In **Table 3**, we report various correlations of the no-growth and PVGO P/Es. The no-growth and PVGO components have a correlation of 0.363, but this correlation has only a small effect on total P/E variation because of the low volatility of no-growth P/E values. Thus, most of the variation in the total P/E is caused by growth opportunities, and not surprisingly, the PVGO P/E and the total P/E are highly correlated, at 0.998. Both the growth P/E and the total P/E decrease when risk-free rates and earnings growth increase. The correlation of the total P/E with earnings growth is particularly strong at –0.766. High earnings growth by itself increases earnings, which is the denominator of the P/E, and causes P/Es to decrease, resulting in the high negative correlation between earnings growth and the P/E. But another discount rate effect occurs because high earnings growth causes discount rates to significantly increase (see Table 1). This effect also causes P/Es to decrease. High payout ratios, as expected, are positively correlated with the P/E at 0.713. Finally, the latent factor, f, is negatively correlated with the P/E because it is only a discount rate factor: By construction, P/Es are high when f is low.

Table 3. Correlation of Growth (PVGO) and No-Growth Components of the P/E

	No Growth P/E	PVGO P/E
PVGO P/E:	0.363	
Data P/E:	0.421	0.998
rf	–0.353	–0.426
g	–0.051	–0.766
po	–0.292	0.713
ip	0.114	–0.303
term	0.027	0.390
f	–0.903	–0.538

Conclusion

We decomposed the P/E into a no-growth component (the perpetuity value of future earnings held constant with full payout) and a component termed PVGO that reflects the growth opportunities and real options a firm has to invest in the future. We valued both components in a dynamic stochastic environment where risk premiums and earnings growth are stochastic. We found that discount rates exhibit significant variation: 27.5 percent of the variation in total returns is caused by persistent, time-varying expected return components. However, although the variation of discount rates is large, these rates are highly mean reverting. The result is that the no-growth value of earnings exhibits relatively little volatility. The PVGO component dominates; it accounts for the bulk of the level and variation of P/Es in the data: Approximately 80 percent of the level and 95 percent of the variance of P/Es are a result of time-varying growth opportunities.

We thank Geert Bekaert, Sigbjørn Berg, and Tørres Trovik for helpful discussions.

Appendix A

Here, we provide the coefficients a_i and b_i and the definition of the P/E as used by the S&P 500. All the formulas are derived in the online appendix at www.columbia.edu/~aa610.

Full and No-Growth P/Es. The coefficients a_i and b_i for the P/E in Equation 12 are given by

$$a_{i+1} = -\delta_0 + a_i + \left(\mathbf{e}_2 + \boldsymbol{b}_i\right)' \boldsymbol{\mu} + \frac{1}{2}\left(\mathbf{e}_2 + \boldsymbol{b}_i\right)' \Sigma\Sigma' \left(\mathbf{e}_2 + \boldsymbol{b}_i\right)$$

and

$$b_i = -\boldsymbol{\delta}_1 + \Phi'\left(\mathbf{e}_2 + \boldsymbol{b}_i\right),$$

where $\mathbf{e}_n$ is a vector of 0s with a 1 in the nth position. The initial conditions are

$$a_1 = -\delta_0 + \left(\mathbf{e}_2 + \mathbf{e}_3\right)' \boldsymbol{\mu} + \frac{1}{2}\left(\mathbf{e}_2 + \mathbf{e}_3\right)' \Sigma\Sigma' \left(\mathbf{e}_2 + \mathbf{e}_3\right)$$

and

$$b_1 = -\boldsymbol{\delta}_1 + \Phi'\left(\mathbf{e}_2 + \mathbf{e}_3\right).$$

The coefficients in the no-growth P/E, P/E_t^{ng}, in Equation 13 are given by

$$a_{i+1}^* = -\delta_0 + a_i^* + \boldsymbol{b}_i^{*\prime} \boldsymbol{\mu} + \frac{1}{2} \boldsymbol{b}_i^{*\prime} \Sigma\Sigma' \boldsymbol{b}_i^*$$

and

$$\boldsymbol{b}_{i+1}^* = -\boldsymbol{\delta}_1 + \Phi' \boldsymbol{b}_i^*,$$

where a_i^* and b_i^* have initial values $a_i^* = -\delta_0$ and $\boldsymbol{b}_i^* = -\delta_1$.

Data. The P/E defined by Standard & Poor's is the market value at time t divided by trailing 12-month earnings reported from t to $t-1$. To back out earnings growth from P/Es, we used the following transformation:

$$\begin{aligned}\exp\left(g_{t+1}\right) &= \frac{EA_{t+1}}{EA_t} \\ &= \left(\frac{P/E_t}{P/E_{t+1}}\right)\left(\frac{P_{t+1}}{P_t}\right),\end{aligned}$$

where P_{t+1}/P_t is the price gain (capital gain) on the market from t to $t+1$.

The dividend yield reported by Standard & Poor's is also constructed from trailing 12-month summed dividends. We computed the log payout ratio from the ratio of the dividend yield, $dy_t = D_t/P_t$, to the inverse P/E:

$$\begin{aligned}\exp\left(po_t\right) &= \frac{dy_t}{1/\left(P/E\right)_t} \\ &= \frac{D_t}{EA_t}.\end{aligned}$$

For the risk-free rate, r_t^f, we used one-year zero-coupon yields expressed as a log return, which we obtained from the Fama Files derived from the CRSP U.S. Government Bond Files. For the macro variables, we expressed industrial production growth, *ip*, as a log year-on-year growth rate for which we used the industrial production index from the St. Louis Federal Reserve. We defined the term spread, *term*, as the difference in annual yields between 10-year and 1-year government bonds, which we obtained from CRSP.

Bibliography

Ang, A., and J. Liu. 2001. "A General Affine Earnings Valuation Model." *Review of Accounting Studies*, vol. 6, no. 4 (December):397–425.

Bakshi, G., and Z. Chen. 2005. "Stock Valuation in Dynamic Economies." *Journal of Financial Markets*, vol. 8, no. 2 (May):115–151.

Bansal, R., and A. Yaron. 2004. "Risk for the Long Run: A Potential Resolution of Asset Pricing Puzzles." *Journal of Finance*, vol. 59, no. 4 (August):1481–1509.

Bekaert, G., E. Engstrom, and S.R. Grenadier. 2010. "Stock and Bond Returns with Moody Investors." *Journal of Empirical Finance*, vol. 17, no. 5 (December):867–894.

Bodie, Z., A. Kane, and A.J. Marcus. 2009. *Investments*. 8th ed. New York: McGraw-Hill/Irwin.

Campbell, J.Y., and R.J. Shiller. 1988. "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors." *Review of Financial Studies*, vol. 1, no. 3 (July):195–228.

Claus, J., and J. Thomas. 2001. "Equity Premia as Low as Three Percent? Evidence from Analysts' Earnings Forecasts for Domestic and International Stock Markets." *Journal of Finance*, vol. 56, no. 5 (October):1629–1666.

Fama, E.F., and K.R. French. 2002. "The Equity Premium." *Journal of Finance*, vol. 57, no. 2 (April):637–659.

Jagannathan, R., E.R. McGrattan, and A. Scherbina. 2000. "The Declining U.S. Equity Premium." *Federal Reserve Bank of Minneapolis Quarterly Review*, vol. 24, no. 4 (Fall):3–19.

Miller, M.H., and F. Modigliani. 1961. "Dividend Policy, Growth, and the Valuation of Shares." *Journal of Business*, vol. 34, no. 4 (October):411–433.

van Binsbergen, J., and R.S.J. Koijen. 2010. "Predictive Regressions: A Present-Value Approach." *Journal of Finance*, vol. 65, no. 4 (August):1439–1471.

Vuolteenaho, T. 2002. "What Drives Firm-Level Stock Returns?" *Journal of Finance*, vol. 51, no. 1 (February):233–264.

Long-Term Stock Returns Unshaken by Bear Markets

Jeremy J. Siegel

Russell E. Palmer Professor of Finance

Wharton School of the University of Pennsylvania

The first Equity Risk Premium Forum, sponsored by CFA Institute, was held on 8 November 2001, not long after the September 11 terrorist attacks and coincident with the first of two devastating bear markets in the first decade of the new millennium. At the time of the first forum, stocks had already fallen by more than half of what would become a nearly 50 percent decline from the peak reached in March 2000 to the low in October 2002. Over the four years after the low, the equity market recovered all of its losses and moved into new all-time-high territory. But the 2008 financial crisis precipitated a more severe bear market than 2000–2002 and the worst since the Great Crash of 1929–1932. In the financial crisis, the S&P 500 Index plunged 57 percent from October 2007 to March 2009 and non-U.S. equity markets fell more than 60 percent. As of this writing (May 2011), stocks worldwide have made a strong recovery and are now within 15 percent of their all-time highs.

Nevertheless, the returns for stocks during the past decade have not been good. Since the first forum was held, the stock returns on the broad-based Russell 3000 Index have averaged 5.6 percent per year; when offset against 2.5 percent annual inflation, the real return is only a little more than 3 percent per year. The nominal yields on Treasuries have averaged 2.2 percent during the decade, leaving a real return of –0.2 percent per year on those instruments. These returns mean that the realized equity premium, or excess return of stocks over T-bills, has been between 3 percent and 3.5 percent. These numbers are not far from the predictions that I made at the first forum 10 years ago. At the time, I expected real returns of equities to be 4.5–5.5 percent and an equity risk premium of 2 percent (200 bps).

As I read through my analysis from 10 years ago, I could see that the main reason I overestimated the real return on stocks was that I overestimated the price-to-earnings ratio (P/E) that investors would pay for stocks. There were good reasons back then for why the P/E of stocks should be higher than its historical average of 15, a level computed from earnings data extending back to 1871, and should instead range between 20 and 25. First, the sharp decline in transaction costs caused by the development of index funds and the plunge in commission prices gave investors a much more favorable realized risk–return trade-off than they received in earlier years. Another reason I conjectured that the P/E would be higher than its historical level was the decline in the volatility of real economy variables. This increase in macroeconomic stability was termed by economists at the time as the "Great Moderation."

Of course, the 2007–09 recession dispelled the idea that the business cycle had been tamed. It is my opinion that the Great Moderation was indeed real, but the long period of macroeconomic stability led to an excessive decline in risk premiums, particularly in housing-related securities. So, when real estate prices unexpectedly fell, the entire financial system came crashing down. The financial crisis greatly increased the risk aversion of investors, and that result brought the P/E back down to historical levels and led to the poor stock returns of the past decade.

This observation can be confirmed by examining the data. When the first forum was held in November 2001, the reported earnings of the S&P 500 over the preceding 12 months were $15.90, which yielded a P/E of 36.77. The trailing 12-month earnings on the S&P 500 at the time of the second forum in January 2011 were $81.47, more than a threefold increase. Yet the index itself was up by only 30 percent, and the P/E had fallen to 16.66. If the P/E had fallen only to 22.5, the middle of my valuation range, stock returns would have been about 3 percentage points per year higher.

Another prediction that did not materialize was my estimate of future bond yields. I believed that the real yields on bonds would remain between 3 and 4 percent, the level that prevailed when Treasury Inflation-Protected Securities (TIPS) were first issued in 1997. I also believed that the realized bond returns in the period after World War II (WWII) were biased downward because of the unanticipated inflation from the late 1960s through the early 1980s. So, I did not consider historical returns on bonds; instead, I used the current yield on TIPS in making my forecast for future bond yields.

Instead, real yields fell dramatically, especially in the wake of the financial crisis. As of early 2011, 10-year TIPS yields are less than 1 percent and 5-year TIPS yields are negative. The two primary reasons for the drop in real yields are the slowdown in economic growth and the increase in the risk aversion of the investing public, which, in turn, is caused by both the aging of the population and the shocks associated with the financial crisis. The decline in inflation has caused the yields on nominal bonds to drop even more, generating very large realized returns for nominal bond investors. Over the last decade, realized bond returns were 4.7 percent per year after inflation, swamping stock returns. Over the past 20 years, realized bond returns were 6.0 percent per year, 1 percentage point less than the 7.0 percent real returns on stocks.

Updated Return Data

Table 1 shows historical returns for stocks, bonds, and T-bills from 1802 through April 2011. The past decade has shaved one-tenth of a percent off of the annualized real returns on stocks from 1802 through April 2001; three-tenths off of the equity returns from 1871, which is when the Cowles Foundation for Research in Economics data became available; and five-tenths off of the real return since 1926, which is the period that Ibbotson and Sinquefield popularized in their research.[1] Over all long-term periods, the real return on stocks remained in the 6–7 percent range. Over the past 30 years, the real annual return on stocks has been 7.9 percent, and over the past 20 years, the real return has been 7.0 percent. In fact, the numbers that now fill the table are almost identical to those that I calculated when I started my research in the late 1980s. In essence, the poor returns of the past 10 years just offset the very high returns of the previous decade.

Table 2 summarizes some of the important statistics about the equity market, such as the P/E, earnings growth, and dividend growth, for 1871–April 2011. The average P/E has changed very little over the past decade. In the version of Table 2 prepared for the 2001 forum, the average P/E was 14.45; adding the subsequent 10 years of data increased it by 0.06 to 14.51. The earnings yield, which is the reciprocal of the P/E, obviously also changes little.

One important issue that was in contention in the first forum is still debated today. Finance theory, particularly that of Modigliani and Miller (M&M), predicts that when the dividend payout ratio declines, the dividend yield will also decline, but this decline will be offset by an increase in the growth rate of future earnings and dividends.[2] Cliff Asness, at the 2001 forum, and Rob Arnott, at the most recent forum, cite research, which they performed together, that suggests that a lower payout ratio, in contrast to what finance theory would predict, does not actually lead to faster earnings growth.[3] At the first forum, I claimed that this finding was a result of the cyclical behavior of earnings. Asness and Arnott claimed to have run further tests to contest this point. Notwithstanding their results, my data clearly show that over long periods of time, the payout ratio is inversely correlated with dividend and earnings growth as predicted by finance theory.

---

[1]Roger G. Ibbotson and Rex A. Sinquefield, "Stocks, Bonds, Bills, and Inflation: Year-by-Year Historical Returns (1926–1974)," *Journal of Business*, vol. 49, no. 1 (January 1976):11–47.

[2]Franco Modigliani and Merton H. Miller, "The Cost of Capital, Corporation Finance and the Theory of Investment," *American Economic Review*, vol. 48, no. 3 (June 1958):261–297.

[3]Robert D. Arnott and Clifford S. Asness, "Surprise! Higher Dividends = Higher Earnings Growth," *Financial Analysts Journal*, vol. 59, no. 1 (January/February 2003):70–87.

Table 1. Historical Returns for Stocks, Bonds, and T-Bills, 1802–April 2011

	Real Return						Stocks' Excess Return Over			
	Stocks		Bonds		T-Bills		Bonds		T-Bills	
	Geometric	Arithmetic	Geometric	Arithmetic	Geometric	Arithmetic	Geometric	Arithmetic	Geometric	Arithmetic
Periods										
1802–2011	6.7%	8.2%	3.6%	3.9%	2.7%	2.9%	3.1%	4.3%	3.9%	5.3%
1870–2011	6.5	8.2	3.0	3.3	1.6	1.7	3.5	4.9	4.9	6.5
Major subperiods										
1802–1870	7.0%	8.3%	4.8%	5.1%	5.1%	5.4%	2.2%	3.2%	1.9%	2.9%
1871–1925	6.6	7.9	3.7	3.9	3.2	3.3	2.9	4.0	3.5	4.7
1926–2011	6.4	8.4	2.5	3.0	0.6	0.7	4.0	5.4	5.8	7.7
After World War II										
1946–2011	6.4%	8.3%	1.8%	2.2%	0.5%	0.6%	4.6%	6.0%	6.0%	7.6%
1946–1965	10.0	11.4	–1.2	–1.0	–0.8	–0.7	11.2	12.3	10.9	12.1
1966–1981	–0.4	1.4	–4.2	–3.9	–0.2	–0.1	3.8	5.2	–0.2	1.5
1982–1999	13.6	14.3	8.5	9.3	2.9	2.9	5.1	5.0	10.7	11.4
1982–2011	7.9	9.1	7.5	7.9	1.8	1.7	0.4	1.2	6.1	7.4
1991–2011	7.0	8.5	6.0	6.3	0.9	0.9	0.9	2.1	6.1	7.6
2001–2011	0.8	2.8	4.7	4.7	–0.3	–0.3	–4.0	–1.9	1.1	3.0

Table 2. Historical Equity Market Statistics, 1871–April 2011

	Real Stock Return	Average P/E	Inverse of Average P/E	Real Earnings Growth	Real Dividend Growth	Dividend Yield	Real Capital Gains	Average Payout Ratio
1871–2011	6.51%	14.51	6.89%	1.81%	1.22%	4.47%	1.55%	59.92%
1871–1945	6.39	13.83	7.23	0.67	0.74	5.31	1.11	70.81
1946–2011	6.44	15.29	6.54	3.14	1.76	3.50	2.85	47.42

In fact, the evidence in favor of M&M has been strengthened by the addition of the past 10 years of data. In the 1871–1945 data, annual real per share earnings growth was only 0.67 percent per year and the payout ratio averaged nearly 72 percent. In the post-WWII period, real earnings growth was 3.14 percent and the payout ratio was only 47.42 percent.[4]

It is true that adding the past 10 years increases post-WWII real per share dividend growth only marginally because the payout ratio is still declining and has not yet reached a new "steady state" in which dividend growth will increase to the level of earnings growth.

Projections for the Next Decade

I hope a third forum will be held in 2021 so we can look back on our predictions in 2011, either nursing our wounds or congratulating ourselves on our astuteness. Using the current P/E as a basis, I expect real stock returns to be between 6 and 7 percent. But I will not be surprised if they are higher because the same factors that influenced my prediction of P/Es in the range of 20–25 are as operative in 2011 as they were at the time of the first forum in 2001.

Real bond returns are on track to be much lower. Ten-year TIPS are now yielding about 1 percent, so the excess returns of stocks over bonds should be in the 5–6 percent range, which is higher than the historical average. And the bias, if any, will be toward a higher equity premium if real bond yields rise from their extremely low levels, as I think they should. In short, relative to bonds, stocks look extraordinarily attractive, and I expect stock investors will look back a decade from now with satisfaction.

[4]Note that the 3.14 percent growth rate is more than 1 percentage point higher than the post-WWII real earnings growth rate presented at the first forum; the addition of the past 10 years also reduces the post-WWII average payout ratio from 50.75 percent to 47.42 percent.

The Equity Premium Puzzle Revisited

Rajnish Mehra

E.N. Basha Arizona Heritage Chair Professor of Finance and Economics, Arizona State University

Research Associate, NBER

In the two and a half decades since "The Equity Premium: A Puzzle" (Mehra and Prescott 1985) was published, attempts to successfully account for the equity premium have become a major research impetus in finance and economics. In an effort to reconcile theory with observations, I will elaborate on the appropriateness of three crucial abstractions in that article. In particular, I will argue that our finding (i.e., the premium for bearing nondiversifiable aggregate risk is small) is not inconsistent with the average equity premium over the past 120 years.

The three abstractions that I address here are

- using T-bill prices as a proxy for the expected intertemporal marginal rate of substitution of consumption;
- ignoring the difference between borrowing and lending rates (a consequence of agent heterogeneity and costly intermediation);
- abstracting from life-cycle effects and borrowing constraints on the young.

I examine each of these in detail below.

Using T-Bill Prices as a Proxy for the Expected Intertemporal Marginal Rate of Substitution of Consumption

An assumption implicit in Mehra and Prescott (1985) is that agents use both equity and the riskless asset to smooth consumption intertemporally. This assumption is a direct consequence of the first-order condition (see Equation 1) for the representative household in our model. It implies that agents save by optimally allocating resources between equity and riskless debt.

$$0 = E_t\left[\frac{U_c(c_{t+s})}{U_c(c_t)}\left(r^e_{t,t+s} - r^d_{t,t+s}\right)\right]. \tag{1}$$

Author Note: This paper draws widely on my collaborations with George Constantinides, John Donaldson, and Edward Prescott. Quite independently of our joint work, they have made substantial contributions to the literature on the equity premium puzzle. Consequently, the views expressed in this paper do not necessarily reflect their views.

Equation 1 is the standard asset-pricing equation in macroeconomics and finance. $U_c(c_{t+s})$ is the marginal utility of consumption at time $t + s$; $r^e_{t,t+s}$ and $r^d_{t,t+s}$ are, respectively, the return on equity and the return on the riskless asset over the period t, $t + s$; and E_t is the expectation conditional on the agent's information set at time t.

If the results from the model are to be compared with data, it is crucial to identify the empirical counterpart of the riskless asset that is actually used by agents to smooth consumption. In Mehra and Prescott (1985), we used the highly liquid T-bill rate, corrected for expected inflation, as a proxy for this asset. But one might ask: Is it reasonable to assume that T-bills are an appropriate proxy for the riskless asset that agents use to save for retirement and smooth consumption? Do households actually hold T-bills to finance their retirement? *Only if this question is empirically verified would it be reasonable to equate their expected marginal rate of substitution of consumption to the rate of return on T-bills.*

This question cannot be answered in the abstract without reference to the asset holdings of households, so a natural next step is to examine the assets held by households. **Table 1** details these holdings for U.S. households. The four big asset-holding categories of households are tangible assets, pension and life insurance holdings, equity (both corporate and noncorporate), and debt assets.

Table 1. Household Assets and Liabilities as a Fraction/ Multiple of GDP
(average of 2000 and 2005)

Assets (GDP)		Liabilities (GDP)	
Asset	GDP (×)	Liability	GDP (×)
Tangible household	1.65	Liabilities	0.7
Corporate equity	0.85	Net worth	4.15
Noncorporate equity	0.5		
Pension and life insurance reserves	1.0		
Debt assets	0.85		
Total	4.85		4.85

In 2000, privately held government debt was only 0.30 times GDP, a third of which was held by foreigners. The amount of interest-bearing government debt with maturity less than a year was only 0.085 times GDP, which is a small fraction of total household net worth. Virtually no T-bills are directly owned by households.[1] Approximately one-third of the T-bills outstanding are held by foreign central banks, and two-thirds are held by U.S. financial institutions.

[1] See Table B-89, *Economic Report of the President* (2005).

Although large amounts of debt assets are held, most of these are in pension fund and life insurance reserves. Some are in demand deposits, for which free services are provided. Most government debt is held indirectly; a small fraction is held as savings bonds.

Thus, much of intertemporal saving is in debt assets, such as annuities and mortgage debt, held in retirement accounts and as pension fund reserves. Other assets, not T-bills, are typically held to finance consumption in retirement. *Hence, T-bills and short-term debt are not reasonable empirical counterparts to the risk-free asset priced in Equation 1*, and it would be inappropriate to equate the return on these assets to the expected marginal rate of substitution for an important group of agents.

An inflation-indexed, default-free bond portfolio with a duration similar to that of a well-diversified equity portfolio would be a reasonable proxy for a risk-free asset used for consumption smoothing.[2] For most of the 20th century, equity has had an implied duration of about 25 years, so a portfolio of TIPS (Treasury Inflation-Protected Securities) of a similar duration would be a reasonable proxy.

