a greater rate of return. A successful business may appear not to survive because it is acquired at a high price.

In international business, using growth and survival as a performance metric can be a particularly tricky issue for at least two main reasons. First, continuing with the previous argument, the meaning of growth and survival is often context dependent and is linked to corporate governance issues. The use of growth and survival as a performance metric is often done for firms located in advanced countries with strong institutions. In locations with poorer institutions, where market disciplinary forces are inadequate and where corporate governance is weak, we must exercise extra care in interpreting growth and survival data. Second, even in environments where market disciplinary forces are robust and corporate governance is strong, we have to recognize that a subsidiary's growth and survival have a context dependent relationship with performance. While a failing subsidiary may indeed shrink and exit, such experience may also provide great learning to management and indeed the whole process may be ‘planned’. Or, the shrinkage and exit of some units may not indicate failure; they may be planned by a successful multinational which is rearranging its portfolio of international presences.

Our point is that we always need to go back to the basics; metrics have to be developed with care and tie in carefully with both our theories and the empirical context.

28.4 Survey Data

Some measure performance by conducting surveys. Attempt to generate first hand data is admirable. We make a few brief comments here.

(p. 810)

The first is that survey results can be affected by ‘framing’. For example, if we first give a presentation on the positive spillover effects of inward foreign direct investment and then survey local business partners on their satisfaction with the foreign partners, we believe that the survey result will be affected. Also, it is well known that the way a question is raised can affect the answer.

Second, in an international setting, language and culture affect how people interpret and answer questions. A Chinese's ‘yes’ to the question ‘You do not like your job?’ means that she indeed dislikes her job. There are many other examples. Asians often say ‘no’ so subtly that Westerners interpret the subtle answers as that a positive possibility exists.

Third, even without cultural and language gaps, survey answers could be ambiguous. For example, an answer ‘I disagree’ to the question that ‘this is a good business strategy’ may mean that I do not think the strategy is good at all, or that I think the strategy is better than good, it is excellent.

Fourth, survey answers are often affected by *ex post* reasoning or cognitive resonance. We may *ex post* construct smart reasons to explain what we do while ex ante we choose to do what we do by chance. We may not like a business plan. However, after it is implemented, we grow to like it and we answer evaluation questions accordingly. Managers tend to sing praise to their own business plans *ex post*. Frankly, will a prosecutor convict a murderer by just directly asking in court whether the murderer has committed the crime?

All these concerns suggest that validation of survey data is critical for the methodology to yield sensible results. Do our survey data capture what they are supposed to capture? Could the survey observations be driven by factors other than the latent forces we have in our measurement model?

28.5 *Ex Post* Reasoning

Finally, we need to address the impact of survival bias and *ex post* reasoning in empirical inferences.

Much international business empirical work relies on contemporary cross- sectional data. That clearly is related to data availability—historical data are harder to find and may not even exist. The data availability issue also tilts our attention towards developed countries, which have more readily available data.

We can infer a time series pattern from a cross-sectional pattern if the two mimic each other, sometimes that is referred to as ‘ergodicity’. But, often they do not and misleading inferences can be drawn from cross-sectional data. Consider the following. Currently, there are many state-owned Chinese firms conducting outward foreign direct investment. For simplicity, let us artificially assume that they are the only outward investing Chinese firms; other good Chinese firms are not able to (p. 811) do so because of financial constraints. These firms pursue outward foreign direct investment because they have excess cash which their decision-makers do not want to keep idle or to pay out as dividends. Choosing amongst investment alternatives, these corporate decision-makers conduct outward foreign direct investment because that fits the state's economic policy. Obviously, some of these firms learn over time. In particular, some lucky firms with strong support from the state will be able to absorb overseas losses, survive, and learn; they become strong multinational firms possessing legitimate capabilities. At the same time, the rest may just retreat back to China. In the future, relying only on cross-section data, we shall not see any Chinese firms which conduct outward foreign direct investment for the wrong reasons. We shall find that home grown Chinese MNSs are strong firms with high levels of capabilities originally backed or owned by the government. It would be tempting to

draw a wrong *ex post* rationalization of China's SOEs' outward foreign direct investment based on the internalization theory or the resource-based view of the firm. (See Morck, Yeung, and Zhao 2008.)