Because TIPS have only recently (1997) been introduced in U.S. capital markets, it is difficult to get accurate estimates of the mean return on this asset class. The average return for the 1997–2005 period is 3.7 percent. An alternative (though imperfect) proxy would be to use the returns on indexed mortgages guaranteed by Ginnie Mae (Government National Mortgage Association) or issued by Fannie Mae (Federal National Mortgage Association). I conjecture that if these indexed default-free securities are used as a benchmark, the equity premium will be closer to 4 percent than to the 6 percent equity premium relative to T-bills. By using a more appropriate benchmark for the riskless asset, I can account for 2 percentage points of the "equity premium."

Ignoring the Difference between Borrowing and Lending Rates

A major disadvantage of the homogeneous household construct is that it precludes the modeling of borrowing and lending among agents. In equilibrium, the shadow price of consumption at date $t + 1$ in terms of consumption at date t is such that the amount of borrowing and lending is zero. However, there is a large amount of costly intermediated borrowing and lending between households, and as a consequence, borrowing rates exceed lending rates. When borrowing and lending rates differ, a question arises: Should the equity premium be measured relative to the riskless borrowing rate or the riskless lending rate?

[2]McGrattan and Prescott (2003) use long-term high-grade municipal bonds as a proxy for the riskless security.

To address this question, Mehra, Piguillem, and Prescott (2011) constructed a model that incorporates agent heterogeneity and costly financial intermediation. The resources used in intermediation (3.4 percent of GNP) and the amount intermediated (1.7 percent of GNP) imply that the average household borrowing rate is at least 2 percentage points higher than the average household lending rate. Relative to the level of the observed average rates of return on debt and equity securities, this spread is far from being insignificant and cannot be ignored when addressing the equity premium.

In this model,[3] a subset of households both borrow money and hold equity. Consequently, a no-arbitrage condition is that the return on equity and the borrowing rate are equal (5 percent). The return on government debt, the household lending rate, is 3 percent. If I use the conventional definition of the equity premium—the return on a broad equity index less the return on government debt—I would erroneously conclude that in this model, the equity premium is 2 percent. The difference between the government borrowing rate and the return on equity is not an equity premium; it arises because of the wedge between borrowing and lending rates. Analogously, if borrowing and lending rates for equity investors differ, and they do in the U.S. economy, the equity premium should be measured relative to the investor borrowing rate rather than the investor lending rate (the government's borrowing rate). Measuring the premium relative to the government's borrowing rate artificially increases the premium for bearing aggregate risk by the difference between the investor's borrowing and lending rates.[4] If such a correction is made to the benchmark discussed earlier, the "equity premium" is further reduced by 2 percentage points. Thus, I have accounted for 4 percentage points of the equity premium reported in Mehra and Prescott (1985) by factors other than aggregate risk.

Abstracting from Life-Cycle Effects and Borrowing Constraints on the Young

In Constantinides, Donaldson, and Mehra (2002), we examined the impact of life-cycle effects, such as variable labor income and borrowing constraints, on the equity premium. We illustrated these ideas in an overlapping-generations exchange economy in which consumers live for three periods. In the first period, a period of human capital acquisition, the consumer receives a relatively low endowment income. In the second period, the consumer is employed and receives wage income subject to large uncertainty. In the third period, the consumer retires and consumes the assets accumulated in the second period.

[3]There is no aggregate uncertainty in our model.

[4]For a detailed exposition of this and related issues, see Mehra and Prescott (2008).

In the article, we explored the implications of a borrowing constraint by deriving and contrasting the stationary equilibriums in two versions of the economy. In the *borrowing-constrained* version, the young are prohibited from borrowing and from selling equity short. The *borrowing-unconstrained* economy differs from the borrowing-constrained one only in that the borrowing constraint and the short-sale constraint are absent.

The attractiveness of equity as an asset depends on the correlation between consumption and equity income. Because the marginal utility of consumption varies inversely with consumption, equity will command a higher price (and consequently, a lower rate of return) if it pays off in states when consumption is high and vice versa.[5]

A key insight of ours in the article is that as the correlation of equity income with consumption changes over the life cycle of an individual, so does the attractiveness of equity as an asset. Consumption can be decomposed into the sum of wages and equity income. Young people looking forward at the start of their lives have uncertain future wage and equity income; furthermore, the correlation of equity income with consumption will not be particularly high as long as stock and wage income are not highly correlated. This is empirically the case, as documented by Davis and Willen (2000). Equity will, therefore, be a hedge against fluctuations in wages and a "desirable" asset to hold as far as the young are concerned.

The same asset (equity) has a very different characteristic for the middle-aged. Their wage uncertainty has largely been resolved. Their future retirement wage income is either zero or deterministic, and the innovations (fluctuations) in their consumption occur from fluctuations in equity income. At this stage of the life cycle, equity income is highly correlated with consumption. Consumption is high when equity income is high, and equity is no longer a hedge against fluctuations in consumption; hence, for this group, equity requires a higher rate of return.

The characteristics of equity as an asset, therefore, change depending on the predominant holder of the equity. Life-cycle considerations thus become crucial for asset pricing. If equity is a desirable asset for the marginal investor in the economy, then the observed equity premium will be low relative to an economy where the marginal investor finds it unattractive to hold equity. The *deus ex machina* is the stage in the life cycle of the marginal investor.

[5]This is precisely the reason why high-beta stocks in the simple capital asset pricing model framework have a high rate of return. In that model, the return on the market is a proxy for consumption. High-beta stocks pay off when the market return is high—that is, when marginal utility is low and, hence, their price is (relatively) low and their rate of return high.

We argued that the young, who should be holding equity in an economy without frictions, are effectively shut out of this market because of borrowing constraints. The young are characterized by low wages; ideally, they would like to smooth lifetime consumption by borrowing against future wage income (consuming a part of the loan and investing the rest in higher return equity). However, they are prevented from doing so because human capital alone does not collateralize major loans in modern economies for reasons of moral hazard and adverse selection.

Therefore, in the presence of borrowing constraints, equity is exclusively priced by middle-aged investors because the young are effectively excluded from the equity markets and a high equity premium is thus observed. If the borrowing constraint is relaxed, the young will borrow to purchase equity, thereby raising the bond yield. The increase in the bond yield induces the middle-aged to shift their portfolio holdings from equities to bonds. The increase in demand for equity by the young and the decrease in demand for equity by the middle-aged work in opposite directions. On balance, the effect is to increase both the equity and the bond return, while shrinking the equity premium.

The results suggest that, depending on the parameterization, between 2 and 4 percentage points of the observed equity premium can be accounted for by incorporating life-cycle effects and borrowing constraints.

Conclusion

I have argued that using an appropriate benchmark for the risk-free rate, accounting for the difference between borrowing and lending rates, and incorporating life-cycle features can account for the equity premium. That this can be accomplished without resorting to risk supports the conclusion of Mehra and Prescott (1985) that the premium for bearing systematic risk is small.

My projection for the equity premium is that at the end of the next decade, it will be higher than that observed in the past. During the next 10 years, the ratio of the retired population to the working-age population will increase. These retired households, in an attempt to hedge against outliving their assets, will likely rebalance their portfolios by substituting annuity-like products for equity. Because, in equilibrium, all assets must be held, this substitution will lead to an increase in the expected equity premium. Consequently, during this adjustment process, the realized equity premium will probably be lower than the historical average.

REFERENCES

Constantinides, G.M., J.B. Donaldson, and R. Mehra. 2002. "Junior Can't Borrow: A New Perspective on the Equity Premium Puzzle." *Quarterly Journal of Economics*, vol. 117, no. 1 (February):269–296.

Davis, Stephen J., and Paul Willen. 2000. "Using Financial Assets to Hedge Labor Income Risk: Estimating the Benefits." Working paper, University of Chicago.

McGrattan, E.R., and E.C. Prescott. 2003. "Average Debt and Equity Returns: Puzzling?" *American Economic Review*, vol. 93, no. 2 (May):392–397.

Mehra, R., and E.C. Prescott. 1985. "The Equity Premium: A Puzzle." *Journal of Monetary Economics*, vol. 15, no. 2 (March):145–161.

———. 2008. "Non-Risk-Based Explanations of the Equity Premium." In *Handbook of Investments: The Handbook of the Equity Risk Premium*. Edited by R. Mehra. Amsterdam: Elsevier.

Mehra, R., F. Piguillem, and E.C. Prescott. 2011. "Costly Financial Intermediation in Neoclassical Growth Theory." *Quantitative Economics*, vol. 2, no. 1 (March):1–36.

RESEARCH FOUNDATION CONTRIBUTION FORM

☑ **Yes,** I want the Research Foundation to continue to fund innovative research that advances the investment management profession. Please accept my tax-deductible contribution at the following level:

Contributing Research Fellow ....$25,000 to $49,999
Research Fellow ....$10,000 to $24,999
Contributing Donor ....$1,000 to $9,999
Donor .... Up to $999

I would like to donate $______________.

☐ My check is enclosed (payable to the Research Foundation of CFA Institute).
☐ I would like to donate appreciated securities (send me information).
☐ Please charge my donation to my credit card.
☐ VISA ☐ MC ☐ Amex ☐ Diners ☐ Corporate ☐ Personal

Card Number

___/___
Expiration Date

Name on card PLEASE PRINT

☐ Corporate Card
☐ Personal Card

Signature

☐ This is a pledge. Please bill me for my donation of $______________.

☐ I would like recognition of my donation to be:
☐ Individual donation ☐ Corporate donation ☐ Different individual

PLEASE PRINT NAME OR COMPANY NAME AS YOU WOULD LIKE IT TO APPEAR

PLEASE PRINT ☐ Mr. ☐ Mrs. ☐ Ms. MEMBER NUMBER______________

Last Name (Family Name) First Middle Initial

Title

Address

City State/Province Country ZIP/Postal Code

11ERP

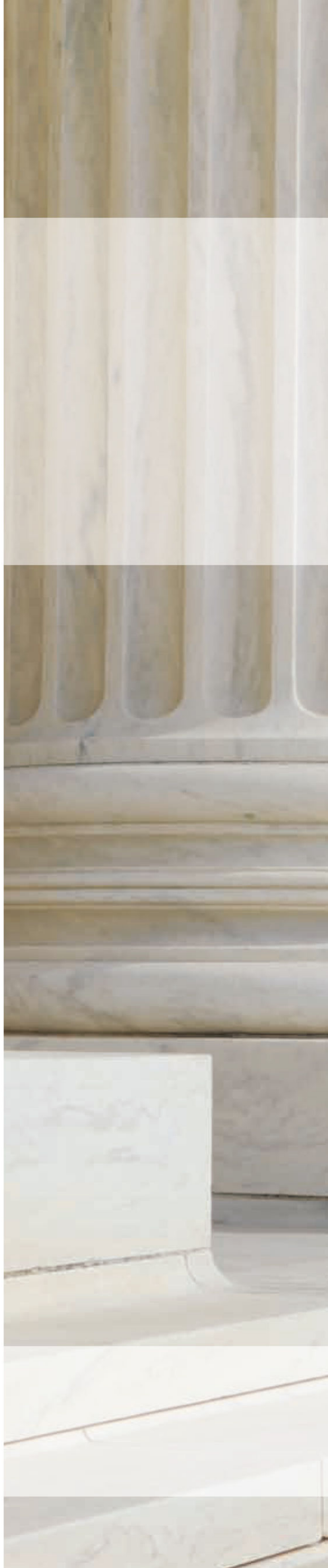
Please mail this completed form with your contribution to:
The Research Foundation of CFA Institute • P.O. Box 2082
Charlottesville, VA 22903-0638 USA

For more on the Research Foundation of CFA Institute, please visit www.cfainstitute.org/about/foundation/.

RESEARCH FOUNDATION OF CFA INSTITUTE

available online at
www.cfapubs.org

9 781934 667446
90000



"An Opening of Minds"

"I think investors are starting to come around to the view that stocks aren't quite as special as they once thought," says Rob Arnott

By Jonathan Barnes

"My career has largely been successful as a consequence of the fact that I love to test ideas," says Rob Arnott, chairman and CEO of Research Affiliates and former editor in chief of the *Financial Analysts Journal*. Arnott's reputation for testing conventional investment wisdom made him one of the key contributors when the Research Foundation of CFA Institute gathered leading academics and practitioners in 2011 to discuss the equity risk premium (ERP), the expected return for equities in excess of a risk-free rate. He delivered a presentation titled "Equity Risk Premium Myths," which was subsequently included in the book *Rethinking the Equity Risk Premium*. In this interview with *CFA Institute Magazine*, Arnott corrects some of the misconceptions about the ERP, argues that "a cult of equities is worshipping a false idol," deconstructs the notion of a risk-free rate, and explains why "our industry, both on the practitioner and on the academic sides, has tremendous inertia, a resistance to new ideas."

Do we need a stronger definition of the equity risk premium?
All too often, the term "equity risk premium" is attached to widely different concepts. It is applied to the historical difference in returns between stocks and bonds—or between stocks and cash—and it is also applied to forward-looking expectational return differences. Really, a risk premium is an *expectational* return, so when we look at historical returns, I think it is important to use different terminology. I prefer the term "historical excess return," not risk premium.

If we turn attention from past to future, the equity risk premium should be the expected incremental return that an investor will likely earn from a willingness to hold stocks instead of bonds or cash. So, one needs to further define one's terms. The risk premium versus bonds and the risk premium versus cash are very different. Today, cash yields nothing; 30-year bonds have yields around 3%.

Which measure is more widely used?
Academia tends to think of the equity risk premium relative to a risk-free rate (never mind that there is nothing that is really risk free in life), and typically that is thought of as a cash yield. A much more relevant measure is equities versus long bonds because they both have a long investment horizon. Cash is very risky for the long-term investor!

ALL TOO OFTEN, THE TERM "EQUITY RISK PREMIUM" IS ATTACHED TO WIDELY DIFFERENT CONCEPTS.

When we look at stocks relative to long bonds, we can do some very simple arithmetic as it relates to expectational returns. Thirty-year bonds have yields around 3%, and the real return as indicated by long-term Treasury Inflation-Protected Securities (TIPS) is 0.5%, give or take.

Stocks produce returns in a real return form because earnings and dividends grow with inflation, plus a real growth kicker. Historically, going back a hundred years, you find earnings and dividends have grown a little less than 1.5% above the rate of inflation. If you add that to the current yield, you get something on the order of a 3.5% expected real return, as against 0.5% for long TIPS. That gives you a 3% risk premium. And that assumes that past rates of growth can continue, given the headwinds from our aging population, as well as our burgeoning debt and deficits.

So when we reframe the definition in terms of forward-looking return expectations for stocks (relative to forward-looking real return expectations for long bonds), we get a comparison of two relatively similar-horizon investments and a comparison that has some real economic meaning. That's my preferred way of thinking about the equity risk premium.

Is more standardization of the ERP needed?
Discussions about the equity risk premium often occur in vague terms: How much more do you expect to earn from a willingness to bear equity market risk? How much more return relative to what? Over what investment horizon? These questions are left ambiguous in all too many examinations of the equity risk premium. If they are defined with any precision, you get much more reasonable apples-with-apples comparisons. Then, you have an ability to examine the underlying assumptions.

There is an annual academic survey of estimates on the equity risk premium in which the ERP is defined as a long-term return against T-bills. But you still have to factor inflation expectations, and on a long-term basis, inflation is anyone's guess, not to mention the future real T-bill yields. So, even with studies that define their terms, if you have a gap in return horizon—cash has a horizon that is measured in weeks or months, stocks have a horizon that is measured in decades—then again, you get into ambiguous comparisons of apples and oranges and a relatively meaningless phenomenon.

Can you explain the myth that the equity risk premium is 5%?
The notion that stocks beat bonds by 5% was embraced in the 1990s by much of the consulting community (and through the consulting community, by much of the plan sponsor community). It is something of a core belief in the practitioner community. This myth is very dangerous because the long-term historical excess return—while not far from 5%—is driven in large measure by a change in valuation multiples for equities. The long-term historical average dividend yield for stocks going back a hundred or more years is about 4%. If the yield now is 2%—a rise in valuation multiples from 25 years of dividends to 50 years of dividends—that is a big change in valuation multiples. So, it creates an inflated historical excess return, which people then translate into an inflated expectational risk premium.

How does your estimate of 3% compare historically?
It's above the historic norms. In 2002, I wrote a paper with Peter Bernstein for the *Financial Analysts Journal* that showed that the reasonable historical equity risk premium—not the excess

return—but what would reasonably have been expected historically for stocks relative to long bonds—was 2.4%.

So, if we are looking at 3% today, that means that right now we have a modestly outsized equity risk premium (if future economic growth matches past growth). It's predicated on negative real yields at the long end of the bond market, so that is a big problem. If you are looking at anemic real returns on bonds (and less-anemic real returns on stocks), you get a positive risk premium through the unfortunate path of generally dismal returns.

I THINK THE MYTHS ARE A CONSEQUENCE OF INERTIA. OUR INDUSTRY, BOTH ON THE PRACTITIONER AND ON THE ACADEMIC SIDES, HAS TREMENDOUS INERTIA, A RESISTANCE TO NEW IDEAS.

Another myth is that the ERP is static over time, companies, and markets. Can you say more?

There are respected academics who build their theories on the notion that the equity risk premium must be static. Yet, on the other hand, there are those who argue that the equity risk premium varies from one stock to another. If it varies from one stock to another, why shouldn't it vary from one month or year to another? The notion of a static equity risk premium is another unfortunate myth.

The risk premium is really a function of pricing. When bond yields are high, the risk premium can get very skinny indeed. Ever so briefly in 2000, you could buy TIPS, *long-term* TIPS, extending out 20–30 years that had a yield of over 4%. I believe the top was 4.3%. A 4.3% real return guaranteed with full faith and credit of the U.S. Treasury is a marvelous default risk–free return. To have that available in bonds at a time when stocks had a yield of 1% is really quite breathtaking. So, what we find is that the risk premium is dynamic. It changes over time.

And across companies and markets.

Yes, let's look across companies. Bank of America is a huge company and comprises less than 1% of the U.S. stock market. Apple is a much smaller company that comprises over 4% of the U.S. stock market. Is it reasonable to assume that Apple—with wonderful growth, no serious competition, and viewed widely as a safe haven—should have the same risk premium as Bank of America, a company that has in recent years seemed to lose its way strategically and is facing daunting headwinds in the years ahead? Should they be priced at the same forward-looking rate of return? Probably not.

By the same token, compare the risk premium when people were worried about financial Armageddon in early 2009 and the risk premium when people felt that things were getting solidly back on track in early 2011. Should that risk premium be the same from one year to the next? Of course not.

So, yes, risk premia vary cross-sectionally, across time, across markets, across companies. Is the Greek risk premium higher than the U.S. risk premium today? Yeah, I would think so, which means that investors in Greek stocks should be expecting a higher return than investors in U.S. stocks because of the higher expected uncertainty.

Why are these myths so enduring?

I think the myths are a consequence of inertia. Our industry, both on the practitioner and on the academic sides, has tremendous inertia, a resistance to new ideas. Once people are taught a particular way of thinking, there is a resistance to questioning that way of thinking. One could characterize it even as a bit of intellectual laziness. People embrace an idea that they have been taught, and they hang on to that idea. They are reluctant to relinquish it in favor of something else.

People are taught the normal risk premium is 5%. In early 2001, Ron Ryan and I wrote a paper titled "The Death of the Risk Premium," which was first published as a First Quadrant "President's Letter" and later published in the *Journal of Portfolio Management*, where we suggested that the equity risk premium was now negative. That created a firestorm of controversy and even outrage in some quarters—to suggest that stocks would produce a lower return than bonds. But if stocks have a dividend yield of 1% and bonds have a yield of 6% in an environment of 2% inflation, that points to a negative risk premium, unless stocks can deliver long-term earnings and dividend growth north of 5%. There is nothing written into contract law in the finance community that says, "Stocks must have a positive risk premium."

WE DO OURSELVES A GREAT FAVOR IF WE ABANDON THE NOTION OF A RISK-FREE RATE AND REPLACE IT WITH A NOTION OF A RISK-MINIMIZING ASSET OR PORTFOLIO OVER A HORIZON MATCHING THE INTENDED LIABILITIES.

Why are you so interested in these myths?
My career has largely been successful as a consequence of the fact that I love to test ideas. The more widely accepted an idea is, the more I am inclined to say, "Let's test it and see if it is true."

One of the things that startled me over the course of my career is how few people pursue that line of reasoning—"If an idea is well accepted, maybe we should test it"—and how many people resist those tests when they turn out to suggest that conventional wisdom is wrong. Conventional wisdom isn't always wrong; it's just not always *right*.

How risk free is the risk-free rate?
I think the whole notion of a risk-free rate is a distraction which takes our eye off of the ball in terms of how people think about investments. First, risk free in what context?

The risk of a 30-day Treasury bill defaulting is, for all intents and purposes, zero. The risk of it producing a real return that is less than we expect—that is a much bigger risk because the uncertainty about next month's CPI has a certain standard deviation that makes that so-called risk-free asset a little less risk free than we might think or hope.

Try to persuade any investor with a long-term liability—a typical pension fund, for instance—that owning and rolling T-bills is a risk-free way to fund those pensions. Come on! We don't know what the rates are going to be over the coming years. We don't know what the inflation is going to be, and we don't know what the growth of the liability itself will be. There is *no such thing* as a risk-free rate. The sooner we abandon the notion that there is a risk-free rate, the better off we will be.

If not risk free, then what?
For most long-term investors, the risk-*minimizing* asset—not risk *free*—is something that is duration-matched to your intended spending stream and to your liabilities. If you are a pension fund, for instance, if those liabilities have an inflation kicker to them—if they are sensitive to the rates of inflation—then long TIPS are your risk-minimizing asset.

If we think in terms of risk-minimizing assets over a horizon long enough to matter, we arrive at very, very different answers. All of a sudden, what feels low risk (a cash-dominated portfolio) turns out to be very high risk measured in terms of long-term return expectations and long-term liabilities. Something that feels pretty volatile, a 30-year TIPS instrument, winds up being very low risk measured against long-term liabilities. So, I think we do ourselves a great favor if we abandon the notion of a risk-free rate and replace it with a notion of a risk-minimizing asset or portfolio over a horizon matching the intended liabilities.

Would that alter the traditional asset-pricing models that evaluate risk–return trade-offs?
Peter Bernstein and I published a paper way back in 1988 in the *Harvard Business Review* (they assigned the title "The Right Way to Manage Your Pension Fund," which I thought was a pretty arrogant title). The paper simply said, "If you redefine your efficient frontier to characterize risk as the mismatch between your assets and liabilities, you wind up with a very different efficient frontier and a very different portfolio mix." We urged consultants and pension funds to consider optimizing their holdings on the basis of a redefinition of risk. To this day, I believe that makes absolute sense, and to this day, hardly anyone does it.

How does the LIBOR scandal tie in to this?
I think that the LIBOR scandal is simultaneously a big deal and much ado about nothing, which sounds contradictory.

I say much ado about nothing because when people price swaps off LIBOR, when it is a gamed LIBOR, they figure out what they want to charge for the swap and they price it relative to that gamed LIBOR. The gaming of the LIBOR has nothing to do with the rate that they are charging. The rate that they are charging relative to LIBOR is really an outcome of setting a rate that you want to charge and subtracting the gamed LIBOR from it. So if the gaming of LIBOR is much the same from one period to the next, no one is harmed.

But it was a very big deal in the sense that people trusted that it was a fair interbank borrowing rate. We have had so many damaging body blows to the public's sense of trust in the capital markets. How useful are the capital markets if we can't trust them? How effective is the capitalist system that is predicated on trust? When we do a deal, we trust that the other side will honor their side of the deal.

You attended the CFA Institute forums on the equity risk premium in 2001 and 2011. What did you learn? What was your experience at the forums?

They were fun. As I mentioned, when Ron Ryan and I wrote the paper "The Death of the Equity Risk Premium" in 2000, we ran into a buzz saw of resistance. Today, you don't get that push-back. One thing that has changed is that people, probably by dint of the pain of the last dozen years, are beginning to recognize that the cult of equities is itself promulgating huge myths.

The notion that double-digit returns are natural for stocks, the notion that lower yields are the market's way of telling you to expect faster growth, the notion that stocks are assuredly going to produce higher returns than long bonds for those patient enough to stay the course over the course of one or two economic cycles and that stocks are less risky than bonds for the truly long-term investor—these are all myths that are fast dissipating.

My view that a cult of equities is worshipping a false idol is no longer a fringe view that gets one consigned to our industry's virtual lunatic asylum. It's becoming an acceptable view. So I think we are seeing an opening of minds. The opening of minds is unfortunately a dozen years too late to avert damage, but it is important and interesting to see that it is happening.

You've written on the necessity of challenging deeply rooted assumptions of finance theory. Can you explain?

Neoclassical finance and the capital asset pricing model are predicated on an array of powerful theories and, in many cases, mathematical proofs that demonstrate that if the market behaves in thus and such a fashion, it will have thus and such implications.

Take the capital asset pricing model. If markets are efficient and if investors share a common view on forward-looking risks and returns, if investors trade for free with no taxes and no trading costs, and if all investors have a similar utility function, then the market-clearing portfolio will be the "mean–variance-efficient portfolio" and you can't beat it on a risk-adjusted basis.

That is a very powerful conclusion—deservedly winning a Nobel Prize for Bill Sharpe—built on a foundation of heroic and clearly inaccurate assumptions. I think finance theory is *wonderful*, but I think it is important that we acknowledge that finance theory is theory. It is not the real world. Theory is designed to tell us how the world *ought* to work. The more we can learn from theory and conform theory to better match the real world, the deeper our understanding of markets.

I think, with the coming quarter century, it will be marvelous if we see a marriage—and it will be an uncomfortable marriage—of neoclassical finance with behavioral finance, a theoretical foundation for the empirical observations of behavioral finance. The big issues in finance theory are really simple. If you assume that the theory is correct and true, then we are tacitly assuming that the assumptions are correct and true. And yet nobody would argue that the assumptions are true. I think we need to back off from the notion that theory is reality.

FINANCE THEORY IS THEORY. IT IS NOT THE REAL WORLD.

Are equities worth the risk, given the potentially low equity risk premium?

I think investors are starting to come around to the view that stocks aren't quite as special as they once thought. The sad irony is that the more extravagantly expensive stocks are, the more members you will have in the cult of equities. The reason for that is simple. Stocks become extravagantly expensive by performing brilliantly. After they have performed brilliantly, it is painful to argue the case that stocks are a lousy investment. People come around to the view that stocks aren't guaranteed a premium return *after* equities have underperformed badly for a long period of time. That is unfortunate and it is ironic, but it is a simple fact.

Jonathan Barnes is a financial journalist and author of the novel *Reunion*.

The Level and Persistence of Growth Rates
Author(s): Louis K. C. Chan, Jason Karceski, Josef Lakonishok
Source: *The Journal of Finance*, Vol. 58, No. 2 (Apr., 2003), pp. 643-684
Published by: Blackwell Publishing for the American Finance Association
Stable URL: http://www.jstor.org/stable/3094553
Accessed: 04/02/2010 09:57

THE JOURNAL OF FINANCE • VOL. LVIII, NO. 2 • APRIL 2003

The Level and Persistence of Growth Rates

LOUIS K. C. CHAN, JASON KARCESKI, and JOSEF LAKONISHOK*

ABSTRACT

Expectations about long-term earnings growth are crucial to valuation models and cost of capital estimates. We analyze historical long-term growth rates across a broad cross section of stocks using several indicators of operating performance. We test for persistence and predictability in growth. While some firms have grown at high rates historically, they are relatively rare instances. There is no persistence in long-term earnings growth beyond chance, and there is low predictability even with a wide variety of predictor variables. Specifically, IBES growth forecasts are overly optimistic and add little predictive power. Valuation ratios also have limited ability to predict future growth.

THE EXPECTED RATE of growth in future cash flows (usually proxied by accounting earnings) plays a pivotal role in financial management and investment analysis. In the context of aggregate market valuation, for example, projections about future growth are instrumental in predicting the equity risk premium. Much current controversy surrounds the appropriate level of the equity risk premium, as well as whether recent market valuation levels (at least as of year-end 1999) can be justified (Asness (2000), Welch (2000), Fama and French (2002)). Debate also revolves around how much of the performance of equity asset classes, such as large glamour stocks, can be attributed to changes in profitability growth (Fama and French (1995), Chan, Karceski, and Lakonishok (2000)). When applied to the valuation of individual stocks, projected growth rates have implications for the cross-sectional distribution of cost of capital estimates (Fama and French (1997), Claus and Thomas (2001), Gebhardt, Lee, and Swaminathan (2001)), as well as widely followed valuation ratios like price-to-earnings and price-to-book ratios.

Common measures of expected growth in future earnings, such as valuation ratios and analysts' growth forecasts, vary greatly across stocks. In the case of price-to-earnings multiples for the IBES universe of U.S. firms, for example, at

*Chan is with the Department of Finance, College of Commerce and Business Administration, University of Illinois at Urbana-Champaign; Karceski is with the Department of Finance, Warrington College of Business Administration, University of Florida; and Lakonishok is with the Department of Finance, College of Commerce and Business Administration, University of Illinois at Urbana-Champaign, and NBER. We thank the editor, Rick Green; Cliff Asness; Kent Daniel; Ken French; an anonymous referee; and seminar participants at Dartmouth, Duke University, the London School of Economics Financial Markets Group, the NBER Behavioral Finance Fall 2000 workshop, the University of Illinois, Washington University, and the Western Finance Association 2001 meetings.

year-end 1999, the distribution of the stock price relative to the consensus forecast of the following year's earnings has a 90th percentile of 53.9, while the 10th percentile is 7.4, yielding a difference of 46.5. Firms with a record of sustained, strong past growth in earnings are heavily represented among those trading at high multiples. Security analysts issue positive recommendations for these stocks and forecast buoyant future prospects. Other stocks with a history of disappointing past growth are shunned by the investment community. They are priced at low multiples and analysts are unexcited about their outlook. Putting aside the possibility of mispricing, one reason for the disparity in multiples is differences in risk. At the level of individual stocks, however, the relation between risk and expected return is weak (Fama and French (1992)). It is thus unlikely that the large dispersion is driven primarily by risk (the evidence in Beaver and Morse (1978) also supports this view). Rather, if the pricing is rational, most of the cross-sectional variation reflects differences in expected growth rates. A more direct measure of the market's expectations, security analysts' forecasts of long-term growth in earnings, also displays large differences across stocks. For example, the 90th percentile of the distribution of IBES five-year forecasts is 40 percent as of year-end 1999, compared to the 10th percentile of 8.9 percent. If analysts and investors do not believe that future earnings growth is forecastable, they would predict the same growth rate (the unconditional mean of the distribution) for all companies, and it is unlikely that the dispersion in forecasts or price–earnings ratios would be as large as it actually is.

Based on market valuations and analysts' forecasts, then, there is a widespread belief among market participants that future earnings growth is highly predictable. However, economic intuition suggests that there should not be much consistency in a firm's profitability growth. Following superior growth in profits, competitive pressures should ultimately tend to dilute future growth. Exit from an unprofitable line of business should tend to raise the remaining firms' future growth rates. Some support for this logic comes from Fama and French (2002). Their evidence for the aggregate market suggests that while there is some short-term forecastability, earnings growth is in general unpredictable.

In short, there may be a sharp discrepancy between share valuations along with analysts' predictions on the one hand, and realized operating performance growth on the other. The discrepancy may reflect investors' judgmental biases or agency distortions in analysts' behavior. In any event, the divergence is potentially large, judging from current market conditions. For instance, take a firm with a ratio of price to forecasted earnings of 100. Such cases are by no means minor irregularities: based on values at year-end 1999, they represent about 11.9 percent of total market capitalization. To infer the growth expectations implicit in such a price earnings ratio, we adopt a number of conservative assumptions. In particular, suppose the multiple reverts to a more representative value of 20 in 10 years, during which time investors are content to accept a rate of return on the stock of zero (assume there are no dividends). A multiple of 20 is conservative, since Siegel (1999) argues that a ratio of 14 may not be an unreasonable long-term value. Further, an adjustment period of 10 years is not short, in light of the fact that many of the largest firms at year-end 1999 did not exist 10 years ago. These assumptions imply that earnings must grow by a factor of five, or at a rate of about 17.5 percent per year, for the next 10 years. Alternatively, suppose investors put up with a paltry 10 percent rate of return (Welch (2000), reports that financial economists' consensus expected return is considerably higher). Then earnings must grow at an even more stellar rate (29.2 percent per year) over 10 years to justify the current multiple.

The above example highlights the two questions we tackle in this paper. How plausible are investors' and analysts' expectations that many stocks will be able to sustain high growth rates over prolonged periods? Are firms that can consistently achieve such high growth rates identifiable ex ante? We begin by documenting the distribution of growth rates realized over horizons of 1, 5, and 10 years. This evidence lets us evaluate the likelihood of living up to the expectations of growth that are implicit in market valuation ratios. To justify rich valuations, investors must believe that high growth persists over many years. Accordingly, we also examine whether there is persistence in operating performance growth. Individual firms' earnings and incomes can be very erratic, so a robust empirical design is a crucial consideration. We employ nonparametric tests on multiple indicators of operating performance across a large cross section of stocks over relatively long horizons. In addition, we focus our tests for persistence by examining subsets of firms where future growth is more likely to be predictable (e.g., stocks in the technology sector and stocks which have displayed persistence in past growth). To give the benefit of the doubt to the possibility of persistence, we relax the definition of consistency in growth and redo our tests. Finally, we expand the list of variables to forecast growth beyond past growth rates. We examine whether valuation measures, such as earnings yields and ratios of book-to-market equity and sales-to-price, are associated with growth on an ex ante as well as ex post basis. Security analysts' earnings forecasts are also widely used as measures of the market's expectations of growth in future earnings. As a check on the quality of analysts' predictions, we evaluate how well realized growth rates align with IBES consensus forecasts.

Our main findings are as follows. Our median estimate of the growth rate of operating performance corresponds closely to the growth rate of gross domestic product over the sample period. Although there are instances where firms achieve spectacular growth, they are fairly rare. For instance, only about 10 percent of firms grow at a rate in excess of 18 percent per year over 10 years. Sales growth shows some persistence, but there is essentially no persistence or predictability in growth of earnings across all firms. Even in cases that are popularly associated with phenomenal growth (pharmaceutical and technology stocks, growth stocks, and firms that have experienced persistently high past growth), signs of persistent growth in earnings are slim. Security analysts' long-term growth estimates tend to be overoptimistic and contribute very little to predicting realized growth over longer horizons. Market valuation ratios have little ability to discriminate between firms with high or low future earnings growth. An expanded set of forecasting variables also has scant success in predicting future earnings growth. All in all, our evidence on the limited predictability of earnings growth suggests that investors should be wary of stocks that trade at very high multiples. Very few firms are able to live up to the high hopes for consistent growth that are built into such rich valuations.

Related prior research in the financial literature on the behavior of earnings growth is meager. Little (1962) and Little and Rayner (1966) examine the growth in earnings of a limited sample of U.K. firms in the 1950s. Early evidence for U.S. firms is provided by Lintner and Glauber (1967) and Brealey (1983). Beaver (1970) and Ball and Watts (1972) start a long line of papers that apply time-series models to earnings. However, few firms have sufficiently long earnings histories to allow precise estimation of model parameters, and the emphasis in this line of work has been on short-term forecasting. More recently, Fama and French (2002) examine the time-series predictability of aggregate earnings for the market. Our work is closest in spirit to that of Fama and French (2000), who look at the cross-sectional predictability of firms' earnings, but even they focus on one-year horizons.

A much larger number of studies by academics and practitioners rely on estimates of expected long-term earnings growth for stock valuation, or for estimating firms' cost of capital. A selective list includes Bakshi and Chen (1998), Lee, Myers, and Swaminathan (1999), Claus and Thomas (2001), and Gebhardt et al. (2001). In particular, many studies use long-term consensus IBES forecasts for expected growth rates (see, e.g., Mezrich et al. (2001)). Given the widespread use of IBES long-term estimates, it is important to evaluate their correspondence with realized growth rates.

The rest of the paper is organized as follows. Section I discusses our sample and some basics of the methodology. The cross-sectional distribution of firms' growth rates is reported in Section II. Section III presents the results of runs tests for consistency in growth of operating performance. Section IV takes up the issue of survivorship bias. Although our main focus is not on the determinants of valuation multiples, Section V examines the relation between growth and valuation ratios such as earnings yields and book-to-market ratios, on both an ex ante and ex post basis. We compare IBES long-term forecasts with realized growth rates in Section VI. Section VII uses cross-sectional regressions to forecast future growth using variables including past growth, valuation ratios, and IBES estimates. A final section concludes.

I. Sample and Methodology

Our sample of firms comprises all domestic common stocks with data on the Compustat Active and Research files. Firms are selected at the end of each calendar year from 1951 to 1997. The earlier years are included for the sake of completeness, even though there is a backfill bias in the earlier part of the sample period (see Chan, Jegadeesh, and Lakonishok (1995)), which may impart an upward bias to growth rates in the beginning of the sample. The number of eligible firms grows from 359 in the first sample selection year to about 6,825 in the last year; on average, the sample comprises about 2,900 firms.

We consider three indicators of operating performance: net sales (Compustat annual item number 12), operating income before depreciation (item 13), and income before extraordinary items available for common equity (item 237). While researchers and practitioners tend to focus exclusively on income before extraordinary items, measuring growth in this variable is beset with pitfalls. In many cases, earnings before extraordinary items is negative, so prospective growth rates are undefined (for our sample, in an average year, 29 percent of firms have negative values for earnings before extraordinary items). In other cases, firms grow from low positive values of base-year net income, introducing large outliers.[1] These include such disparate cases as beaten-down companies with depressed earnings and growing startup companies that are beginning to generate profits. To avoid hanging all our inferences on such a noisy variable, therefore, we also consider growth in net sales and growth in operating income before depreciation. These are relatively better-behaved measures of operating performance.

Researchers have adopted different conventions for calculating growth rates. Given our focus on the predictability of growth rates, we measure growth on a per share basis so as to strip out any predictability due to changes in the scale of the firm's operations. This also corresponds to the measurement convention in the investment industry.[2]

Thus, we take the perspective of an investor who buys and holds one share of a stock over some horizon and track the growth in sales or income that accrues to one share, after adjusting for stock splits and dividends. Moreover, two firms can offer the same expected return, but have different earnings growth rates because of their dividend payout policies. From an investor's standpoint, these two stocks would be considered equivalent. To put firms with different dividend policies on an equal footing, therefore, all cash dividends as well as any special distributions (such as when a firm spins off assets) are reinvested in the stock.

II. The Distribution of Growth Rates of Operating Performance

This section documents the distribution of historical growth rates over relatively long horizons (5 and 10 years). For the sake of completeness, results are also provided for 1-year horizons. At each calendar year-end over the sample period, we measure rates of growth in future operating performance for all eligible

[1] Some of these complications may be alleviated by averaging earnings over a number of years and measuring growth in these averages. Since our focus is on point-in-time growth rates, we do not explore this alternative procedure. In unreported work, we also experiment with other ways to calculate growth rates. These include value-weighted growth rates for portfolios, estimated growth rates from least-squares fits of linear and quadratic time trends through sales and income, and growth rates without dividend reinvestment. Generally speaking, the results are robust to how we measure growth rates.

[2] Lakonishok, Shleifer, and Vishny (1994) calculate growth in a firm's overall sales and earnings, while Daniel and Titman (2001) calculate growth on a per share basis. These studies focus on the impact of investor sentiment on stock returns. The hypothesis is that investors tend to favor companies with strong past performance, those in a glamorous line of business, or those which are perceived to be well managed. From this standpoint, it might be argued that it is the performance of the overall company that is relevant, and not just the profits earned per share.

stocks. Percentiles are calculated for the distribution obtained at each year-end. Table I reports the percentiles averaged across years in the sample period, as well as the most recent distribution corresponding to the last selection year of the sample period.

Several points are important as background to the results in Table I. First, since we include reinvestment of dividends and special distributions, the growth rates we report are typically higher than conventionally measured growth rates. The median dividend yield for our sample (averaged across all years) is about 2.5 percent. A second caveat is that the tabulated growth rates are based only on firms who survive for the following 1, 5, or 10 years. The survivorship bias may induce an upward bias in our reported growth rates. Moreover, we follow the conventional approach and do not calculate growth rates for operating income before depreciation or income before extraordinary items when the base-year value is negative.[3] To illustrate the potential magnitude of these complications, on average there are about 2,900 firms available for inclusion in the sample at each year-end. Of these, 2,782 firms survive at the end of the next year and have a reported value for income before extraordinary items. The calculations for 1-year growth in earnings before extraordinary items are based on 1994 of these firms; the remaining 788 firms have negative values for income in the base year. At the 5-year horizon, there are on average 1884 surviving firms. Growth rates are calculated for 1,398 of these; 486 have negative base-year values. At the 10-year horizon, there are 1,265 surviving firms: 1,002 and 263 with positive and negative base-year values, respectively. In a subsequent section, we examine the performance of nonsurviving firms.

Since negative base-year values are quite common for income before extraordinary items, valid growth rates are unavailable in many cases. These observations are symptomatic of another problem. In particular, the high frequency of cases with negative base values suggests that the neighboring portion of the distribution (with low, positive base-year values) contains a large fraction of the observations as well. These instances give rise to some very high growth rates. For growth over five years, for example, the 98th percentile value for growth in income before extraordinary items averages 62.4 percent per year. Hence, while growth in income before extraordinary items captures much of the investment community's interest, its behavior is the most questionable. While the same problem applies to operating income before depreciation, the frequency of negative base-year values is comparatively lower and growth in this variable is less problematic.[4] For growth in this variable, the 98th percentile is 51.2 percent on average. In comparison, sales growth is relatively well behaved, with a 98th percentile value of 40.5 percent on average. These comparisons suggest that looking at

[3] Note, however, that even if we are unable to calculate growth in income before extraordinary items in such a case, we still get a reading on a firm's operating performance growth from sales (or operating income before depreciation if it is positive).

[4] For example, of the firms surviving after one year and with a reported value for income before depreciation, about 14 percent on average have negative base-year values. The corresponding percentage for income before extraordinary items is 29 percent.

other indicators beyond income before extraordinary items helps to give a more robust picture of growth in operating performance.

The results in Table I serve as cautionary flags to analysts and investors who pursue stocks with rich price–earnings multiples. Take our original example of a stock with a current price–earnings multiple of 100, which declines to 20 in 10 years' time with an expected return of 10 percent per year. Earnings must grow at 29.2 percent per year over 10 years to justify the current multiple. This is a tall order by historical standards. In particular, the required growth rate corresponds to about the 95th percentile of the distribution of 10-year growth rates, even putting aside the inclusion of dividends. Put differently, suppose earnings grow at a historically more representative, but still healthy, annual rate of 14.7 percent (the 75th percentile of the distribution from Part I). Then the current ratio of 100 would be justified if the time it takes for the multiple to fall to 20 is stretched out to 38 years.

Small firms start from a smaller scale of operations and so have more room for potential growth, possibly justifying a high current multiple. However, high multiples also apply to many large, well-known firms. To see whether large firms in general can also achieve high growth, Table II reports the distribution of growth rates for large firms (companies ranked in the top two deciles of year-end equity market capitalization, based on NYSE breakpoints). Bigger firms have a larger scale of operations and, hence, are more likely to face limits on their growth, so extremely high growth rates are less prevalent in Table II compared to Table I. For example, the 90th percentiles of growth rates over 10 years for income before extraordinary items, operating income before depreciation, and sales are all close to 16 percent per year. Also, note that dividend yields are generally higher for large firms.

Our estimated median growth rate is reasonable when compared to the overall economy's growth rate. On average over the sample period, the median growth rate over 10 years for income before extraordinary items is about 10 percent for all firms. The behavior over the last 10-year period in the sample roughly matches the overall average. Growth in the other two indicators also exhibit comparable medians. After deducting the dividend yield (the median yield is 2.5 percent), as well as inflation (which averages 4 percent per year over the sample period), the growth in real income before extraordinary items is roughly 3.5 percent per year. This is consistent with the historical growth rate in real gross domestic product, which has averaged about 3.4 percent per year over the period 1950 to 1998. It is difficult to see how the profitability of the business sector over the long term can grow much faster than overall gross domestic product.

Looking forward, if we project future growth using the median of the distribution of historical growth rates, the implication is that the expected future return on stocks is not very high. For example, in a simple dividend discount model with constant growth rates and constant payout ratio, the expected return is equal to the dividend yield plus the expected future growth rate of earnings. Given the low level of current dividend yields (below 1.5 percent) and expected inflation of 2.5 percent, the expected return is only about 7.5 pecent. This is lower than the

Table I

Distribution of Growth Rates of Operating Performance over 1, 5 and 10 Years: All Firms

At every calendar year-end over the sample period, growth rates in operating performance are calculated over each of the following one, five, and ten years for all firms in the sample. The sample period is 1951 to 1998, and the sample includes all domestic firms listed on the New York, American, and Nasdaq markets with data on the Compustat files. Operating performance is measured as sales, operating income before depreciation, or income before extraordinary items available to common equity. Growth in each variable is measured on a per share basis as of the sample selection date, with the number of shares outstanding adjusted to reflect stock splits and dividends; cash dividends and special distributions are also reinvested. Percentiles of the distribution are calculated each year-end; the simple average over the entire sample period of the percentiles is reported, along with the distribution of growth rates over horizons ending in the last year of the sample period.