The example is a bit trivial; we believe that good international business researchers do not easily fall into such a simple-minded trap. The example, however, illustrates that cross-section and time-series observations may not mimic each other. For thoroughness in our research programmes, we advocate paying attention to historical studies. See Jones and Khanna (2006) and Morck and Yeung (2007) for more extensive discussions.

Historical studies require careful data collection and development. The payoff, however, can be significant. First, besides the above problem, cross-sectional relations often do not readily allow clean causal interpretation. Historical documents can shed light on motivations and exogenous incidents that might allow us to make causal statements. Second, historical data and documents enable us to identify differences and similarities in the path of events across geography and time. That usually deepens our understanding of the phenomenon we are investigating. Hence, we advocate the complementarities of large-scale cross-section data studies, time-series studies, and historical studies.

28.6 Conclusion

In this chapter, we discuss the metrics for international business research. To have a manageable discussion, we focus on performance measures. We start with bottom- line performance measures based on accounting data and capital market valuation. These measures could be particularly problematic in an international setting because of a variety of reasons. There are non-trivial variations in accounting regulations and practices across countries. There are gaps between apparent and actual information content in both accounting and capital market data. These gaps are systematically (p. 812) related to ownership structure, corporate governance, protection for investor rights, and overall market efficiency, all of which vary across countries. These gaps are systematically related to inflation and other macro-economic conditions too. Finally, when comparing rates of return across countries, we need to make returns in different currencies compatible.

We also comment on performance measures based on growth, productivity, survival, and survey data. Estimating productivity in less-developed countries is a particularly daunting exercise for two reasons. First, assumptions that empower solution strategy to overcome 'endogeneity' may be invalid for a significant portion of firms in these environments. Second, reliable firm-level data are difficult to find. The relations between growth, survival, and economic performance are often context dependent and are related to corporate governance. The relations are particularly tenuous in locations with weak institutional environment, weak market disciplinary forces, and poor corporate governance.

We take this opportunity to raise some methodological issues. International business is an applied social science. We need to carefully develop credible empirical results that lead to cumulative learning. The mission of our empirical efforts is to filter out incorrect theories

and to build up robust results that enable *cumulative* learning and theory building. To carry out our mission, we need to take our empirical work extremely seriously. That takes several important steps.

We ask ourselves to be demanding in the validity in our measures. Are we sure of the accuracy in the definition and compilation of our data and variables? Are we sure that our data measure what we intend them to measure, independent of data definitions on face value? Do the theoretical content of our variables fit the empirical content? Do we understand the regularities and irregularities in our data and variables?

We also ask ourselves to be demanding in the empirical and statistical specifications in our empirical work. Do our theory, empirical specification, data, and statistical method converge? For example, have we set up the right empirical specification so that our empirical results answer our theoretical questions? Have we used the right data and variables that properly link the empirical specification to the theoretical inquiries? Are there other interpretations of the empirical relationships?

We also emphasize attention to the statistical details. For our empirical results to be robust, we need to ascertain that we understand the error structure of our empirical specification. Are there any self-selection and endogeneity issues? Any time series and cross section correlations in the error terms? Any possibility of misrepresenting the statistical significance of your results?

Also, we need to be most certain of the robustness of our results. Would our results be driven by outliers? Would small perturbations in our variable definitions and construction qualitatively change our results? Finally, is the theoretical interpretation of our results robust?

The empirical process is labour intensive and painstaking. It is also an art. Rock solid empirical effort is the driver of an applied field's progress.

References

ALCACER, **JUAN** and **GITTELMAN**, **MICHELLE** (2006). ‘Patent Citations as a Measure of Knowledge Flows: The Influence of Examiner Citations’, *Review of Economics and Statistics*, 88/4: 774–9.

AITKEN, **BRIAN** and **HARRISON**, **ANN** E. (1999). ‘Do Domestic Firms Benefit from Direct Foreign Investment?’ Evidence from Venezuela’, *American Economic Review*, 89/3: 605–18.

BLOMSTRÖM, **MAGNUS** (1986). ‘Foreign Investment and Productive Efficiency: The Case of Mexico’, *Journal of Industrial Economics*, 35/1: 97–110.

CAMPA, **JOSE MANUEL** and **KEDIA**, **SIMI** (2002).‘Explaining the Diversification Discount’, *Journal of Finance*, 57/4: 1731–62.

CAMPBELL, **JOHN** Y. **LO**, **ANDREW** W. and **MC KINLEY**, A. **CRAIG** (1996). *The Econometrics of Financial Markets* (Princeton: Princeton University Press).