	Percentile								
Sample period	2%	10%	25%	40%	50%	60%	75%	90%	98%
Part I: Annualized Growth Rate over 10 Years									
(A) Sales									
Average	−9.6	0.1	5.5	8.7	10.2	11.5	13.8	18.0	27.6
Ending 1998	−16.1	−3.4	2.9	6.2	7.9	9.5	12.7	19.2	32.9
(B) Operating Income before Depreciation									
Average	−13.3	−2.3	4.1	7.6	9.5	11.2	14.1	19.4	31.3
Ending 1998	−14.6	−3.3	3.3	7.2	9.0	10.9	14.1	21.5	38.6
(C) Income before Extraordinary Items									
Average	−15.6	−3.1	3.9	7.7	9.7	11.6	14.7	20.4	33.4
Ending 1998	−21.2	−6.3	2.3	6.9	9.0	11.4	15.3	24.4	48.8
Part II: Annualized Growth Rate over 5 Years									
(A) Sales									
Average	−18.7	−4.1	4.3	8.2	10.2	12.0	15.3	22.1	40.5
Ending 1998	−22.7	−6.2	2.9	8.0	10.2	12.4	17.1	27.6	56.3
(B) Operating Income before Depreciation									
Average	−26.8	−8.4	1.9	7.2	9.8	12.4	17.1	26.7	51.2
Ending 1998	−24.4	−7.8	3.5	8.7	11.5	14.4	19.9	33.4	64.4
(C) Income before Extraordinary Items									
Average	−30.9	−10.3	1.5	7.4	10.5	13.4	18.8	30.4	62.4
Ending 1998	−35.1	−11.5	2.8	9.1	12.4	15.7	23.1	40.1	88.2

Part III: 1-Year Growth Rate									
(A) Sales									
Average	−47.3	−12.9	1.2	7.6	10.9	14.2	21.0	38.7	121.7
Ending 1998	−58.3	−20.8	−1.4	6.3	10.3	14.5	24.9	54.1	181.9
(B) Operating Income before Depreciation									
Average	−69.4	−30.7	−5.6	5.9	11.8	17.7	30.6	67.4	253.3
Ending 1998	−74.1	−34.7	−4.9	6.7	12.2	18.5	32.2	76.5	273.2
(C) Income before Extraordinary Items									
Average	−76.8	−37.9	−7.4	6.9	13.3	19.9	35.8	90.2	435.3
Ending 1998	−87.3	−48.2	−13.7	5.4	13.7	21.3	40.4	115.0	727.2

Table II

Distribution of Growth Rates of Operating Performance over 1, 5 and 10 Years: Large Firms

At every calendar year-end over the sample period, growth rates in operating performance are calculated over each of the following one, five, and ten years for large firms (in the top two deciles of year-end equity market capitalization, based on NYSE breakpoints). The sample period is 1951 to 1998, and the sample includes all domestic firms listed on the New York, American, and Nasdaq markets with data on the Compustat files. Operating performance is measured as sales, operating income before depreciation, or income before extraordinary items available to common equity. Growth in each variable is measured on a per share basis as of the sample formation date, with the number of shares outstanding adjusted to reflect stock splits and dividends; cash dividends and special distributions are also reinvested. Percentiles of the distribution are calculated each year-end; the simple average over the entire sample period of the percentiles is reported, along with the distribution of growth rates over horizons ending in the last year of the sample period.

	Percentile								
Sample period	2%	10%	25%	40%	50%	60%	75%	90%	98%
Part I: Annualized Growth Rate over 10 Years									
(A) Sales									
Average	− 3.4	2.5	6.8	9.4	10.7	11.7	13.3	16.3	22.0
Ending 1998	− 7.7	− 0.2	4.4	6.7	8.5	9.5	11.1	15.0	21.5
(B) Operating Income before Depreciation									
Average	− 8.3	0.6	5.4	8.1	9.5	10.8	12.9	16.1	22.6
Ending 1998	− 11.6	− 1.7	4.3	7.4	8.7	10.4	11.8	16.3	21.4
(C) Income before Extraordinary Items									
Average	− 12.8	− 0.9	4.5	7.5	9.3	10.8	13.1	16.6	23.8
Ending 1998	− 25.6	− 3.8	1.7	6.1	8.2	9.9	13.3	18.5	36.4
Part II: Annualized Growth Rate over 5 Years									
(A) Sales									
Average	− 9.7	− 0.6	6.9	9.4	10.8	11.9	14.1	18.1	27.9
Ending 1998	− 13.6	− 3.0	4.0	8.8	10.2	11.5	13.7	19.6	32.5
(B) Operating Income before Depreciation									
Average	− 16.9	− 3.5	4.3	7.9	9.8	11.5	14.3	19.3	32.1
Ending 1998	− 13.6	− 6.6	4.5	7.5	10.8	12.7	15.6	19.9	32.0
(C) Income before Extraordinary Items									
Average	− 26.4	− 6.4	2.8	7.6	9.8	12.0	15.3	21.3	37.2
Ending 1998	− 39.5	− 10.1	4.3	9.5	11.8	14.4	19.6	30.4	57.4
Part III: 1-Year Growth Rate									
(A) Sales									
Average	− 36.4	− 2.4	5.7	9.3	11.3	13.3	17.0	25.2	47.7
Ending 1998	− 49.8	− 14.7	1.5	6.6	8.9	11.8	18.1	29.1	53.0
(B) Operating Income before Depreciation									
Average	− 52.3	− 15.2	0.2	7.1	10.6	13.8	19.8	33.7	82.3
Ending 1998	− 60.0	− 30.3	− 1.9	6.6	11.1	14.0	20.8	33.4	73.1
(C) Income before Extraordinary Items									
Average	− 67.5	− 25.3	− 2.8	6.9	11.0	14.9	23.1	45.9	216.6
Ending 1998	− 80.0	− 46.9	− 13.5	4.7	11.5	15.5	27.1	56.7	213.6

consensus forecast of professional economists (see Welch (2000)), but is in line with Fama and French (2002).

III. Persistence in Growth

Differences in valuations indicate a pervasive belief that stocks with high or low future growth are easily identifiable ex ante. For example, analysts and investors seem to believe that a firm that has grown rapidly in the past for several years in a row is highly likely to repeat this performance in the future. Conversely, stocks that have done poorly over prolonged periods are shunned and trade at low multiples. This section checks whether there is consistency in growth. We examine whether past growth or other characteristics, such as industry affiliation or firm size, help to predict future growth.

A. Consistency across All Firms

Tables I and II suggest that year-to-year growth in income can take on quite extreme values. As a result, multiyear growth rate levels may look impressive because of one or two isolated years of sharp growth, although growth in other years may be unremarkable. However, many of the firms with lofty multiples grow rapidly every year for several years. Accordingly, we test for consistency in growth using a design that does not rely heavily on the level of growth rates.[5] In our first set of tests, we define consistency as achieving a growth rate above the median for a consecutive number of years: Such cases are labeled as runs.[6]

At each year-end over the sample period, we calculate how many firms achieve runs over horizons of 1 to 10 years in the future. A run over 5 years, for example, denotes a case where in each of the subsequent 5 years, a firm's growth rate exceeds the median growth rate that year. Each year's median is calculated over all growth rate observations available in that year. Again, note that survivorship bias affects our runs tests. To see how many firms achieve runs above the median for 5 years in a row, we necessarily look at firms that survive over the full 5 years. In each of these years, we compare the survivors to a median which is based on all available firms that year, including those that do not survive for the full 5 years,

[5] Brealey (1983) uses a similar procedure.

[6] We want to avoid discarding an entire sequence of observations because one year's growth rate cannot be calculated when earnings are negative. Instead, we handle such cases as follows, taking growth in operating income per share OI_t as an example. In addition to calculating the percentage growth rate of operating income as $(OI_{t+1} - OI_t)/OI_t$ for each firm, we also scale the change in operating income by the stock price as of the base year t, $(OI_{t+1} - OI_t)/P_t$. All firms in a given year are ranked by their values of change in income relative to stock price. For any firm with negative income in a base year, we find its percentile rank based on income change relative to price. We then look up the corresponding percentile value from the distribution of growth rates of income (based on firms with positive base-year values) for that year. This growth rate is then assigned to the firm with negative base-year income. At the same time, however, it would be dangerous to pin our estimates of growth over a 5- or 10-year horizon in Tables I and II on some imputed value of base-year earnings. Accordingly, we do not impute growth rates in those tables for cases with negative base-year values.

Table III

Persistence in Growth Rates of Operating Performance: All Firms

At every calendar year-end over the sample period, growth rates in operating performance are calculated over each of the following one to ten years (or until delisting) for all firms in the sample. The sample period is 1951 to 1998, and the sample includes all domestic firms listed on the New York, American, and Nasdaq markets with data on the Compustat files. Operating performance is measured as sales (panel A), operating income before depreciation (panel B), or income before extraordinary items available to common equity (panel C). Growth in each variable is measured on a per share basis as of the sample formation date, with the number of shares outstanding adjusted to reflect stock splits and dividends; cash dividends and special distributions are also reinvested. For each of the following ten years, the number of firms with valid growth rates, the number of firms whose growth rate exceeds the median growth rate each year for the indicated number of years, the percentage these firms represent relative to the number of valid firms, and the percentage expected under the hypothesis of independence across years, are reported. Statistics are provided for the entire sample period, and for the ten-year horizon corresponding to the last sample formation year.

	Firms with Above-Median Growth each year for Number of Years									
Variable	1	2	3	4	5	6	7	8	9	10
					(A) Sales					
Average Number of Valid Firms	2771	2500	2263	2058	1878	1722	1590	1471	1364	1265
Average Number above Median	1386	721	382	209	118	70	42	26	17	11
Percent above Median	50.0	28.8	16.9	10.2	6.3	4.0	2.7	1.8	1.3	0.9
1989–1998	50.0	30.0	18.6	11.9	7.8	5.6	3.4	2.4	1.5	1.2
					(B) Operating Income before Depreciation					
Average Number of Valid Firms	2730	2456	2219	2014	1833	1678	1546	1428	1322	1223
Average Number above Median	1365	628	290	136	67	34	18	10	6	4
Percent above Median	50.0	25.6	13.0	6.8	3.6	2.0	1.2	0.7	0.5	0.3
1989–1998	50.0	25.0	13.1	7.0	4.0	2.1	1.3	0.8	0.5	0.5
					(C) Income before Extraordinary Items					
Average Number of Valid Firms	2782	2509	2271	2065	1884	1727	1593	1473	1365	1265
Average Number above Median	1391	625	277	125	57	28	14	7	4	2
Percent above Median	50.0	24.9	12.2	6.0	3.0	1.6	0.9	0.5	0.3	0.2
1989–1998	50.0	24.8	12.2	5.7	2.6	1.3	0.8	0.5	0.2	0.0
Expected Percent above Median	50.0	25.0	12.5	6.3	3.1	1.6	0.8	0.4	0.2	0.1

and newly listed firms. Since the survivors are likely to have better performance than the population, they tend to have a greater chance of being above the median. Section IV examines differences between the growth rates of surviving and nonsurviving firms.

Table III reports the counts of runs, averaged across the year-ends. For growth in sales (Panel A), for example, out of an average number of 2,900 firms available for sample selection at each year-end, 2,771 firms on average survive until the end of the following year. Over the following 10 years, there are on average 1,265 surviving firms. Of these, 11 have sales growth rates that exceed the median in each of the 10 years, representing 0.9 percent of the eligible firms. If sales growth is independent over time, we should expect to see 0.5^{10} (about 0.1 percent) of the surviving firms achieve runs above the median over 10 years (see the last row of the table). To give a flavor of what happens in the more recent years, we also report the percentage of firms with runs over the 10-year period ending in the last year of our sample period.

There is a great deal of persistence in sales growth. Over a five-year horizon, for example, on average 118 firms, or 6.3 percent of the 1878 firms who exist over the full five years, turn in runs above the median. The number expected under the hypothesis of independence over time is about 59 (3.1 percent of 1,878), so roughly twice more than expected achieve runs over five years.

The persistence in sales growth may reflect shifts in customer demand, which are likely to be fairly long-lasting. A firm can also sustain momentum in sales by expanding into new markets and opening new stores, by rolling out new or improved products, or by granting increasingly favorable credit terms. Persistence in sales may also arise from managers' "empire-building" efforts, such as expanding market share regardless of profitability. In all these cases, however, profit margins are likely to be shrinking as well, so growth in profits may not show as much persistence as sales growth.

While it may be relatively easy for a firm to generate growth in sales (by selling at a steep discount, for example), it is more difficult to generate growth in profits. The recent experience of Internet companies, where sales grew at the same time losses were accumulating, provides a stark example. Panel B confirms that there is less persistence in operating income before depreciation compared to sales. On average, 67 firms a year, or 3.6 percent of 1,833 surviving firms, have above-median runs for 5 consecutive years. The expected frequency of runs is 3.1 percent or 57 firms. There are, thus, 10 firms more than expected out of 1,833, so the difference is unremarkable. An average of 4 firms a year (or 0.3 percent of 1,223 survivors), which is only 3 more than expected, pull off above-median growth for 10 years in a row. The patterns in the more recent years do not deviate markedly from the averages across the entire sample period.

Any sign of persistence vanishes as we get closer to the bottom line (Panel C). On average, the number of firms who grow faster than the median for several years in a row is not different from what is expected by chance. An average of 57 firms out of 1,884 survivors (3 percent) beat the median for 5 years in a row, while 59 (3.1 percent) are expected to do so. Runs above the median for 10 years occur in 0.2 percent of 1,265 cases (or 2 firms), roughly matching the expected frequency (0.1 percent, or 1 firm). To sum up, analysts and investors seem to believe that many firms' earnings can consistently grow at high rates for quite a few years. The evidence suggests instead that the number of such occurrences is not much different from what might be expected from sheer luck. The lack of consistency in earnings growth agrees with the notion that in competitive markets, abnormal profits tend to be dissipated over time.

Table IV

Persistence in Growth Rates of Operating Performance: Selected Equity Classes

At every calendar year-end over the sample period, growth rates in operating performance are calculated over each of the following one to ten years (or until delisting) for all firms in the sample. The sample period is 1951 to 1998, and the underlying sample includes all domestic firms listed on the New York, American, and Nasdaq markets with data on the Compustat files. Operating performance is measured as sales, operating income before depreciation, or income before extraordinary items available to common equity. Growth in each variable is measured on a per share basis as of (the sample formation date, with the number of shares outstanding adjusted to reflect stock splits and dividends; cash dividends and special distributions are also reinvested. For each of the following ten years, the number of firms whose growth rate exceeds the median growth rate each year for the indicated number of years is expressed as a percentage of the number of firms with valid growth rates. Statistics are provided for the following sets of stocks: technology stocks (panel A), comprising stocks whose SIC codes begin with 283, 357, 366, 38, 48, or 737; value stocks (panel B), comprising stocks ranked in the top three deciles by book-to-market value of equity; glamour stocks (panel C), comprising an equivalent number as in panel B of the lowest-ranked stocks by book-to-market value of equity; large stocks (panel D), comprising stocks ranked in the top 2 deciles by equity market value; mid-cap stocks (panel E), comprising stocks ranked in the third through seventh deciles by equity market value; and small stocks (panel F), comprising stocks ranked in the bottom three deciles by equity market value. All decile breakpoints are based on domestic NYSE stocks only.

	Percent of Firms with Above-Median Growth each Year for Number of Years									
Variable	1	2	3	4	5	6	7	8	9	10
				(A) Technology Stocks						
Sales	51.6	30.7	19.1	12.5	8.5	5.9	4.2	3.0	2.3	1.7
Operating Income	51.0	27.2	14.9	8.7	5.3	3.3	2.2	1.4	1.0	0.7
Income before Extraordinary Items	50.9	25.9	13.5	7.3	4.1	2.5	1.5	0.9	0.5	0.4
				(B) Value Stocks						
Sales	50.6	30.0	18.2	11.1	6.9	4.3	2.8	1.9	1.3	0.9
Operating Income	49.3	25.3	13.2	6.8	3.5	1.8	0.9	0.5	0.3	0.2
Income before Extraordinary Items	48.3	23.8	11.4	5.4	2.5	1.2	0.7	0.4	0.3	0.2
				(C) Glamour Stocks						
Sales	48.3	26.6	15.1	8.5	4.7	2.7	1.7	1.0	0.8	0.6
Operating Income	50.1	25.2	11.9	5.9	3.3	1.7	1.0	0.6	0.4	0.3
Income before Extraordinary Items	50.7	25.2	12.0	5.8	2.9	1.6	0.9	0.4	0.2	0.1
				(D) Large Stocks						
Sales	53.2	31.3	18.9	11.7	7.5	4.8	3.2	2.2	1.6	1.1
Operating Income	49.4	25.2	13.0	6.9	3.7	2.0	1.1	0.6	0.4	0.3
Income before Extraordinary Items	46.7	21.9	10.0	4.7	2.2	1.2	0.7	0.4	0.3	0.2
				(E) Mid-cap Stocks						
Sales	53.9	32.4	19.8	12.1	7.6	4.9	3.3	2.2	1.5	1.0
Operating Income	50.5	26.6	13.9	7.5	4.2	2.4	1.5	1.0	0.7	0.4
Income before Extraordinary Items	49.4	24.9	12.4	6.2	3.1	1.6	0.9	0.5	0.3	0.2
				(E) Small Stocks						
Sales	47.0	26.1	14.7	8.6	5.2	3.2	2.1	1.4	1.0	0.7
Operating Income	50.1	25.2	12.6	6.4	3.3	1.8	1.0	0.6	0.4	0.2
Income before Extraordinary Items	51.0	25.5	12.6	6.3	3.2	1.7	0.9	0.4	0.2	0.1
Expected Percent above Median	50.0	25.0	12.5	6.3	3.1	1.6	0.8	0.4	0.2	0.1

B. Consistency for Subsets of Firms

While Table III suggests that there may not be much consistency in growth across all firms, it is possible that consistency may show up more strongly in subsets of firms. Table IV focuses our tests by looking at the performance of subsamples of firms. For a subsample such as small stocks, we consider a "run" as a case where the firm's growth rate exceeds the median for a consecutive number of years, where each year the median is calculated across all firms in the entire sample, not just small stocks. This explains why the percentage of runs is not identically 50 percent in the first year.

Many observers single out technology and pharmaceutical firms as instances of consistently high growth over long horizons. Such firms may be able to maintain high growth rates because of their intangible assets, such as specialized technological innovations or drug patents. Panel A examines firms in these sectors. Specifically, the sample comprises firms that are relatively heavily engaged in research and development activity, and are predominantly drawn from the computer equipment, software, electrical equipment, communications, and pharmaceutical industries.[7] Growth in sales and operating income for the set of technology firms both display strong persistence. However, the percentage of runs in income before extraordinary items does not differ markedly from the expected frequency. For example, over a five-year horizon, 14 firms (or 4.1 percent of the 331 surviving technology stocks) have above-median runs. This is only 4 more than the expected number of runs (10 firms, or 3.1 percent). The recent experience of Internet companies provides numerous examples where sales grow rapidly for several years, at the same time that losses are mounting.

Panel A may exaggerate the degree of persistence in growth for technology stocks on two accounts. First, the technology stocks are evaluated against the median growth rate of the entire sample of firms, which would include, for example, utility stocks with relatively unexciting growth rates. Second, technology stocks are relatively more volatile, so survivorship bias may be a particularly acute problem in this subsample.

Technology stocks that are intensive in research and development also tend to be glamour stocks with low ratios of book-to-market value of equity. The popular sentiment regarding persistence in growth applies to glamour stocks generally. These stocks typically enjoy higher past growth in operating performance than value stocks with high book-to-market ratios (see Lakonishok et al. (1994)). The evidence from psychology suggests that individuals tend to use simple heuristics in decision making. As LaPorta et al. (1997) argue, investors may think that there is more consistency in growth than actually exists, so they extrapolate glamour stocks' past good fortunes (and value stocks' past disappointments) too far into the future. Panels B and C of Table IV test for consistency in growth for value and glamour stocks, respectively. Value stocks comprise stocks that are ranked

[7] Specifically, the sample includes all firms whose SIC codes begin with 283, 357, 366, 38, 48, or 737. See Chan, Lakonishok, and Sougiannis (2001).

in the top three deciles by book-to-market ratio based on NYSE breakpoints, while glamour stocks represent an equivalent number of stocks with the lowest positive book-to-market ratios. Growth in sales is persistent for both sets of stocks. The results for the other measures of operating performance, however, are not markedly different across the two sets of stocks.

The remaining panels perform our runs tests for large, midcapitalization, and small stocks. Large stocks include stocks in the top two deciles of market capitalization based on NYSE breakpoints as of June in the sample selection year, midcapitalization stocks fall in the next five deciles, and small stocks include the bottom three deciles. While sales growth tends to be more persistent for large firms, it does not translate into persistent growth in income. Of the large stocks, 2.2 percent achieve five-year runs in growth of income before extraordinary items, while 3.2 percent of small stocks achieve the same result (the expected fraction is 3.1 percent).

C. Runs Tests Conditional on Past Growth

It might be expected that firms that have demonstrated consistently superior past growth would be able to maintain their growth in the future. In the case of firms such as Microsoft and EMC, their valuations at year-end 1999 reflected investors' bets that these firms will beat the odds and continue the streak. Table V checks whether firms that have demonstrated consistently high (or low) past growth have continued success in the future.

Part I of Table V applies runs tests to those firms that have achieved superior past growth. In Panel A, at every year-end, we select those firms with above-median growth in each of the prior five years (or three years), and examine their subsequent growth.

Superior past growth in sales carries over into the future. In Panel A1, out of all firms whose sales grow above the median rate each year over the prior three years, on average 305 firms survive over the three years following sample selection. Of these, 70 firms have above-median growth rates in each of the three postselection years. They represent 22.8 percent of the survivors, compared to the 12.5 percent that is expected by chance. Growth in income, on the other hand, is an entirely different matter (Panels A2 and A3). For example, there are 222 firms with the impressive track record of above-median growth in income before extraordinary items in each of the three prior years and that survive over the following three years. Yet over the postselection period, only 28 or 12.5 percent manage to repeat and beat the median over all available firms each year. This matches the number expected under the null hypothesis of independence. Although sample sizes become much smaller in the case of firms with favorable growth over the past five years, the findings are similar. Starting out with roughly 2,900 eligible firms on average, 43 firms enjoy a run over the preceding five years for growth in income before extraordinary items and survive over the subsequent five years. In these five years, the percentage of firms who manage to repeat the run is 5.1 percent, while the percentage expected by chance is 3.1 percent. This corresponds to only one run more than expected, however, so the difference is not outstanding.

Table V

Persistence in Growth Rates of Operating Performance: Firms with Superior and Poor Past Growth

At every calendar year-end over the sample period, growth rates in operating performance are calculated over each of the following one to five years (or until delisting) for firms with superior (part I of the table) or inferior (part II) past growth in operating performance. Firms with superior (inferior) past growth include: firms with above-median (below-median) operating performance growth each year over the past five or past three years; firms whose average rank on growth rate each year over the past five or past three years falls in the top (bottom) quartile. The sample period is 1951 to 1998, and eligible firms include all domestic firms listed on the New York, American, and Nasdaq markets with data on the Compustat files. Operating performance is measured as sales (panel 1), operating income before depreciation (panel 2), or income before extraordinary items available to common equity (panel 3). Growth in each variable is measured on a per share basis as of the sample formation date, with the number of shares outstanding adjusted to reflect stock splits and dividends; cash dividends and special distributions are also reinvested. For each of the following five years, the number of firms with valid growth rates, the number of firms whose growth rate exceeds the median growth rate each year for the indicated number of years, the percentage these firms represent relative to the number of valid firms, and the percentage expected under the hypothesis of independence across years are reported.

	Part I: Firms with Superior Past Growth (A) Firms with Past Above-Median Run									
	Firms with Above-Median Growth each Year for Past 5 Years and Above-Median Growth each Year for Number of Future Years:					Firms with Above-Median Growth each Year for Past 3 Years and Above-Median Growth each Year for Number of Future Years:				
	1	2	3	4	5	1	2	3	4	5
					(A1) Sales					
Average Number of Valid Firms	110	103	96	90	83	355	329	305	285	265
Average Number above Median	70	42	26	17	11	209	118	70	42	26
Percent above Median	63.3	41.0	27.3	19.0	13.7	58.9	35.6	22.8	14.8	9.9
					(A2) Operating Income before Depreciation					
Average Number of Valid Firms	61	57	53	50	47	267	245	227	210	194
Average Number above Median	34	18	10	6	4	136	67	34	18	10
Percent above Median	55.9	32.3	19.4	12.2	8.0	51.1	27.2	15.1	8.8	5.3
					(A3) Income before Extraordinary Items					
Average Number of Valid Firms	53	50	47	44	43	259	240	222	207	193
Average Number above Median	28	14	7	4	2	125	57	28	14	7
Percent above Median	51.9	27.8	15.1	8.4	5.1	48.3	23.7	12.5	6.7	3.6
Expected Percent above Median	50.0	25.0	12.5	6.3	3.1	50.0	25.0	12.5	6.3	3.1

Table V—*continued*

	(B) Firms with Past Average Growth Rank in Top Quartile									
	Firms with Average Growth Rank over Past 5 Years in Top Quartile and Above-Median Growth each Year for Number of Future Years					Firms with Average Growth Rank over Past 3 Years in Top Quartile and Above-Median Growth each Year for Number of Future Years				
	1	2	3	4	5	1	2	3	4	5
(B1) Sales										
Average Number of Valid Firms	78	71	66	61	56	204	187	172	159	147
Average Number above Median	47	27	16	10	6	120	67	39	24	15
Percent above Median	60.8	37.7	24.4	16.6	11.4	58.9	35.8	22.8	14.8	9.9
(B2) Operating Income before Depreciation										
Average Number of Valid Firms	35	32	30	27	25	133	121	110	100	91
Average Number above Median	18	8	4	2	1	65	31	15	8	4
Percent above Median	50.6	26.4	15.0	8.9	5.9	49.0	25.4	13.6	7.6	4.7
(B3) Income before Extraordinary Items										
Average Number of Valid Firms	29	27	25	23	22	121	112	103	94	86
Average Number above Median	13	5	3	1	0	56	24	11	5	2
Percent above Median	44.0	19.6	10.2	4.8	2.1	46.4	21.5	10.4	5.5	2.6

Part II. Firms with Inferior Past Growth
(C) Firms with Past Below-Median Run

	Firms with Below Median Growth each Year for Past 5 Years and Above-Median Growth each Year for Number of Future Years:					Firms with Below Median Growth each Year for Past 3 Years and Above-Median Growth each Year for Number of Future Years:				
	1	2	3	4	5	1	2	3	4	5
(C1) Sales										
Average Number of Valid Firms	106	92	82	73	66	343	302	270	244	221
Average Number above Median	35	15	7	4	2	125	59	28	14	7
Percent above Median	33.0	16.3	8.6	4.9	2.5	36.4	19.4	10.6	5.9	3.4

				(C2) Operating Income before Depreciation						
Average Number of Valid Firms	39	35	32	30	28	229	206	186	170	156
Average Number above Median	20	9	5	2	1	122	58	27	13	6
Percent above Median	51.4	25.7	14.3	6.3	3.5	53.3	28.0	14.7	7.6	3.6
				(C3) Income before Extraordinary Items						
Average Number of Valid Firms	33	30	28	26	25	220	201	184	170	157
Average Number above Median	18	9	4	2	1	127	61	28	13	5
Percent above Median	56.2	30.2	14.8	6.7	3.0	57.7	30.4	15.3	7.7	3.4
Expected Percent above Median	50.0	25.0	12.5	6.3	3.1	50.0	25.0	12.5	6.3	3.1

	(D) Firms with Past Average Growth Rank in Bottom Quartile									
	Firms with Average Growth Rank over Past 5 Years in Bottom Quartile and Above-Median Growth each Year for Number of Future Years					Firms with Average Growth Rank over Past 3 Years in Bottom Quartile and Above-Median Growth each Year for Number of Future Years				
	1	2	3	4	5	1	2	3	4	5
					(D1) Sales					
Average Number of Valid Firms	86	74	65	57	51	202	175	154	137	123
Average Number above Median	29	12	6	3	1	71	32	14	6	3
Percent above Median	33.1	16.7	8.6	4.4	2.3	35.2	18.1	9.3	4.5	2.3
				(D2) Operating Income before Depreciation						
Average Number of Valid Firms	23	20	17	15	14	111	97	86	77	70
Average Number above Median	15	7	3	1	1	68	33	15	7	3
Percent above Median	63.8	34.8	19.8	8.9	4.2	61.8	33.7	17.5	8.7	4.1
				(D3) Income before Extraordinary Items						
Average Number of Valid Firms	18	16	14	13	12	100	89	80	72	66
Average Number above Median	13	7	4	2	1	68	34	16	7	3
Percent above Median	73.5	47.1	25.1	12.1	5.3	68.1	38.9	20.7	10.3	9.8

The results caution against extrapolating past success in income growth into the future.

A firm may have extraordinary past growth even though it slips below the median for one or two years, as long as growth in the other years is very high. To include such cases of successful past growth, we use a different criterion for what qualifies as superior past growth. In particular, we also classify firms by their average growth ranks. At every calendar year-end over the sample period, we assign each firm a score based on its past growth. The score is obtained by looking back over each of the preceding five (or three) years, ranking the firm's growth rate each year relative to all available firms (where the firms with the highest growth rate and the lowest growth rate get ranks of one and zero, respectively), and then averaging the ranks over five (or three) years. Firms whose average ranks fall in the top quartile are classified as firms with superior past growth in Panel B. While high past sales growth foretells high future sales growth, there are still no signs of persistence in growth of income before extraordinary items in Panel B3. Out of the firms who survive for three years following sample selection, 103 firms have an average rank based on growth over the preceding three years falling in the top quartile. Only 11 or 10.4 percent of them have above-median runs in the three postselection years, amounting to 2 less than the expected number.

In Part II of Table V, Panel C performs the same analysis for firms with below-median growth over each of the past five or past three years. However, survivorship bias is a particularly grave concern here. After a long period of lackluster performance, the firms that are left standing at the end of the following period are particularly likely to be those who post relatively high growth rates. From Panel C1, future sales growth is persistently low. The fraction of above-median runs in sales growth is notably lower than the expected percentage. On the other hand, they are not less likely to achieve favorable above-median runs with regard to future growth in income. For example, looking at firms with a below-median run for the past three years, over the following three- and five-year horizons, the actual (expected) proportions of above-median runs are 15.3 (12.5) and 3.4 (3.1) percent for growth in income before extraordinary items. While survivorship bias makes it difficult to draw a definitive conclusion, it does not appear that, going forward, the firms with disappointing past growth differ notably from the more successful firms with respect to growth in income.

D. Alternative Criteria for Consistency in Growth

Given the large transitory component of earnings, investors may consider a firm to show persistent growth even if its growth fades for a few years, as long as there is rapid growth for the rest of the time. Even a celebrated example of a growth stock such as Microsoft, for example, falls short of delivering above-median growth in income before extraordinary items for 10 years in a row.[8]

[8] In the 10-year period preceding the latest sample selection date, Microsoft's growth rank of 0.49 in 1994 narrowly misses the median that year.

In Table VI, we adopt more relaxed criteria for defining consistency in growth. In particular, we check whether a firm beats the median for most years over the horizon, but allow it to fall short of the median for one or two years. For example, looking forward from a sample selection date, 269 firms on average have sales growth rates that exceed the median in five out of the following six years. These firms represent 15.6 percent of the surviving firms, more than the expected value of 9.4 percent. In the case of income before extraordinary items, the departures from what is expected under independence are slender, especially over longer horizons. For instance, an average of 9.9 percent have income before extraordinary items growing at a rate above the median for five out of six years, which is close to the expectation of 9.4 percent. Similarly, if we let a firm falter for two years, 4.8 percent of the surviving firms have growth in income before extraordinary items that exceeds the median in 8 out of 10 years, compared to an expected value of 4.4 percent.

As another way to single out cases of sustained high growth while allowing for some slack, we require a firm to post an average annual growth rank over the subsequent five years that falls in the top quartile (where in any year a growth rank of one denotes the highest realized growth rate that year, and zero denotes the lowest rate). The results for this definition of consistency are provided in the last column of Table VI. On average, 1.4 percent of the surviving firms (27 firms) pass this criterion with respect to growth of income before extraordinary items. Assuming independence, the expected value is 2.5 percent.

In summary, analysts' forecasts as well as investors' valuations reflect a widespread belief in the investment community that many firms can achieve streaks of high growth in earnings. Perhaps this belief is akin to the notion that there are "hot hands" in basketball or mutual funds (see Camerer (1989) and Hendricks, Patel, and Zeckhauser (1993)). While there is persistence in sales growth, there is no evidence of persistence in terms of growth in the bottom line as reflected by operating income before depreciation and income before extraordinary items. Instead, the number of firms delivering sustained high growth in profits is not much different from what is expected by chance. The results for subsets of firms, and under a variety of definitions of what constitutes consistently superior growth, deliver the same verdict. Put more bluntly, the chances of being able to identify the next Microsoft are about the same as the odds of winning the lottery. This finding is what would be expected from economic theory: Competitive pressures ultimately dissipate excess earnings, so profitability growth reverts to a normal rate.

IV. The Behavior of Nonsurvivors

Survivorship bias is a serious concern in our tests. By necessity, we condition on surviving into the future in order to calculate growth rates and to carry out our runs tests. Moreover, in our runs tests, the survivors are compared each year to all firms (survivors and nonsurvivors) available that year. To gauge the poten-

Table VI

Distribution of Firms Classified by Above-Median Growth in Operating Performance over Indicated Horizon: All Firms

At every calendar year-end over the sample period, growth rates in operating performance are calculated over each of the following one to ten years (or until delisting) for all firms in the sample. The sample period is 1951 to 1998, and the sample includes all domestic firms listed on the New York, American, and Nasdaq markets with data on the Compustat files. Operating performance is measured as sales (panel A), operating income before depreciation (panel B), or income before extraordinary items available to common equity (panel C). Growth in each variable is measured on a per share basis as of the sample formation date, with the number of shares outstanding adjusted to reflect stock splits and dividends; cash dividends and special distributions are also reinvested. The table reports the average number of firms with above-median growth in each of the indicated categories, as well as the percentage these firms represent relative to the number of valid firms; the last row reports the percentage expected under the hypothesis of independence across years. Statistics are provided for the entire sample period and for the ten-year horizon corresponding to the last sample formation year.

	Firms with Above-Median Growth							
Variable	3 out of 4 years	4 out of 5 years	5 out of 6 years	6 out of 7 years	6 out of 8 years	7 out of 9 years	8 out of 10 years	Firms with Average Growth Rank in Top Quartile over 5 Years
					(A) Sales			
Average Number	697	432	269	170	287	191	127	79
Percent	33.9	23.0	15.6	10.7	19.5	14.0	10.0	4.2
1989–1998	36.6	26.0	18.0	12.6	21.4	16.0	12.7	5.6
					(B) Operating Income before Depreciation			
Average Number	629	341	184	100	205	119	70	34
Percent	31.2	18.6	10.9	6.5	14.4	9.0	5.7	1.9
1989–1998	31.7	19.3	11.5	7.4	15.1	10.4	8.0	2.0
					(C) Income before Extraordinary Items			
Average Number	634	334	171	88	190	109	61	27
Percent	30.7	17.7	9.9	5.5	12.9	8.0	4.8	1.4
1989–1998	29.9	16.5	8.4	5.0	12.8	8.4	5.7	0.9
Expected Percent	25.0	15.6	9.4	5.5	10.9	7.0	4.4	2.5

tial magnitude of the problem, in this section, we replicate some of our tests on firms who do not survive over the entire future horizon.

Specifically, we examine two sets of stocks. Given our focus on long-horizon growth, we first select at each year-end a sample of firms who survive over the full 10-year following period. The behavior of these (the survivors) is compared to a second set (the nonsurvivors) that also includes firms who do not last for the full period. To strike a balance between the mix of survivors and nonsurvivors in this second set, we require firms to survive for the first five years after sample selection, but they may drop out between the 6th to 10th year of the postselection period.

The results are reported in Panels A and B of Table VII. The survivors have a higher chance than expected for achieving runs above the median in growth of income before extraordinary items. Conversely, the fraction of runs is lower for the set of nonsurvivors. Of the survivors, for example, 3.4 percent sustain runs for five years of growth in income before extraordinary items above the median (where the expected proportion is 3.1 percent). The corresponding percentage for nonsurvivors is 2.3 percent. Nonetheless, the differences across the two sets are generally not substantial. Panels C and D apply the same procedure to the technology stocks considered in Table IV. Here the differences across the two sets are more notable. At the five-year horizon, for example, 5.2 percent of the survivors achieve runs above the median for growth in income before extraordinary items, compared to 3.2 percent of the nonsurvivors.

Finally, Panels A and B of Part II of Table VII give the distribution of one-year growth rates for the two sets of firms (where the percentiles are averaged across all sample selection years). The results confirm that survivors realize higher growth rates than nonsurvivors. For example, the median growth in income before extraordinary items for the survivors averages 10.6 percent, compared to 8.2 percent for nonsurvivors.

V. The Predictability of Growth: Valuation Ratios

Based on the historical record, it is not out of the question for a firm to enjoy strong growth in excess of 20 percent a year for prolonged periods. The issue, however, is whether such firms are identifiable ex ante. Our attempts in the previous sections to uncover cases of persistently high future growth using information such as past growth, industry affiliation, value–glamour orientation, and firm size have limited success. In this section, we expand our search for predictability by investigating whether valuation indicators such as earnings-to-price, book-to-market, and sales-to-price ratios distinguish between firms with high or low future growth. Further, several studies suggest that investors are prone to judgmental biases, so they respond to past growth by extrapolating performance too far into the future (see, e.g., La Porta (1996) and La Porta et al. (1997)). Consequently, after a period of above- or below-average growth, the valuations of firms with high (low) realized growth may be pushed too high (or too low).

In Table VIII, stocks are sorted into deciles at each year-end on the basis of their growth rate in income before extraordinary items over the following five years (Panel A) or over the following 10 years (Panel B). Within each decile, we

Table VII

Results for Surviving versus Non-Surviving Firms: Persistence Tests and Growth Rates

At every calendar year-end over the sample period, two sets of firms are selected: firms that survive over the following ten years (survivors), and firms that survive over the following five years but thereafter fail to survive until the tenth year (nonsurvivors). For each set of firms, growth rates in operating performance are calculated over each of the following ten years. The sample period is 1951 to 1998, and all domestic firms listed on the New York, American, and Nasdaq markets with data on the Compustat files are eligible. Operating performance is measured as sales, operating income before depreciation, or income before extraordinary items available to common equity. Growth in each variable is measured on a per share basis as of the sample formation date, with the number of shares outstanding adjusted to reflect stock splits and dividends; cash dividends and special distributions are also reinvested. Part I provides runs tests of persistence over each of the following ten years for the two sets of firms: the average number of firms whose growth rate exceeds the median growth rate each year for the indicated number of years is expressed as a percentage of the number of firms with valid growth rates. Part II reports the distribution of annualized growth rates realized over the sixth to tenth year (or until delisting) following sample selection for the two sets of firms. The simple average over the entire sample period of the percentiles is reported.

Part I: Runs Tests for Persistence										
	Percent of Firms with Above-Median Growth each Year for Number of Years:									
Variable	1	2	3	4	5	6	7	8	9	10
				(A) Survivors (1265 firms)						
Sales	52.8	30.9	18.1	10.8	6.6	4.2	2.7	1.8	1.3	0.9
Operating Income before Depreciation	51.5	26.8	13.7	7.0	3.8	2.1	1.2	0.7	0.5	0.3
Income before Extraordinary Items	51.7	26.9	13.5	6.7	3.4	1.8	1.0	0.5	0.3	0.2
				(B) Non-Survivors						
Number of Firms	445	445	445	445	445	344	250	165	86	0
Sales	48.7	26.6	14.6	8.1	4.5	2.8	1.7	1.1	0.8	—
Operating Income before Depreciation	50.0	24.2	11.5	5.5	2.5	1.3	0.7	0.5	0.3	—
Income before Extraordinary Items	49.1	23.8	11.1	5.1	2.3	1.1	0.6	0.3	0.1	—
				(C) Survivors, Technology (195 firms)						
Sales	54.6	33.2	20.5	12.9	8.4	5.8	4.2	3.0	2.3	1.7
Operating Income before Depreciation	53.6	29.7	16.5	9.6	5.9	3.6	2.2	1.4	1.0	0.7
Income before Extraordinary Items	54.1	29.9	16.3	9.0	5.2	3.1	1.9	1.1	0.6	0.4

				(D) Non-Survivors, Technology						
Number of Firms	100	100	100	100	100	77	55	37	20	0
Sales	51.5	28.6	16.7	10.6	6.5	4.6	3.1	2.0	1.4	—
Operating Income before Depreciation	49.5	24.3	12.4	6.6	3.3	2.0	1.4	1.3	1.0	—
Income before Extraordinary Items	50.1	25.0	12.4	6.7	3.2	1.7	1.0	0.5	0.0	—
Expected Percent above Median	50.0	25.0	12.5	6.3	3.1	1.6	0.8	0.4	0.2	0.1

Part II: Annualized Growth Rates

	Percentile									
Variable	2%	10%	25%	40%	50%	60%	75%	95%	98%	
					(A) Survivors					
Sales	− 15.4	− 2.0	5.6	9.1	10.9	12.5	15.5	21.7	37.6	
Operating Income before Depreciation	− 23.3	− 6.8	2.8	7.6	10.1	12.5	16.9	25.5	48.0	
Income before Extraordinary Items	− 28.6	− 8.6	2.1	7.7	10.6	13.3	18.1	28.4	56.4	
					(B) Non-Survivors					
Sales	− 18.5	− 7.0	1.0	6.0	8.4	10.4	13.9	20.3	36.8	
Operating Income before Depreciation	− 26.1	− 12.5	− 2.6	4.7	8.1	11.5	16.3	25.7	47.9	
Income before Extraordinary Items	− 27.4	− 14.5	− 3.3	4.4	8.2	11.9	17.9	28.6	55.9	-

Table VIII

Valuation Ratios and Characteristics at Beginning and End of Horizon for Firms Classified by Growth in Income before Extraordinary Items

At every calendar year-end over the sample period, growth rates in income before extraordinary items available to common equity are calculated over the following five and ten years for all firms in the sample. The sample period is 1951 to 1998, and the sample includes all domestic firms listed on the New York, American, and Nasdaq markets with data on the Compustat files. Growth rates are measured on a per share basis as of the sample selection date, with the number of shares outstanding adjusted to reflect stock splits and dividends; cash dividends and special distributions are also reinvested. Firms are classified into one of ten equally-sized categories based on their realized five- and ten-year growth rates. The following statistics are calculated for firms within each category: the median realized annual growth rate over the horizon; the average size decile rank at the beginning and end of the growth horizon; median valuation ratios at the beginning and at the end of the horizon. The ratios are the prior year's income before extraordinary items to price (*EP*), net sales to price (*SP*), and book value to market value of common equity (*BM*). Results are averaged over all years in the sample period, and are also reported for the last five- or 10-year period. Panel A of the table provides results for firms classified by growth rates over five years and for firms with above-median growth each year for five consecutive years; Panel B provides results for firms classified by ten-year growth rates.

Panel A: Classified by Annualized Growth Rate over 5 Years

	Decile										
Variable	1	2	3	4	5	6	7	8	9	10	5-year run above median
Median Growth Rate	−18.9	−5.0	1.5	5.8	9.1	12.0	15.1	18.9	25.1	41.7	40.9
Beginning Size Decile Rank	4.118	4.773	5.087	5.423	5.447	5.526	5.338	4.989	4.273	3.272	3.699
Ending Size Decile Rank	3.526	4.414	4.831	5.275	5.452	5.668	5.652	5.482	5.056	4.243	5.163
Beginning Median EP Ratio	0.083	0.085	0.086	0.083	0.084	0.082	0.082	0.082	0.079	0.068	0.061
At Start of Last 5-year Period	0.050	0.056	0.059	0.055	0.060	0.055	0.052	0.047	0.037	0.021	0.033
Ending Median EP Ratio	0.055	0.073	0.078	0.080	0.082	0.081	0.080	0.079	0.077	0.075	0.066
At End of Last 5-year Period	0.033	0.047	0.052	0.053	0.052	0.052	0.049	0.050	0.046	0.042	0.040
Beginning Median BM Ratio	0.650	0.654	0.678	0.665	0.685	0.679	0.694	0.726	0.777	0.880	0.694
At Start of Last 5-year Period	0.465	0.485	0.476	0.465	0.494	0.430	0.458	0.437	0.452	0.537	0.446
Ending Median BM Ratio	1.115	0.927	0.845	0.789	0.755	0.700	0.669	0.610	0.574	0.560	0.369
At End of Last 5-year Period	0.549	0.495	0.501	0.461	0.402	0.367	0.350	0.337	0.291	0.292	0.200
Beginning Median SP Ratio	1.723	1.576	1.473	1.304	1.370	1.276	1.328	1.530	1.791	2.323	1.684
At Start of Last 5-year Period	0.962	1.022	1.079	0.825	0.890	0.807	0.822	1.065	1.052	1.423	0.914
Ending Median SP Ratio	2.606	2.062	1.783	1.501	1.422	1.288	1.274	1.305	1.377	1.503	1.012
At End of Last 5-year Period	1.174	0.860	0.972	0.638	0.653	0.587	0.573	0.649	0.563	0.681	0.460

Table VIII—*continued*

Panel B: Classified by Annualized Growth Rate over 10 years

Median Growth Rate	−10.8	−3.4	−0.3	2.1	3.9	5.6	7.4	9.4	12.4	19.3
Beginning Size Decile Rank	4.565	5.223	5.577	5.641	5.597	5.508	5.563	5.480	5.040	3.890
Ending Size Decile Rank	3.950	5.087	5.608	5.818	5.882	5.921	5.981	6.100	5.851	5.100
Beginning Median EP Ratio	0.088	0.088	0.087	0.087	0.087	0.086	0.085	0.081	0.080	0.069
At Start of Last 10-year Period	0.072	0.070	0.077	0.073	0.074	0.065	0.068	0.066	0.056	0.039
Ending Median EP Ratio	0.057	0.072	0.076	0.079	0.081	0.083	0.084	0.082	0.082	0.079
At End of Last 10-year Period	0.035	0.047	0.050	0.053	0.048	0.054	0.056	0.049	0.044	0.049
Beginning Median BM Ratio	0.653	0.699	0.696	0.699	0.726	0.707	0.723	0.706	0.742	0.817
At Start of Last 10-year Period	0.550	0.605	0.548	0.564	0.595	0.543	0.609	0.504	0.597	0.724
Ending Median BM Ratio	1.048	0.860	0.796	0.761	0.748	0.734	0.725	0.673	0.647	0.622
At End of Last 10-year Period	0.626	0.482	0.382	0.439	0.392	0.396	0.409	0.321	0.343	0.337
Beginning Median SP Ratio	1.664	1.560	1.470	1.392	1.429	1.399	1.415	1.408	1.503	2.022
At Start of Last 10-year Period	1.405	1.417	1.164	1.285	1.054	1.106	1.211	1.133	1.455	1.409
Ending Median SP Ratio	2.619	1.928	1.648	1.531	1.535	1.477	1.478	1.411	1.385	1.468
At End of Last 10-year Period	1.520	0.941	0.735	0.853	0.758	0.826	0.805	0.664	0.724	0.756

calculate the median realized growth rate, as well as median characteristics such as size decile rank and valuation ratios. This is done at the beginning of the 5- or 10-year growth horizon and also at the end of the horizon. We report results averaged across all sample selection years, as well as results for the most recent 5-year or 10-year growth horizon in our sample period.