CAVES, **RICHARD** E. (1974). 'Multinational Firms, Competition, and Productivity in Host- Country Markets', *Economica*, 41: 176–93.

CHRISTOPHE, **STEPHEN** (1997). 'Hysteresis and the Value of the US Multinational Corporation',*Journal of Business*, 70/3: 435–62.

CHUNG, **WILBUR MITCHELL**, **WILL** and **YEUNG**, **BERNARD** (2003). 'Foreign Direct Investment and Host Industry Productivity: The Case of the American Automotive Parts Industry', *Journal of International Business Studies*, 34/2: 199–218.

DE LONG, J. B. **SHLEIFER**, A. **SUMMERS**, L. H. and **WALDMANN**, R. J. (1990). 'Noise Trader Risk in Financial Markets', *Journal of Political Economy*, 98/4: 703–38.

DONALD, S. and **LANG**, K., (2007). 'Inference with Difference-in-Differences, and Other Panel Data', *Review of Economics and Statistics*, 89/2: 221–33.

FAMA, **EUGENE** F. **FISHER**, **LAWRENCE**, **JENSEN**, **MICHAEL** C. and **ROLL**, **RICHARD** (1969). 'The Adjustment of Stock Prices to New Information', *International Economic Review*, 10/1: 1–21.

——and **FRENCH**, **KENNETH** R. (1992). 'The Cross-Section of Expected Stock Returns', *Journal of Finance*, 47/2: 427–65.

GOMPERS, **PAUL** A. and **METRICK**, **ANDREW** (2001). 'Institutional Investors and Equity Prices', *Quarterly Journal of Economics*, 116/1: 229–59.

GROSSMAN, **SANFORD** and **STIGLITZ**, **JOSEPH** (1980). 'On the Impossibility of Informationally Efficient Markets', *American Economic Review*, 70: 393–408.

(p. 814) **HALL**, **BRONWYN** (1993). 'Industrial Research during the 1980s: Did the Rate of Return Fall?' *Brookings Papers on Economic Activity. Microeconomics*, 2: 289–343.

HADDAD, **MONA** and **HARRISON**, **ANN** (1993). 'Are There Positive Spillovers From Direct Foreign Investment? Evidence from Panel Data for Morocco', *Journal of Development Economics*, 42/1: 51–74.

HARRIS, **DAVID MORCK**, **RANDALL**, **SLEMROD**, **JOEL**, and **YEUNG**, **BERNARD** (1993). 'Income Shifting in U.S. Multinational Corporations' in *Studies in International Taxation*, Alberto Giovinnini, Glen Hubbard, and Joel Slemrod, eds., NBER Working Paper No. 3924 (Chicago: University of Chicago Press).

HEALY, **PAUL** M. and **WAHLEN**, **JAMES** M. (1999). 'A Review of the Earnings Management Literature, and its Implications for Standard Setting', *Accounting Horizons*, 13/4: 365–83.

HORNSTEIN, **ABIGAIL** (2006). 'Essays on Multinational Enterprises' Capital Budgeting Decisions', PhD thesis, Stern Business School, New York University, New York.

JOHN, **KOSE LITOV**, **LUBOMIR** and **YEUNG**, **BERNARD** (2008). 'Corporate Governance and Managerial Risk Taking: Theory and Evidence', *Journal of Finance*, 62, forthcoming.

JOHNSON, **SIMON**, **LA PORTA**, **RAFAEL**, **LOPEZ DE SILANES**, **FLORENCIO** and **SHLEIFER**, **ANDREI** (2000). 'Tunneling', *American Economic Review*, 90/2: 22–7.

JONES, **GEOFFREY** and **KHANNA**, **TARUN** (2006). 'Bringing History (Back) into International Business', *Journal of International Business Studies*, 37/4: 453–68.

KOKKO, **ARI** (1994). 'Technology, Market Characteristics, and Spillovers', *Journal of Development Economics*, 43/2: 279–93.

LEVINSOHN, **JAMES** and **PETRIN**, **AMIL** (2003). 'Estimating Production Functions Using Inputs to Control for Unobservables', *Review of Economic Studies*, 70/2: 317–42.

LINDENBERG, E. B. and **ROSS**, S. A. (1981). 'Tobin's *q* Ratio and Industrial Organization', *Journal of Business (New York)*, 54/1: 1–32.