We focus the discussion on Panel A of the table (the results are similar for the 10-year horizon). In line with the results from Tables I and II, the stocks in the extreme growth deciles tend to be smaller firms. The median firm in the top decile (with a growth rate of 41.7 percent a year) falls in the third size decile, while the median firm in the bottom decile (with a growth rate of -18.9 percent) ranks in the fourth size decile. Over the following 5 years, however, the high-growth firms perform relatively well, resulting in a surge in their market values. Conversely, the market values of the low-growth firms show a relative slump.

Sorting by realized future growth induces a mechanical association between growth rates and the level of earnings at the beginning and end of the growth horizon. To weaken this link, we measure earnings one year prior to the base year (or one year before the final year) of the growth horizon. The price is measured at the start or end of the horizon, so the numbers correspond to the conventional measure of trailing earnings yield that is widely used in practice and research. There is reason to be wary about relying too heavily on the earnings yield variable, however, because net income is the most problematic of our measures of operating performance. For example, a firm may have a low earnings yield because its price impounds investors' expectations of high growth in future earnings, but another reason may be its recent performance has been poor and its earnings are currently depressed. On this account, earnings-to-price ratios are not generally used in academic research, or investment industry analysis, to classify firms as "value" or "glamour" stocks. Instead other, better-behaved, indicators such as the book-to-market ratio, are favored.

The top decile of growth firms at the beginning of the growth horizon has a median earnings-price ratio (0.068) that is much lower than the others (which cluster around 0.08). The low earnings yield for this group is consistent with the notion that the market's valuation accurately incorporates future growth. On the other hand, decile portfolios 8 and 9, which also show relatively strong growth, do not have notably low earnings yields. Rather, the association for the highest-growth decile may reflect cases where firms grow from a depressed level of income. At the end of the growth horizon, only the earnings-price ratio of the bottom decile of firms is eye-catching. Contrary to intuition, however, these firms have comparatively low earnings yields so they appear to be relatively "expensive." Instead, the explanation here may also lie in their low earnings levels, since they have gone through a period of disappointing growth.

Given the shortcomings of the earnings yield variable, we also look at valuation measures that tend to be better-behaved. Table VIII provides median ratios of book-to-market and sales-to-price at the beginning and end of the growth horizon for each decile. Firms which are ranked in the highest decile by earnings growth have relatively high sales-to-price and book-to-market ratios at the beginning. For example, their median book-to-market ratio is 0.880 (compared to 0.690

averaged across the other groups) and the median sales-to-price multiple is 2.323 (compared to 1.486 for the other groups). The modest ex ante valuations suggest that the market fails to anticipate their subsequent growth.

On the other hand, ex post valuations closely track prior growth. The top decile of high-growth firms have ending book-to-market and sales-to-price ratios of 0.560 and 1.503, respectively. These are substantially lower than the averages across all the other groups. This finding fits in with earlier evidence on the existence of extrapolative biases in investors' expectations about future growth (see La Porta (1996) and La Porta et al. (1997)).

The last column in Panel A of Table VIII provides corresponding statistics for firms whose income before extraordinary items grows above the median rate for five consecutive years. The difference between these firms' valuation ratios at the beginning and end of the growth horizon is striking. At the beginning, their book-to-market and sales-to-price ratios are not too far out of line from the average, suggesting that their future performance is not foreseen by the market. However, at the end of the growth horizon, the median book-to-market and sales-to-price ratios of this group are the lowest in Table VIII. The rich ending multiples such firms command highlight the importance investors attach to consistently superior growth, and not just high growth per se. Investors handsomely reward firms that have achieved several consecutive years of strong growth, and believe they will continue the streak (counterfactually, as the results in Table V indicate).

In summary, the results suggest that market valuation ratios have little ability to sort out firms with high future growth from firms with low growth. Instead, in line with the extrapolative expectations hypothesis, investors tend to key on past growth. Firms that have achieved high growth in the past fetch high valuations, while firms with low past growth are penalized with poor valuations.

VI. Comparisons with IBES Consensus Forecasts

Security analysts' estimates of near-term earnings are widely disseminated and receive much attention. Dramatic movements in a stock's price can arise when an influential analyst issues a revised earnings estimate. Possibly, therefore, analysts' estimates of long-term earnings growth may also be useful in forecasting future growth over longer horizons. Analysts are not shy about making aggressive growth forecasts either (the dispersion between the top and bottom decile of IBES long-term forecasts is about 31 percent), so they apparently are confident in their own ability to pick the future success stories.

The current dividend yield on a stock may also have predictive power for future growth in earnings per share. Standard textbook analysis suggests that, given a firm's investment policy and ignoring tax effects, it is a matter of indifference to a shareholder whether earnings are paid out as current dividends or retained for growth in future dividends. For example, a firm may choose to raise the amount paid out from earnings as dividends to current shareholders. To maintain investment, however, it must use external financing, thereby diluting current shareholders' claims to future profits. In other words, high current dividends come at the expense of low future growth per share. To use a simple constant-growth dividend discount model as an illustration, given investors' required rate of return, there is a one-to-one trade-off between future growth per share and the dividend yield. Furthermore, a firm's dividend payout may signal whether it has attractive investment projects available to fuel future growth.

To allow a cleaner comparison with analysts' forecasts, which do not include dividends, in the remainder of the paper, we drop our convention of reinvesting dividends when we calculate growth rates. Analysts' predictions refer to growth in income before extraordinary items, but realized growth in this variable is highly prone to measurement problems (such as the exclusion of cases with negative base-year values for income). For this reason, we also report realized growth in sales and operating income before depreciation. Growth rates in these variables are correlated with growth in income before extraordinary items, but are better behaved and are available for a much larger fraction of the sample.

A. *Individual Firm Growth Rates*

Table IX relates IBES consensus long-term growth forecasts to realized future growth. At each year-end, we rank all domestic firms with available IBES long-term forecasts and sort them into quintiles. IBES long-term estimates do not become available until 1982, so the sample period in Table IX runs from 1982 to 1998. The breakpoints for the sort use all NYSE firms available as of the sample selection date (regardless of whether they survive in the future). In Table IX, we track the subsequent growth rates of firms who survive over the next one, three, or five years in each quintile. The median realized growth rate over firms in each quintile is then averaged across all sample selection dates.

The dispersion in IBES consensus growth forecasts is large, so analysts are boldly distinguishing between firms with high and low growth prospects. The median estimate in quintile 1 averages 6 percent, while the median estimate in quintile 5 is 22.4 percent on average.[9] Notably, analysts' estimates are quite optimistic. Over the period 1982 to 1998, the median of the distribution of IBES growth forecasts is about 14.5 percent, a far cry from the median realized five-year growth rate of about 9 percent for income before extraordinary items.[10]

Near-term realized growth tends to line up closely with the IBES estimate (Panel A). In the first postranking year, the median growth rate in income before extraordinary items is 18.3 percent on average for quintile 5, and 5.1 percent on average for quintile 1. The difference between the growth rates for the other quintile portfolios is much milder, however. Comparing quintiles 4 and 2, median growth rates in income before extraordinary items are apart by only 2.5 percent.

A naive model for predicting future growth uses the dividend yield, and is based on the trade-off between current dividends and future growth. Suppose,

[9] Note that since the breakpoints are based on NYSE stocks only, the number of stocks differs across the quintiles. In particular, many firms penetrate the top quintile.

[10] To sharpen the point, note that the median realized growth rate of nine percent (without dividends reinvested) is based on all firms, including smaller firms that tend to be associated with somewhat higher growth rates. IBES forecasts, on the other hand, predominantly cover larger firms.

Table IX

Realized Median Growth Rates of Operating Performance for Stocks Classified by IBES Long-Term Growth Forecasts

At every calendar year-end t over the sample period, stocks are ranked and classified to one of five groups based on IBES forecasts of long-term earnings growth. Results are reported for individual stocks and for portfolios. For individual stocks, growth rates in operating performance are calculated over each of the five subsequent years (years $t+1$ to $t+5$) for all firms in the sample with available data. The sample period is 1982 to 1998, and all domestic firms listed on the New York, American, and Nasdaq markets with data on the Compustat files are eligible. Operating performance is measured as sales, operating income before depreciation, or income before extraordinary items available to common equity. Growth in each variable is measured on a per share basis as of the sample formation date, with the number of shares outstanding adjusted to reflect stock splits and dividends. The median realized growth over all stocks in each classification is calculated each year, and the simple average over the entire sample period is reported. For portfolios, a value-weighted portfolio is formed at each year-end from all the stocks in each quintile sorted by IBES forecasts. The portfolio's income before extraordinary items is calculated over each of the subsequent five years, with the proceeds from liquidating delisted stocks reinvested in the surviving stocks. Growth rates for each portfolio are calculated in each formation year, and the simple average over the entire sample period of the growth rates is reported. Also reported are the ratios of the prior year's income before extraordinary items per share to current price, and the prior year's cumulative regular dividends per share to current price.

	Quintile Based on IBES Forecast:				
Growth in:	1 (Low)	2	3	4	5 (High)
(A) Growth Rate in Year t+1					
Sales	1.4	4.5	6.3	8.3	13.7
Operating Income before Depreciation	3.6	6.8	7.6	10.3	16.0
Income before Extraordinary Items	5.1	9.5	10.1	12.0	18.3
Portfolio Income before Extraordinary Items	12.6	4.2	4.5	7.2	13.6
No. with Positive Base & Survive 1 year	242	256	266	318	584
No. with Negative Base & Survive 1 year	71	78	60	88	265
(B) Growth Rate in Year t+2					
Sales	1.7	4.5	6.4	7.8	11.6
Operating Income before Depreciation	3.2	7.0	8.4	9.9	14.0
Income before Extraordinary Items	4.7	9.9	10.5	12.2	16.4
Portfolio Income before Extraordinary Items	6.9	7.5	6.1	9.1	10.6
No. with Positive Base & Survive 2 years	225	235	244	296	497
No. with Negative Base & Survive 2 years	62	75	59	85	252
(C) Annualized Growth Rate over 3 Years					
Sales	1.1	4.0	5.6	7.3	11.3
Operating Income before Depreciation	2.5	5.2	6.8	8.1	10.9
Income before Extraordinary Items	3.1	7.4	7.0	9.0	11.5
Portfolio Income before Extraordinary Items	9.0	7.3	5.2	7.1	11.4
No. with Positive Base & Survive 3 years	202	209	230	263	439
No. with Negative Base & Survive 3 years	67	70	56	82	217
(D) Annualized Growth Rate over 5 Years					
Sales	1.2	3.4	5.1	6.9	9.9
Operating Income before Depreciation	2.2	5.1	6.8	7.3	9.2
Income before Extraordinary Items	2.0	6.5	6.5	8.0	9.5
Portfolio Income before Extraordinary Items	8.0	10.7	7.2	7.7	11.3
No. with Positive Base & Survive 5 years	182	179	201	233	356
No. with Negative Base & Survive 5 years	57	63	50	68	170
Median IBES Forecast	6.0	10.2	12.3	15.1	22.4
Median Stock Dividend Yield, %	6.0	3.4	2.7	1.5	0.1
Portfolio Dividend Yield, %	6.9	4.6	3.3	2.5	1.3
Median Stock Earnings to Price Ratio, %	10.0	8.9	7.9	7.2	5.6

as a first approximation, that all stocks have the same long-term expected return. Given this, the naive model forecasts a spread in future growth across stocks that is identical to the spread in their current dividend yields (but in the opposite direction). The naive forecast is quite successful at picking up differences in growth across the intermediate quintiles. Over the first postranking year, the difference between the dividend yields of quintiles 2 and 4 (3.4 and 1.5 percent, respectively) corresponds roughly to the difference in their growth rates. Once differences in the dividend yield are taken into account, then, IBES estimates have forecast power for realized growth over the first year only at the extremes.

In general, IBES long-term forecasts refer to a three- to five-year horizon, so the behavior of realized growth over these horizons is more interesting. Median realized growth rates over three years and over five years are reported in Panels C and D. These panels highlight the upward bias in analysts' long-term growth estimates. In every quintile, median forecasts exceed median realized growth rates, with the most pronounced bias in quintile 5. For five-year growth in income before extraordinary items, for example, the median forecast in the top quintile is 22.4 percent, much higher than the median realized growth rate, which is only 9.5 percent. Furthermore, the realized growth rate for the firms in the top quintile should be taken with a grain of salt. In the highest-ranked quintile, the percentage of firms who survive for the full five postranking years is lower than for any of the other quintiles. For example, there are 849 firms on average who survive in the first postranking year in quintile 5, but this drops to 526 by the fifth year, so about 38 percent of the firms drop out between the first and fifth years. For quintile 3, the corresponding counts are 326 and 251, respectively, so 23 percent disappear from the sample. The upshot is that realized growth in income before extraordinary items is likely to be somewhat overstated for firms in the top quintile.

Over longer horizons, analysts' growth estimates still do not add much information beyond what is contained in the dividend yield. For example, the median realized five-year growth rate is 9.5 percent for the highest-ranked quintile by IBES forecasts, compared to 2 percent for the lowest-ranked quintile. The difference of 7.5 percent is not much higher than the spread in their dividend yields. The yields are 0.1 percent and 6 percent for the highest and lowest ranked quintiles, respectively, so the dividend yield spread is 5.9 percent. The results for growth in operating income before depreciation yield similar conclusions.

To sum up, analysts forecast that long-term earnings growth for the top quintile outperforms the bottom quintile by 16.4 percent. The realized gap in five-year growth rates, however, is only 7.5 percent. Much of the spread in realized growth reflects differences in dividend yields, and some is due to survivorship bias in the top quintile. After accounting for these influences, analyst forecasts add information only over shorter horizons.

B. Portfolio Growth Rates

Issues of survivorship bias and low or negative base-year values for income before extraordinary items are major concerns. Table IX takes another approach to measuring growth rates that tries to work around these concerns. Specifically, after ranking stocks by IBES long-term forecasts at each year-end, we form a value-weighted portfolio of the stocks in each quintile. Value-weighting affords some degree of robustness to our measures, to the extent that problems in measuring growth are less severe for large companies. We then track over the postformation period the income before extraordinary items of the portfolio as a whole. If a stock is delisted in a year after portfolio formation, we assume it generates the average income of the remaining firms in that year. Then, at the end of the year, we take the proceeds from liquidating nonsurviving firms and reallocate them proportionally across the surviving stocks. As a result, we are able to use all eligible companies to calculate growth rates, regardless of whether they survive over the full growth horizon, or whether they have positive earnings in the base year.[11] The portfolio approach, however, is not without its drawbacks. As firms drop out of the sample and the funds from their liquidation are reinvested in the remaining firms, over time, the portfolio can build up large stakes in a relatively small number of surviving firms who tend to have relatively high growth rates. The implication is that long-term portfolio growth rates for cases where survivorship bias is acute, such as the fastest-growing firms in the top quintile by IBES forecasts as noted above, should be interpreted with caution.

The results for the portfolios' long-term growth rates are in line with our earlier findings. IBES long-term forecasts are essentially unrelated to realized growth in income before extraordinary items beyond one or two years out. For example, over the five postformation years (Panel D), the bottom and top quintile portfolios on average experience growth rates of 8 and 11.3 percent per year, respectively. The spread of 3.3 percent in the portfolios' growth rates is smaller than the gap between their dividend yields (5.6 percent).

One difference between our results for individual stocks' growth rates and the portfolios' growth rate concerns the performance of the bottom quintile in the first postranking year. In the year immediately following portfolio formation, the bottom quintile portfolio experiences a strong recovery. Its short-term growth rate (12.6 percent) falls slightly short of the top quintile portfolio's growth rate (13.6 percent). This difference from the earlier results based on individual stocks reflects several methodological details, specifically the use of value-weights, the inclusion in the portfolios of nonsurviving firms as well as firms with negative income, and the use of a time-series average of the yearly portfolio growth rates rather than the cross-sectional medians. In particular, since firms with low IBES forecasts generally tend to start with low or negative values of income before extraordinary items at the portfolio formation date, the growth rate over the following year is likely to be high.[12]

Analysts' forecasts substantially overstate realized long-term growth in the top three quintile portfolios. In the top-ranked quintile, for example, the median projected future growth rate is about 22.4 percent, but the portfolio's realized

[11] The portfolio approach to measuring growth rates is described further in Chan et al. (2000, 2001).

[12] Our results parallel the findings for the prospective earnings growth of beaten-down value stocks documented in Lakonishok et al. (1994).

growth is only 11.4 percent over three years and 11.3 percent over five years. These results suggest that, in general, caution should be exercised before relying too heavily on IBES long-term forecasts as estimates of expected growth in valuation studies. The bottom quintile portfolios by IBES forecasts predominantly comprise firms in mature industries whose growth prospects are relatively unexciting, so analysts' estimates come closer to the mark here. For instance, about 25 percent of the firms in the first quintile are utilities.

The long-term estimates of analysts may be overly optimistic for several reasons. One explanation draws on evidence from studies in psychology that individuals' forecasts are susceptible to cognitive biases.[13] For example, the confirmation bias suggests that individuals tend to focus on evidence that supports their beliefs, while downplaying other information that is inconsistent. In this regard, analysts' estimates will be particularly bullish for glamour stocks that have shown strong past growth and which enjoy favorable investor sentiment. In addition, an analyst is employed by a brokerage firm and is expected to make contributions beyond predicting earnings. Up-beat forecasts may encourage trading by investors and thereby raise commission income, as well as generate investment banking business from firms that receive favorable coverage. The general perception is that these aspects of the brokerage and investment banking business are larger, and their links to analysts closer, in the U.S. market than overseas. As one piece of evidence that such considerations may lead to inflated forecasts, IBES estimates as of mid-2001 for U.S. companies project long-term growth of about 18 percent on average. At the same time, in non-U.S. markets, analysts are forecasting long-term growth for companies of roughly the same size to average 11 percent. Perhaps the close ties that exist in practice between the brokerage and investment banking businesses in the U.S. market foster an environment where analysts tend to be less impartial and err on the side of optimism.

VII. Regression Models

We close out our analysis by gathering all the variables we have previously considered individually into one model in order to take our best shot at forecasting growth. Table X reports the results from cross-sectional regressions to predict future growth in operating profits. The model is

$$\begin{aligned} y_{it+j} = {} & \beta_0 + \beta_1 PASTGS5_{it} + \beta_2 EP_{it-1} + \beta_3 G_{it-1} + \beta_4 RDSALES_{it} \\ & + \beta_5 TECH_{it} + \beta_6 BM_{it} + \beta_7 PASTR6_{it} + \beta_8 IBESLTG_{it} + \beta_9 DP_{it} \\ & + \varepsilon_{it+j}. \end{aligned} \tag{1}$$

The dependent variable, y_{it+j}, is the rate of growth for firm i over year $t+j$ in sales (SALES), operating income before depreciation (OIBD), or income before extraordinary items available to common equity (IBEI). We forecast growth over the first year following sample selection, over the three and five years subsequent to sample selection, and over the second to fifth subsequent years.

[13] The evidence is discussed in Kahnemann and Riepe (1998) and Fisher and Statman (2000).

Table X

Forecasting Regressions for Growth Rates of Operating Performance

At every calendar year-end, a cross-sectional regression model is used to forecast growth rates of operating performance, y_{it+j} for firm i over the following one to five years for all firms in the sample with available data. The model is.

$$y_{it+j} = \beta_0 + \beta_1 PASTGS5_{it} + \beta_2 EP_{it-1} + \beta_3 G_{it-1} + \beta_4 RDSALES_{it} + \beta_5 TECH_{it} + \beta_6 BM_{it} + \beta_7 PASTR6_{it} + \beta_8 IBESLTG_{it} + \beta_9 DP_{it} + \varepsilon_{it+j}.$$

The dependent variable is growth in: sales (SALES); operating income before depreciation (OIBD); or income before extraordinary items available to common equity (IBEI). The variables used to forecast a firm's growth are *PASTGS5*, the growth in sales over the five years prior to the sample selection date; *EP*, the ratio of income before extraordinary items available to common equity to equity market value; *G*, the sustainable growth rate given by the product of return on equity (income before extraordinary items available to common equity relative to book equity) and plowback ratio (one minus the ratio of total dividends to common equity to income before extraordinary items available to common equity); *RDSALES*, the ratio of research and development expenditures to sales; *TECH*, a dummy variable with a value of one for a stock in the technology sector and zero otherwise; *BM*, book-to-market ratio; *PASTRGs*, the stock's prior six-month compound rate of return; *IBESLTGs*, the IBES consensus forecast for long-term growth; and *DP* the dividend yield, accumulated regular dividends per share over the last twelve months divided by current price per share.

Growth in:	*PASTGS5*	*EP*	*G*	*RDSALES*	*TECH*	*BM*	*PASTR6*	*IBESLTG*	*DP*	R^2
					(A) Growth Rate in Year t+1					
SALES	0.0890	0.1641	0.0141	0.0979	−0.0038	−0.0184	0.0365	0.3018	−0.5258	0.0709
	(3.7)	(6.0)	(1.5)	(1.6)	(−0.5)	(−4.7)	(3.0)	(6.1)	(−4.8)	
OIBD	−0.0729	−0.2400	0.0064	0.2047	−0.0045	0.0031	−0.0592	0.2334	−0.5390	0.0274
	(−1.3)	(−3.3)	(0.9)	(1.0)	(−0.3)	(0.4)	(−2.4)	(2.6)	(−3.9)	
OBEI	−0.0971	−0.3982	−0.0242	−0.0024	−0.0162	0.0093	−0.0621	0.1179	−0.9564	0.0263
	(−1.4)	(−3.3)	(−1.5)	(−0.0)	(−0.7)	(0.4)	(−2.0)	(0.9)	(−3.5)	
					(B) Annualized Growth Rate over Years t+1 to t+3					
SALES	0.0469	0.1400	0.0099	0.0974	0.0014	−0.0253	0.0311	0.1901	−0.5758	0.0984
	(1.3)	(5.4)	(1.6)	(3.1)	(0.6)	(−9.2)	(6.8)	(9.3)	(−6.4)	
OIBD	−0.0547	−0.0554	0.0014	0.3453	−0.0127	−0.0073	−0.0089	0.1147	−0.4060	0.0296
	(−1.5)	(−1.8)	(0.1)	(3.1)	(−3.2)	(−1.1)	(−1.7)	(2.0)	(−2.6)	
IBEI	0.0087	−0.1881	0.0011	0.3436	−0.0191	−0.0061	−0.0279	0.0758	−0.0630	0.0257
	(0.5)	(−6.0)	(0.1)	(2.4)	(−2.9)	(−0.4)	(−6.5)	(0.9)	(−0.3)	

Table X—*continued*

				(C) Annualized Growth Rate over Years t+1 to t+5						
SALES	0.0252	0.1074	0.0067	0.0931	0.0014	−0.0260	0.0227	0.1538	−0.5446	0.1175
	(0.7)	(10.5)	(3.6)	(6.8)	(0.4)	(−7.4)	(3.2)	(3.1)	(−16.6)	
OIBD	−0.0645	−0.0146	−0.0035	0.3476	−0.0115	−0.0069	−0.0133	0.1227	−0.2675	0.0367
	(−3.0)	(−0.6)	(−0.5)	(7.6)	(−10.3)	(−1.8)	(−2.3)	(1.5)	(−7.4)	
IBEI	−0.0163	−0.1222	−0.0098	0.2493	−0.0133	−0.0095	−0.0293	0.0729	−0.0917	0.0313
	(−4.2)	(−2.3)	(−0.6)	(3.7)	(−3.0)	(−1.0)	(−2.8)	(0.9)	(−0.7)	
SALES	0.1128	0.0351	0.0628	0.2554						0.0507
	(2.7)	(1.8)	(2.3)	(4.3)						
OIBD	−0.0080	−0.0518	−0.0166	0.3779						0.0150
	(−0.2)	(−3.3)	(−0.7)	(13.1)						
IBEI	0.0311	−0.1295	−0.0675	0.2229						0.0148
	(25.5)	(−3.8)	(−1.5)	(2.4)						
				(D) Annualized Growth Rate over Years t+2 to t+5						
SALES	0.0175	0.0983	0.0060	0.1020	0.0007	−0.0273	0.0218	0.1237	−0.5122	0.0902
	(0.5)	(5.0)	(2.9)	(5.6)	(0.2)	(−6.3)	(3.7)	(2.8)	(−20.1)	
OIBD	−0.0665	0.0136	−0.0147	0.3856	−0.0130	−0.0049	−0.0042	0.1354	−0.3197	0.0335
	(−2.1)	(1.0)	(−1.1)	(4.9)	(−7.7)	(−0.9)	(−0.3)	(1.7)	(−2.7)	
IBEI	0.0119	−0.0932	0.0018	0.2897	−0.0174	−0.0075	−0.0245	0.0809	−0.0538	0.0268
	(0.6)	(−2.6)	(0.1)	(12.8)	(−5.8)	(−0.6)	(−1.8)	(1.0)	(−0.4)	
SALES	0.0962	0.0279	0.0655	0.2515						0.0398
	(2.1)	(1.6)	(3.1)	(5.2)						
OIBD	−0.0097	−0.0255	−0.0023	0.3840						0.0144
	(−0.2)	(−1.2)	(−0.1)	(8.6)						
IBEI	0.0534	−0.1065	−0.0448	0.2310						0.0144
	(3.2)	(−3.3)	(−0.8)	(5.5)						

Growth in each operating peformance variable is measured on a per share basis as of the sample formation date, with the number of shares outstanding adjusted to reflect stock splits and dividends. Values of *PASTGS5, RDSALES, EP, G,* and *PASTR6* are Winsorized at their 5th and 95th percentiles; *IBESLTG* is Winsorized at its 1st and 99th percentiles; and *DP* is Winsorized at its 98th percentile. Stocks with negative values of *BM* are excluded. In the regressions for *OIBD* or *IBEI,* firms with negative values of the operating performance variable in the base year are excluded, as are stocks with ratios of price to the operating performance variable above 100. The reported statistics are the averages over all years of the estimated coefficients, with *t*-statistics in parentheses, as well as the average R^2 of the model. In panels B to D, standard errors are based on the Hansen-Hodrick (1980) adjustment for serial correlation.

To see whether high past growth is a precursor to future growth, we use *PASTGS5*, the growth rate in sales over the five years prior to the sample selection date. Sales growth is correlated with earnings growth, but is much less erratic and so should yield a relatively more reliable verdict on whether past growth helps to predict future growth.[14]

Simple theoretical models of earnings growth suggest one set of variables that, in principle, should help to predict growth. For instance, a firm's earnings-to-price ratio, *EP*, is widely interpreted as impounding the market's expectations of future growth. We measure this as the firm's income before extraordinary items in the year prior to the sample selection date, relative to its price at the sample selection date. Similarly, in the standard constant-growth valuation model, a firm's sustainable growth rate is given by the product of its return on equity and its plowback ratio. Our proxy for this measure is *G*, where return on equity is measured as the firm's earnings before extraordinary items in the year prior to sample selection, divided by book equity in the preceding year; plowback is one minus the ratio in the prior year of dividends to income before extraordinary items.[15] Finally, to capture the firm's investment opportunities, we use the ratio of research and development expenditures to sales, *RDSALES*. The intensity of R&D relative to sales is widely used in practice as an indicator of how much resources a firm is investing in future growth opportunities (see, e.g., Chan et al. (2001)). When a firm has no R&D spending, we set this variable to zero, so all firms are eligible for the regression.

The forecast equation also incorporates variables that are popularly thought to connote high growth. Firms in technologically innovative industries, or more generally, growth stocks as measured by low book-to-market ratios, are popularly associated with high growth. High past returns for a stock may signal upward revisions in investors' expectations of future growth. Analysts' long-term forecasts are another proxy for the market's expectations of future growth. Finally, the dividend yield may provide information on the firm's investment opportunities and hence ability to grow future earnings. Correspondingly, the other forecasting variables are *TECH*, a dummy variable with a value of one for a stock in the pharmaceutical and technology sectors (defined as in Panel A of Table IV) and zero otherwise; *BM*, the firm's book-to-market value of equity; *PASTR6*, the stock's prior six-month compound rate of return; *IBESLTG*, the *IBES* consensus forecast of long-term growth; and *DP*, the ratio of dividends per share cumulated over the previous 12 months to current price. To be eligible for inclusion in the regression at a given horizon, a firm must have nonmissing values for all the predictors. In addition it must have a positive base-year value for the operating performance indicator in question, so as to calculate a growth rate. To screen out

[14] Results using past five-year growth in *OIBD* or *IBEI* as predictor variables indicate that these variables do a worse job in capturing any persistence in growth.

[15] Firms with negative value of book equity are dropped from the sample for the regression. In cases where the measure for sustainable growth is negative (when income is negative, or when dividends to common exceed income so the plowback ratio is negative), we set the sustainable growth rate variable *G* to zero.

outliers due to low values in the base year, we exclude cases where the ratio of the price to the operating performance variable exceeds 100 in the base year.

The model is estimated each year-end, yielding a time series of estimated coefficients and the adjusted R^2. Means for the time series, and t-statistics based on the standard error from the time series, are reported in Table X. Standard errors from the overlapping regressions in Panels B to D use the Hansen–Hodrick (1980) correction for serial correlation.

The results in Table X deliver a clear verdict on the amount of predictability in growth rates. In line with our earlier results, it is much easier to forecast growth in sales than growth in variables such as *OIBD* and *IBEI*, which focus more on the bottom line. For example, the forecasting model that has the highest adjusted R^2 in Table X is the equation for five-year growth in sales (11.75 percent; Panel C). By comparison, the adjusted R^2 in the equations for *OIBD* and *IBEI* barely exceed 3 percent, so there is relatively little predictability for growth in these variables. If anything, our results may be overstating the predictability in growth. Our cross-sectional regressions are reestimated monthly, so we let the coefficients in the model change over time. As a check on the robustness of our results, we also replicated the regressions in the table using growth rate ranks (ranging from zero for the firm with the lowest growth rate in that year to one for the firm with the highest growth rate). The results from the growth rank regressions echo the findings in Table X.

Our full model includes a total of nine predictors, and the correlations between some of them are quite high. As a result, sorting out the relative importance of each variable is not straightforward. Focusing on the models for *OIBD* and *IBEI*, no variable has coefficients that are statistically significant across all forecasting horizons. The coefficient of past sales growth *PASTGS5* is generally negative, suggesting that there are reversals in growth rates. When past sales have been declining, income levels tend to be low in the base year, resulting in relatively higher future growth rates.[16]

At least over longer horizons (Panels B to D), R&D intensity, *RDSALES*, has the strongest forecast power. In accordance with economic intuition, firms that are investing heavily in R&D, and thereby building up their intangible capital base, on average tend to be associated with elevated future growth. Specifically, a firm that spends 10 percent of its sales on R&D tends to have higher five-year growth in *IBEI* by about 2.5 percent, compared to a firm with no R&D (Panel C). However, the high correlation between *RDSALES* and variables like *TECH* or *DP* suggests caution is warranted in interpreting this result.

The variable *IBESLTG* is provided by supposed experts, and is widely used as a proxy for expected future growth. Its coefficient has the expected positive sign, but it is not statistically significant in the equations for *IBEI*. This variable does somewhat better in the equations for *OIBD*, especially over shorter horizons. In general, however, *IBESLTG* does not have higher forecast power than the divi-

[16] The effect of extremely low base-year values is mitigated to some extent because we drop from the regression cases where the ratio of the price to operating performance indicator exceeds 100 in the base year. However, this is only a partial solution.

dend yield, *DP*, which can be viewed as another proxy for the firm's investment opportunities.[17] In terms of predicting long-term growth, the forecasts of highly paid security analysts are about as helpful as the dividend yield, a piece of information that is readily available in the stock listings of most newspapers.

In line with the results in Table VIII, a low earnings yield *EP* is associated with higher future growth rates, especially for *IBEI*. However, the association is driven by a relatively small number of cases with unusually low base-year earnings. Low values of the earnings base result in a low earnings yield, and given that the firm survives, in an unusually high future growth rate. This explanation agrees with the results in Table VIII, where the relation between *EP* and future growth is confined to companies with the highest growth rates. As further confirmation of this line of reasoning, when we use growth in a variable such as *OIBD*, which is less prone to the problem of a low base level, *EP* does a poor job of forecasting in Table X.

The coefficient of the technology dummy *TECH* is highly significant in many cases, but it generally has an unexpected sign. This may be due to the high correlation between *TECH* and *RDSALES*. For example, dropping *RDSALES* from the model substantially reduces the t-statistics for *TECH* (although its coefficient retains a negative sign).

Neither the book-to-market ratio nor our proxy for sustainable growth G reliably predicts growth in *OIBD* and *IBEI*. Contrary to the conventional notion that high past returns signal high future growth, the coefficient of *PASTR6* is negative. The explanation for this result echoes our explanation for our findings with respect to *EP*. When a firm's near-term prospects sour and current earnings are poor, stock returns tend to be disappointing as well. Once again, these cases of low base levels of earnings may induce a negative association between past return and future growth.

Panels C and D also provide results that are based on a simple textbook model for predicting growth. Here the predictor variables are earnings yield, sustainable growth, and R&D intensity. The textbook model has weak forecast power. For example, over a five-year horizon, the adjusted R^2 from the equation for *IBEI* is only 1.48 percent.

VIII. Summary and Conclusions

We analyze historical long-term growth rates across a broad cross section of stocks using a variety of indicators of operating performance. All the indicators yield a median growth rate of about 10 percent per year (with dividends reinvested) over the 1951 to 1998 period. With dividends taken out, the median estimate is the same magnitude as the growth rate of gross domestic product over this period, between 3 and 3.5 percent in real terms. Given the survivorship bias underlying the growth rate calculations, the expected growth rate is likely to be lower. Based on these historical values and the low level of the current dividend

[17] Forecasting models with *IBESLTG* and DP as the only predictors yield qualitatively similar conclusions. In particular, the dividend yield does at least as well as the consensus forecast in forecasting growth.

yield, looking forward, the expected return on stocks in general does not appear to be high. In particular, the expected return using a constant-growth dividend valuation model is about 7.5 percent, assuming there is no mispricing.

Expectations about long-term growth are also crucial inputs in the valuation of individual stocks and for estimating firms' cost of capital. At year-end 1999, a sizeable portion of the market commanded price–earnings multiples in excess of 100. Justifying such a multiple under some relatively generous assumptions requires that earnings grow at a rate of about 29 percent per year for 10 years or more. Historically, some firms have achieved such dazzling growth. These instances are quite rare, however. Going by the historical record, only about 5 percent of surviving firms do better than a growth rate of 29 percent per year over 10 years. In the case of large firms, even fewer cases (less than 1 percent) would meet this cutoff. On this basis, historical patterns raise strong doubts about the sustainability of such valuations.

Nonetheless, market valuation ratios reflect a pervasive belief among market participants that firms who can consistently achieve high earnings growth over many years are identifiable ex ante. The long-term growth expectations of one influential segment of the market, security analysts, boldly distinguish between firms with strong and weak growth prospects. To see whether this belief that many firms can achieve persistently high growth holds up in reality, we use an experimental design that singles out cases where a firm consistently delivers favorable growth for several years in a row. Our results suggest that there is some persistence in sales revenue growth. The persistence in sales does not translate into persistence of earnings, however. Even though we measure consistency against a hurdle that is not particularly challenging (the median growth rate), there are few traces of persistence in growth of operating income before depreciation, or in income before extraordinary items. For example, on average three percent of the available firms manage to have streaks in growth above the median for five years in a row. This matches what is expected by chance. The evidence for persistence is still slim under more relaxed criteria for consistency in growth. All in all, the evidence suggests that the odds of an investor successfully uncovering the next stellar growth stock are about the same as correctly calling coin tosses.

A skeptic might argue that while there is little persistence for the population at large, specific segments of the market are able to improve earnings steadily over long periods. In particular, popular sentiment views firms in the pharmaceutical and technology sectors, along with glamour stocks, as being able to maintain consistently high growth rates. To accommodate this argument, we narrow our search to these subsets of firms. While there is persistence in sales growth, when it comes to growth in bottom-line income, over long horizons, the likelihood of achieving streaks is not much different from sheer luck. Conversely, value firms who are out of favor do not seem to do much worse, although survivorship bias makes it difficult to deliver a definitive verdict. To narrow the search even more, we check whether firms with consistently high past growth manage to maintain their performance going forward. While past growth carries over to future sales growth, the income variables do not display strong persistence.

There is a widespread belief that earnings-to-price ratios signal future growth rates. However, the cross-sectional relation between earnings yields and future growth is weak, except possibly in the cases of firms ranked highest by realized growth. For these firms, an inverse association between ex ante earnings yields and growth may arise because they start from a battered level of earnings in the base year, so future growth is high. In light of the noisiness of the earnings yield measure, academic and practitioner research mainly focuses on other valuation ratios such as book-to-market and sales-to-price. These multiples, which are better behaved, show little evidence of anticipating future growth. On the other hand, firms that enjoy a period of above-average growth are subsequently rewarded by investors with relatively high ratios of sales-to-price and book-to-market. Conversely, investors tend to penalize firms that have experienced poor growth. These results are consistent with the extrapolation hypothesis of La Porta (1996) and La Porta et al. (1997).

Additionally, it is commonly suggested that one group of informed participants, security analysts, may have some ability to predict growth. The dispersion in analysts' forecasts indicates their willingness to distinguish boldly between high- and low-growth prospects. IBES long-term growth estimates are associated with realized growth in the immediate short-term future. Over long horizons, however, there is little forecastability in earnings, and analysts' estimates tend to be overly optimistic. The spread in predicted growth between the top and bottom quintiles by IBES forecasts is 16.4 percent, but the dispersion in realized five-year growth rates is only 7.5 percent. On the basis of earnings growth for portfolios formed from stocks sorted by IBES forecasts, the spread in realized five-year growth rates is even smaller (3.3 percent). In any event, analysts' forecasts do not do much better than a naive model that predicts a one-for-one tradeoff between current dividend yield and future growth per share.

A regression forecasting model which brings to bear a battery of predictor variables confirms that there is some predictability in sales growth, but meager predictability in long-term growth of earnings. Only about three percent of the variation in five-year earnings growth rates is captured by the model. One variable that stands out is the level of research and development intensity, suggesting that a firm's intangible assets may have an important influence on its future performance. On the whole, the absence of predictability in growth fits in with the economic intuition that competitive pressures ultimately work to correct excessively high or excessively low profitability growth.

REFERENCES

Asness, Clifford, 2000, Bubble logic, Working paper, AQR Capital Management.

Bakshi, Gurdip, and Zhiwu Chen, 1998, Stock valuation in dynamic economies, Working paper, University of Maryland.

Ball, Ray, and Ross Watts, 1972, Some time-series properties of accounting income, *Journal of Finance* 27, 663–682.

Beaver, William H., 1970, The time series behavior of earnings, *Journal of Accounting Research* 8 (Suppl.), 62–99.

Beaver, William H., and Dale Morse, 1978, What determines price-earnings ratios?, *Financial Analysts Journal* 34, 65–76.

Brealey, Richard A., 1983, *An Introduction to Risk and Return from Common Stocks* (2nd edition) (MIT Press, Cambridge, MA).

Camerer, Colin, 1989, Does the basketball market believe in the "hot hand"? *American Economic Review* 79, 1257–1261.

Chan, Louis K. C., Narasimhan Jegadeesh, and Josef Lakonishok, 1995, Evaluating the performance of value versus glamour stocks: The impact of selection bias, *Journal of Financial Economics* 38, 269–296.

Chan, Louis K. C., Jason Karceski, and Josef Lakonishok, 2000, New paradigm or same old hype in equity investing? *Financial Analysts Journal* 56, 23–36.

Chan, Louis K. C., Josef Lakonishok, and Theodore Sougiannis, 2001, The stock market valuation of research and development expenditures, *Journal of Finance* 56, 2431–2456.

Claus, James, and Jacob Thomas, 2001, Equity premia as low as three percent? Evidence from analysts' earnings forecasts for domestic and international stocks, *Journal of Finance* 56, 1629–1666.

Daniel, Kent, and Sheridan Titman, 2001, Market reactions to tangible and intangible information, Working paper, Kellogg School of Management, Northwestern University.

Fama, Eugene F., and Kenneth R. French, 1992, The cross section of expected stock returns, *Journal of Finance* 47, 427–465.

Fama, Eugene F., and Kenneth R. French, 1995, Size and book-to-market factors in earnings and returns, *Journal of Finance* 50, 131–155.

Fama, Eugene, F., and Kenneth R. French, 1997, Industry costs of equity, *Journal of Financial Economics* 43, 153–193.

Fama, Eugene F., and Kenneth R. French, 2000, Forecasting profitability and earnings, *Journal of Business* 73, 161–175.

Fama, Eugene F., and Kenneth R. French, 2002, The equity premium, *Journal of Finance* 57, 637–659.

Fisher, Kenneth L., and Meir Statman, 2000, Cognitive biases in market forecasts, *Journal of Portfolio Management* 27, 72–81.

Gebhardt, William R., Charles M. C. Lee, and Bhaskaran Swaminathan, 2001, Toward an implied cost of capital, *Journal of Accounting Research* 39, 135–176.

Hansen, Lars, and Robert Hodrick, 1980, Forward exchange rates as optimal predictors of future spot rates: An econometric analysis, *Journal of Political Economy* 88, 829–853.

Hendricks, Darryll, Jayendu Patel, and Richard Zeckhauser, 1993, Hot hands in mutual funds: Short-run persistence of relative performance, 1974–1988, *Journal of Finance* 48, 93–130.

Kahneman, Daniel, and Mark W. Riepe, 1998, Aspects of investor psychology, *Journal of Portfolio Management* 24, 52–65.

La Porta, Rafael, 1996, Expectations and the cross section of stock returns, *Journal of Finance* 51, 1715–1742.

La Porta, Rafael, Josef Lakonishok, Andrei Shleifer, and Robert Vishny, 1997, Good news for value stocks: Further evidence on market efficiency, *Journal of Finance* 52, 859–874.

Lakonishok, Josef, Andrei Shleifer, and Robert W. Vishny, 1994, Contrarian investment, extrapolation, and risk, *Journal of Finance* 49, 1541–1578.

Lee, Charles M. C., James Myers, and Bhaskaran Swaminathan, 1999, What is the intrinsic value of the Dow? *Journal of Finance* 54, 1693–1741.

Lintner, John, and Robert Glauber, 1967, Higgledy piggledy growth in America, reprinted in James Lorie and Richard Brealey, eds., *Modern Developments in Investment Management* (Dryden Press, Hinsdale, IL).

Little, I. M. D., 1962, Higgledy piggledy growth, *Bulletin of the Oxford University Institute of Economics and Statistics* 4, 387–412.

Little, I. M. D., and Arthur C. Rayner, 1966, *Higgledy Piggledy Growth Again* (Basil Blackwell, Oxford).

Mezrich, Joseph, Qi Zeng, David Nordquist, and Lakshmi Seshadri, 2001, Rebroadcast: Changing of the guard, *Equity Research Quantitative Strategy* (Morgan Stanley, New York, NY).

Siegel, Jeremy J., 1999, The shrinking equity premium, *Journal of Portfolio Management* 26, 10–17.

Welch, Ivo, 2000, Views of financial economists on the equity premium and on professional controversies, *Journal of Business* 73, 501–537.

The Quarterly Review of Economics and Finance 58 (2015) 213–227

Contents lists available at ScienceDirect

The Quarterly Review of Economics and Finance

journal homepage: www.elsevier.com/locate/qref

Alternative errors-in-variables models and their applications in finance research

Hong-Yi Chen[a,*], Alice C. Lee[b,*], Cheng-Few Lee[c]

[a] National Chengchi University, Taiwan
[b] Center for PBBEF Research, USA
[c] Rutgers University, USA

ARTICLE INFO

Article history:
Received 25 February 2014
Received in revised form
29 September 2014
Accepted 2 December 2014
Available online 10 December 2014

JEL classification:
C58
G10

Keywords:
Measurement error
Errors-in-variables
Cost of capital
Capital structure
Investment equation
Capital asset pricing model

ABSTRACT

Specification error and measurement error are two major issues in finance research. The main purpose of this paper is (i) to review and extend existing errors-in-variables (EIV) estimation methods, including classical method, grouping method, instrumental variable method, mathematical programming method, maximum likelihood method, LISREL method, and the Bayesian approach; (ii) to investigate how EIV estimation methods have been used to finance related studies, such as cost of capital, capital structure, investment equation, and test capital asset pricing models; and (iii) to give a more detailed explanation of the methods used by Almeida et al. (2010).