MACKINLEY, A. **CRAIG**. (1997). 'Event Studies in Economics and Finance,' *Journal of Economic Literature*, 35/1: 13–39.

MORCK, **RANDALL** and **NAKAMURA**, **MASAO** (1999). 'Banks and Corporate Control in Japan', *Journal of Finance*, 54/1: 319–40.

—— **SHLEIFER**, **ANDREI** and **VISHNY**, **ROBERT** (1988). 'Management Ownership and Market Valuation: An Empirical Analysis', *Journal of Financial Economics*, 20/1–2: 293–315.

——**STANGELAND**, **DAVID** and **YEUNG**, **BERNARD** (2000). 'Inherited Wealth, Corporate Control, and Economic Growth', in Randall Morck (ed.), *Concentrated Corporate Ownership*, NBER Working Paper No. 6814 (1998) (Chicago: University of Chicago Press), 319–69.

——**WOLFENZON**, **DANIEL** and **YEUNG**, **BERNARD** (2005). 'Corporate Governance, Economic Entrenchment and Growth', *Journal of Economics Literature*, 43/3: 655–720.

——and **YEUNG**, **BERNARD** (1991). 'Why Investors Value Multinationality', *Journal of Business (Chicago)*, 64/2: 165–187.

—— ——(1992). 'Internalization: An Event Study Test', *Journal of International Economics*, 33/1–2: 41–56.

—— ——(2003) 'Why Firms Diversify: Internalization vs. Agency Behavior', in John Hand and Baruch Lev (eds.), *Intangible Assets* (Oxford: Oxford University Press), 269–302.

—— ——(2007). ‘History in Perspective: Comment on Jones and Khanna “Bringing History (Back) into International Business”’, *Journal of International Business Studies*, 38/2: 357–60.

—— ——and **YU**, **WAYNE** (2000). ‘The Information Content of Stock Markets: Why Do Emerging Markets Have Synchronous Stock Price Movements?’, *Journal of Financial Economics*, 59/1–2: 215–60.

—— ——and **ZHAO**, **MINYUAN** (2008). ‘Perspectives on China's Outward Foreign Direct Investment’, *Journal of International Business Studies*, 39 (14 Feb. outline): 1–14.

(p. 815) **OLLEY**, **STEVEN** and **PAKES**, **ARIEL** (1996).‘The Dynamics of Productivity in the Telecommunications Equipment Industry’, *Econometrica*, 64/6: 1263–97.

PETERSEN, **MITCHELL** (2005). ‘Estimating Standard Errors in Financial Panel Data Sets: Comparing Approaches’, NBER Working Paper No. 11280 (Cambridge, Mass.: National Bureau of Economic Research).

RUGMAN, **ALAN** M., and OH, **CHANG HOON** (2007). ‘Methodological Issues in the Measurement of Multinationality’, Working Paper.

SHLEIFER, A. and **VISHNY**, R. W. (1997). ‘The Limits of Arbitrage’, *Journal of Finance*, 52/1: 35–55.

VILLALONGA, **BELEN** (2004). ‘Does Diversification Cause the Diversification Discount?’ *Financial Management*, 33/2: 5–27.

Notes:

(1.) We should notice the differences between earnings and profits. Conceptually, if we knew the right rental expenses to all capitals, profits would be equal to ‘operating income after taxes’ + ‘0interest expenses,’ + ‘depreciation, rental, amortization, and depletion expenses’ − ‘total capital rental expenses.’ Unfortunately, we do not readily know the right total rental expenses. The accounting profitability measure, ‘operating income after taxes’, proxies for shareholders’ earnings after tax and depreciation expenses.

(2.) Accounting tax expenses and actual tax expenses are often very different.

(3.) Linderberg and Ross (1981) provides the estimation details.

(4.) Morck, Wolfenzon, and Yeung (2005) discuss the reasons why investor rights protection may be weak in some countries.

Randall Morck

Randall Morck is Stephen A. Jarislowski Distinguished Chair in Finance and University Professor at the University of Alberta and a Research Associate of the National Bureau of Economic Research. He has published widely on corporate valuations, busi-

ness groups, family firms and corporate lobbying in the Review of Economics and Statistics, Journal of Finance, Journal of Economic Literature and other leading journals.

Bernard Yeung

Bernard Yeung is Stephen Riady Distinguished Professor and Dean at NUS Business School, National University of Singapore.