1. Introduction

Specification error and measurement error are two major issues in applying the econometric model to economic and finance research. Studies by Miller and Modigliani (1966) and Roll (1969) are two of the earliest finance related research studies to apply errors-in-variables (EIV) model in their empirical works. Miller and Modigliani (1966) show that, in determining the cost of capital, anticipated average earnings are unobservable and using accounting estimates of earnings as the proxy may result measurement error problems. Roll (1969, 1977) and Lee and Jen (1978) show that the observed market rate returns in terms of stock market index are measured with errors since the stock market index does not include all assets which can be invested by investors. Roll (1969, 1977) argues that testing capital asset pricing model suffers from an EIV problem and concludes that no correct and unambiguous test of the theory can be accomplished. Lee and Jen (1978) have theoretically shown how beta estimates and Jensen performance measures can be affected by both constant and random measurement errors of the market rate of return and risk free rate. Other issues such as the determination of the capital structure and investment functions also suffer EIV problems.[1]

Understanding the existence of measurement error problems on finance related studies, a large extent of the literature subsequently tries to mitigate biased results from measurement errors. For the issue of the estimation of the cost of capital, Miller and Modigliani (1966) use the instrumental variable approach to resolve the measurement error problem and get consistent estimators in determining the cost of capital. Zellner (1970) and Lee and Wu (1989) also uses various estimation methods to deal with potential EIV problems on estimating the cost of capital. For the issue of the

* Corresponding author.
E-mail addresses: fnhchen@nccu.edu.tw (H.-Y. Chen), alice.finance@gmail.com (A.C. Lee), lee@business.rutgers.edu (C.-F. Lee).

[1] For the measurement problems related to the determinants of the capital structure, please see Titman and Wessels (1988), Chang et al. (2009), and Yang et al. (2009). For the measurement problems related to the investment function, please see Erickson and Whited (2000, 2002) and Almeida et al. (2010).

http://dx.doi.org/10.1016/j.qref.2014.12.002
1062-9769/

capital asset pricing test, Lee and Jen (1978) argue that both market return and beta coefficient are subjected to measurement error, and show how the beta coefficient can be estimated. Lee (1984) shows that the most generalized beta estimate can be decomposed into three components; bias due to specification error, bias due to measurement error, and interaction bias. Therefore, the evidence of failure in capital asset pricing model or the finding of new risk factors might result from model misspecification error or EIV problem. Gibbons and Ferson (1985), Green (1986), Roll and Ross (1994) and Diacogiannis and Feldman (2011) have argued that market portfolio measure with errors is an inefficient portfolio and show how the inefficient benchmark can affect the theoretical CAPM derivation. For the issue of the determinants of the capital structure, Titman and Wessels (1988), Chang, Lee, and Lee (2009) and Yang, Lee, Gu, and Lee (2009) apply structure equation models to investigate determinants of the capital structure. For the measurement error problems related to Tobin's q in investment function, Erickson and Whited (2000) use generalized method of moments (GMM) to obtain consistent estimators in testing q theory. Most recently, Almeida, Campello, and Galvao (2010) propose an alternative instrumental method to deal with measurement error problems in Tobin's q and support the q theory.

The main purpose of this paper is to study existing EIV estimation methods and to discuss how these estimation methods have been used in finance research. We first show how EIV problems affect estimators in the regression model. We further demonstrate seven alternative estimation methods dealing with EIV problems. Classical method, grouping method, instrumental variable method, mathematical programming method, maxima likelihood method, LISREL method, and Bayesian approach will be discussed. Finally, we conduct a survey on various studies and investigate the effect that resulted from EIV problems associated with cost of capital, capital asset pricing model, capital structure, and investment equation. We also investigate the correction models used in such studies to mitigate the problem raised from measurement errors.

The remainder of this paper is organized as follows. Section 2 shows the classical EIV problems and how they affect estimators of the linear regression model. Section 3 provides seven alternative correction methods in dealing with EIV problems. Section 4 presents the effects of EIV problems on the empirical research of cost of capital, asset pricing, capital structure, and investment decision. Finally, Section 5 presents the conclusion.

2. Effects of errors-in-variables in different cases

2.1. Bivariate normal case

Suppose we have a two variate structural relationship

$$V_i = \alpha + \beta U_i \tag{1}$$

Both V_i and U_i are unobserved, while we can observe $Y_i = V_i + \eta_i$ and $X_i = U_i + \varepsilon_i$. We assume that

(a) $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$ and $\eta_i \sim N(0, \sigma_\eta^2)$.
(b) $E(\varepsilon_i U_i) = 0$, $E(\varepsilon_i V_i) = 0$, $E(\varepsilon_i \eta_i) = 0$, $E(\eta_i U_i) = 0$, and $E(\eta_i V_i) = 0$.
(c) $U_i \sim N(E(X), \sigma_U^2)$ and $V_i \sim N(\alpha + \beta E(X), \beta^2 \sigma_U^2)$.

This results in measurement error on the estimates of α and β implying that the asymptotic biases for β and α are

$$\text{plim}\hat{\beta} - \beta = \frac{-\beta\sigma_1^2}{\sigma_U^2 + \sigma_1^2} \tag{2a}$$

$$\text{plim}\hat{\alpha} - \alpha = \frac{\beta\sigma_1^2}{\sigma_U^2 + \sigma_1^2} E(X). \tag{2b}$$

Eq. (2) implies that $\hat{\beta}$ is downward biased and $\hat{\alpha}$ is upward biased.

2.2. Multivariate case

Suppose we have a trivariate structural relationship

$$W_i = \alpha + \beta U_i + \gamma V_i \tag{3}$$

W_i, U_i, and V_i are unobserved, but we can observe $Z_i = W_i + \tau_i$, $X_i = U_i + \varepsilon_i$, and $Y_i = V_i + \eta_i$. U_i and V_i have a joint normal distribution with variances σ_U^2 and σ_V^2 and correlation coefficient ρ_{UV}. In the observed variables X, Y, and Z, the observed errors ε, η, and τ are independent normal variables with zero means and variance $\sigma_\varepsilon^2, \sigma_\eta^2, \sigma_\tau^2$. X, Y, and Z have a multivariate normal distribution.

The asymptotic biases of $\hat{\beta}$ and $\hat{\gamma}$ can be seen from the following:

$$\text{p}\lim \hat{\beta} - \beta = \frac{\sigma_{VW}\sigma_\eta^2 - \beta(\sigma_U^2\sigma_\eta^2 + \sigma_V^2\sigma_\varepsilon^2 + \sigma_\varepsilon^2\sigma_\eta^2)}{(\sigma_U^2\sigma_V^2 - \sigma_{VW}^2) + \sigma_U^2\sigma_\eta^2 + \sigma_V^2\sigma_\varepsilon^2 + \sigma_\varepsilon^2\sigma_\eta^2} \tag{4a}$$

$$\text{p}\lim \hat{\gamma} - \gamma = \frac{\sigma_{WV}\sigma_\varepsilon^2 - \gamma(\sigma_U^2\sigma_\eta^2 + \sigma_V^2\sigma_\varepsilon^2 + \sigma_\varepsilon^2\sigma_\eta^2)}{(\sigma_U^2\sigma_V^2 - \sigma_{UV}^2) + \sigma_U^2\sigma_\eta^2 + \sigma_V^2\sigma_\varepsilon^2 + \sigma_\varepsilon^2\sigma_\eta^2} \tag{4b}$$

The direction of the biases of $\hat{\beta}$ and $\hat{\gamma}$ can be treated according to different assumptions.[2]

Concerning the coefficient of the reliability, Cochran (1970) shows that measurement errors of both explained and explanatory variables will reduce the multiple correlations and increase the residual variance, and the good prediction formula is more sensitive to measurement errors than the poor one. Moreover, from the analysis of the effects of measurement error on both the simple regression coefficient and residual variance, in general, we can conclude that the t statistic of the simple regression coefficient will be downward biased if variables are measured with errors.

3. Estimation methods when variables are subject to error

In this section, we will discuss alternative EIV estimation methods, classical method, grouping method, instrumental variable method, mathematical method, maxima likelihood method, LISREL method, and the Bayesian approach.

3.1. Classical estimation method

3.1.1. The classical method to a simple regression analysis

In general, the classical method considers three cases: (i) either σ_ε^2 or σ_η^2 is known; (ii) $\lambda = \sigma_\eta^2/\sigma_\varepsilon^2$ is known; and (iii) σ_ε^2 and σ_η^2 are known. We can obtain the estimate for β from Eq. (2) under every possible situation as:

$$\text{(i)}\ \hat{\beta} = \frac{S_{XY}}{S_{XX} - \sigma_\varepsilon^2}, \quad \text{when}\ \sigma_\varepsilon^2\ \text{is known.} \tag{5}$$

$$\hat{\beta} = \frac{S_{YY} - \sigma_\eta^2}{S_{XY}}, \quad \text{when}\ \sigma_\eta^2\ \text{is known.} \tag{6}$$

$$\text{(ii)}\ \hat{\beta} = \frac{(S_{YY} - \lambda S_{XX}) + \{(S_{YY} - \lambda S_{XX})^2 + 4\lambda S_{XY}\}^{1/2}}{2S_{XY}},$$

$$\text{when}\ \lambda = \frac{\sigma_\eta^2}{\sigma_\varepsilon^2}\ \text{is known.} \tag{7}$$

[2] Please see Lee (1973) and Chen (2011) for detail.

(iii) When both σ_ε^2 and σ_η^2 are known, Kendall and Stuart (1961) regarded it as an over-identified situation unless a non-zero covariance between U_i and V_i is introduced. Barnett (1967) followed Kiefer's (1964) suggestion and derived a consistent estimator of $\hat{\beta}$ as one of the real roots of Eq. (19).

$$\hat{\beta}^4 - \left(\frac{1}{b_2} - \frac{\lambda}{b_1} - 2\lambda b_2\right)\hat{\beta}^3 - 3\lambda\left(1 - \lambda\frac{b_2}{b_1}\right)\hat{\beta}^2 + \lambda^2\left(\lambda\frac{b_2}{b_1^2} - \frac{1}{b_1} - 2b_2\right)\hat{\beta} - \lambda^3\frac{b_2}{b_1} = 0, \tag{8}$$

where $S_{XX} = (\sum_{i=1}^n (X_i - \bar{X})^2)/n$, $S_{YY} = (\sum_{i=1}^n (Y_i - \bar{Y})^2)/n$, $S_{XY} = (\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}))/n$, $m_{XX} = Var(X) = \sigma_U^2 + \sigma_\varepsilon^2$, $m_{YY} = Var(Y) = \beta^2\sigma_U^2 + \sigma_\eta^2$, $m_{XY} = Var(Y) = \beta\sigma_U^2$, $b_1 = m_{XY}/m_{XX}$, $b_2 = m_{XY}/m_{YY}$, and $\lambda = \sigma_\eta^2/\sigma_\varepsilon^2$.

The advantage of the knowledge of both σ_ε^2 and σ_η^2 is that one obtains a more efficient estimator of $\hat{\sigma}_\eta^2$.

The analysis of Eqs. (5)–(8) furnish us two important implications. First, if only U or V is subject to measurement error, then we know that the maxima likelihood estimator of β is equivalent to fitting a least square line when using the error-free variable as a regressor. Second, if both U and V are measured with errors, then the estimate of β lies between the values estimated from Eqs. (5) and (6). This situation can be further analyzed. The quadratic equation for Eq. (7) is

$$\hat{\beta}^2 S_{XY} + \hat{\beta}(\lambda S_{XX} - S_{YY}) - \lambda S_{XY} = 0. \tag{9}$$

(a) When either $\sigma_\eta^2 \to 0$ or $\sigma_\varepsilon^2 \to \infty$, Eq. (6) shows $\hat{\beta} = \frac{S_{YY}}{S_{XY}}$.

(b) When $\sigma_\varepsilon^2 \to 0$ or $\sigma_\eta^2 \to \infty$, Eq. (5) shows $\hat{\beta} = \frac{S_{XY}}{S_{XX}}$.

3.1.2. The classical method to a multiple regression analysis

It is clear that U is orthogonal to V if $\rho = 0$. It is well-known that this multiple regression reduces to two simple regression relationships. If $\rho \neq 0$, to identify β and γ, we need to know either the actual values of σ_ε^2, σ_η^2, and σ_τ^2 or the relative ratios among σ_ε^2, σ_η^2, and σ_τ^2. We will investigate the following cases:

(i) $\sigma_\eta^2 = 0$, $\sigma_\varepsilon^2 > 0$, $\sigma_\tau^2 < 0$, and $\sigma_\varepsilon^2 = \lambda\sigma_\tau^2$
(ii) $\sigma_\varepsilon^2 > 0$, $\sigma_\eta^2 > 0$, and $\sigma_\tau^2 > 0$

(a) σ_ε^2 and σ_η^2 are known
(b) $\sigma_\eta^2 = \lambda_1\sigma_\varepsilon^2$ and $\sigma_\tau^2 = \lambda_2\sigma_\varepsilon^2$

(**Case i**) Only Z and X are measured with errors.
The estimator of $\hat{\beta}$ can be one of real root of Eq. (10)

$$k_2\hat{\beta}^2 + k_1\hat{\beta} + k_0 = 0, \tag{10}$$

where

$$k_0 = \left(S_{XZ} - \frac{S_{XY}S_{YZ}}{S_{YY}}\right), \quad k_1 = -\left(S_{XX} - \frac{S_{XY}^2}{S_{YY}}\right) + \lambda\left(S_{ZZ} - \frac{S_{YZ}^2}{S_{YY}}\right),$$

and $k_2 = -\lambda k_0 = -\lambda\left(S_{XZ} - \frac{S_{XY}S_{YZ}}{S_{YY}}\right)$.

There are three cases to consider:

(a) $\lambda \to 0$ when $\sigma_\varepsilon^2 \to 0$, then in this case from Eq. (10), we know that

$$\hat{\beta} = \frac{-S_{XY}S_{YZ} + S_{XZ}S_{YY}}{S_{XX}S_{YY} - S_{XY}^2}. \tag{11}$$

(b) $\lambda \to \infty$ when $\sigma_\tau^2 \to 0$, in this case from Eq. (11), we know that

$$\hat{\beta} = \frac{-S_{ZZ}S_{YY} - S_{YZ}^2}{S_{XZ}S_{YY} + S_{XY}S_{YZ}}. \tag{12}$$

(c) When both $\sigma_\varepsilon^2 > 0$ and $\sigma_\tau^2 > 0$, then

$$\hat{\beta} = \frac{k_1 \pm \sqrt{k_1^2 + 4\lambda k_0^2}}{2\lambda k_0}. \tag{13}$$

When β is determined, γ can also be estimated. After both β and γ are estimated, then α can be estimated by

$$\hat{\alpha} = \bar{Z} - \hat{\beta}\bar{X} - \hat{\gamma}\bar{Y}. \tag{14}$$

(**Case ii**) When Z, X, and Y are all observed with errors

(a) Both σ_ε^2 and σ_η^2 are known.
We can obtain the two normal equations as follows:

$$S_{XZ} = \beta(S_{XX} - \hat{\sigma}_1^2) + \gamma S_{XY} \tag{15}$$

$$S_{YZ} = \beta S_{XZ} + \gamma(S_{YY} - \hat{\sigma}_\eta^2) \tag{16}$$

Solving Eqs. (15) and (16) by Cramer′s rule we have

$$\hat{\beta} = \frac{S_{XZ}S_{YY} - S_{XY}S_{YZ} - S_{XZ}\hat{\sigma}_\eta^2}{S_{XX}S_{YY} - \hat{\sigma}_\varepsilon^2 S_{YY} - \hat{\sigma}_\eta^2 S_{XX} + \hat{\sigma}_\varepsilon^2\hat{\sigma}_\eta^2 - (S_{XY})^2}$$
$$\hat{\gamma} = \frac{S_{YZ}S_{XX} - S_{XZ}S_{YY} - S_{YZ}\hat{\sigma}_\varepsilon^2}{S_{XX}S_{YY} - \hat{\sigma}_\varepsilon^2 S_{YY} - \hat{\sigma}_\eta^2 S_{XX} + \hat{\sigma}_\varepsilon^2\hat{\sigma}_\eta^2 - (S_{XY})^2}. \tag{17}$$

(b) Both $\hat{\sigma}_\eta^2 = \lambda_1\hat{\sigma}_\varepsilon^2$ and $\hat{\sigma}_\tau^2 = \lambda_2\hat{\sigma}_\varepsilon^2$ are known.

We can obtain β estimator as one of real roots of the following cubic equation

$$\hat{\beta}^3 H_3 + \hat{\beta}^2 H_2 + \hat{\beta}H_1 + H_0 = 0, \tag{18}$$

where

$$H_3 = S_{XY}S_{YZ}^2 - MS_{XZ}S_{YZ} - \lambda_1 S_{XY}S_{XZ}^2$$
$$H_2 = -S_{YZ}^2 + 2\lambda_2 S_{XY}^2 S_{YZ} + MTS_{YZ} - M\lambda_2 S_{XY}S_{YZ} - \lambda_1 S_{XZ}^2 S_{YZ} + 2\lambda_1 TS_{XZ}S_{XY}$$
$$H_1 = \lambda_2^2 S_{XY}^3 - 2\lambda_2 S_{XY}S_{YZ}^2 + MT\lambda_2 S_{XY} + \lambda_2 S_{XY}S_{YZ} + \lambda_2 MS_{YZ}S_{XZ} - \lambda_1\lambda_2 S_{XZ}^2 S_{XY} + 2\lambda_1 S_{XY}S_{XZ}^2 - \lambda_1 T^2 S_{XY}$$
$$H_0 = \lambda_2^2 S_{YZ}S_{XY}^2 + \lambda_2^2 S_{XZ}S_{XY} + \lambda_1\lambda_2 S_{XZ}^2 S_{YZ} + T\lambda_1\lambda_2 S_{XZ}S_{YZ} - 2\lambda_1\lambda_2 TS_{XZ}S_{XY}$$

When $\lambda_2 = 0$, Eq. (18) will reduce to a quadratic equation in $\hat{\beta}$.

3.1.3. The constrained classical method

Under the classical case (Case ii), if we only know $\hat{\sigma}_\eta = \hat{\sigma}_\varepsilon\lambda$, then we can identify β and γ by imposing $\beta + \gamma = 1$.

We can obtain a quadratic equation in $\hat{\beta}$

$$\hat{\beta}^2(1-\lambda)S_{XY} + \hat{\beta}(S_{YZ} - \lambda S_{XZ} + 2\lambda S_{XY} + S_{YY} - \lambda S_{XX}) + \lambda(S_{XZ} - S_{XY}) = 0. \tag{19}$$

When $\lambda = 0$, Eq. (19) will reduce to

$$\hat{\beta} = -\frac{S_{YZ} + S_{YY}}{S_{XY}}. \tag{20}$$

Imposing $\beta + \gamma = 1$, upon a multiple regression will help to identify the regression coefficients, but it should also be realized that the constrained regression technique will bias the estimates of the regression coefficients if the unrestricted estimator fails to satisfy the restriction $\beta + \gamma = 1$. The advantages and disadvantages of the constrained regression technique have been discussed by Theil (1971) in some detail.

3.2. Grouping method

Following the structural relationship described in Eq. (1)

$V_i = \alpha + \beta U_i.$

Both V_i and U_i are unobserved, and only $Y_i = V_i + \eta_i$ and $X_i = U_i + \varepsilon_i$ can be observed. There exists EIV bias when using Y_i and X_i, to investigate the relationship between V_i and U_i.

Wald (1940) proposes a two-portfolio grouping method in dealing with the EIV problem when both dependent and independent variables are subject to measurement errors. He suggests that the measurement error can be reduced by grouping observations into portfolios. In Wald's two-portfolio grouping method, he groups the independent variable either in descending or ascending order, and divides the observations into two equal groups for both dependent and independent variables; therefore, the first-step estimator of the market model, estimated beta risk, can be written as:

$$\hat{\beta} = \frac{(\bar{Y}_1 - \bar{Y}_2)}{(\bar{X}_1 - \bar{X}_2)}, \tag{21}$$

where $\bar{X}_1$ and $\bar{X}_2$ are the arithmetic means of independent variables for the first and the second groups, respectively; and $\bar{Y}_1$ and $\bar{Y}_2$ are the arithmetic means of independent variables for the first and the second groups, respectively.

Grouping method is widely used in finance related research. For example, to minimize the EIV problem in testing the asset pricing model, Black, Jensen, and Scholes (1972), Blume and Friend (1973), Fama and MacBeth (1973) and Litzenberger and Ramaswamy (1979) use two-pass procedure and the k-portfolio grouping method to examine the capital asset pricing model. By combining securities into portfolios, most of the firm-specific component of the returns can be diversified away and the precision of the beta estimates will be enhanced. The grouping method can, therefore, mitigate the problem raised from measurement errors in estimated beta.

However, some limitations affect the grouping method. First, the grouping method shrinks the range of estimators in the first step and reduces statistical power. To mitigate this problem, in two-pass procedure, the grouping method suggests sorting securities on the first-pass estimator first. Then portfolios are formed by grouping securities with same level of first-pass estimators. This sorting procedure is now standard in empirical tests. Second, a trade-off exists between the bias and the variance of the first-pass estimator according to the number of portfolios. Shanken (1992) argues that the grouping method may cause a larger variation in the portfolio beta. As the number of portfolios (N) increases, the magnitude of the bias becomes greater while the variance of the estimator becomes smaller, and vice versa; therefore, an optimal number of portfolios might exist in which a minimum mean squared error can be obtained. More specifically, when risk premium is estimated by the time-series mean of the cross-sectional regression estimates in testing capital asset pricing model, the mean squared error of the risk premium estimate would be dominated by its bias because its variance would monotonically decrease as the testing period becomes longer. Third, the formation of portfolios for the second-pass estimation might cause a loss of valuable information about cross-sectional behavior among individual securities, because the cross-sectional variations would be smoothed out. Fourth, Ahn, Conrad, and Dittmar (2009) argue that the grouping method, although mitigating measurement error, may yield different results by using different portfolio grouping methods.

Although the grouping method suffers from the limitations discussed above, it still has some clear advantages. With the cross-sectional regression in the second pass, interpreting the results in economic terms is straightforward. Examining model misspecification by testing whether firm characteristics, such as firm size and book-to-market ratio, can explain returns across firms is also convenient. Moreover, the grouping method is intuitive and easy to implement with real data. The grouping method is therefore still preferred in many empirical studies.

3.3. Instrumental variable method

Durbin (1954) proposes an instrumental variable method to deal with the EIV problem in a regression model. In the instrumental variable method, the instrumental variable, T_i, is an observable variable known to correlate with V_i and U_i, but is independent of η_i and ε_i. Then β can be estimated by

$$\hat{\beta} = \frac{\sum_{i=1}^{n}(T_i - \bar{T})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(T_i - \bar{T})(X_i - \bar{X})} = \frac{\sum_{i=1}^{n}(T_i - \bar{T})(U_i - \bar{U}) + \sum_{i=1}^{n}(T_i - \bar{T})(\eta_i - \bar{\eta})}{\sum_{i=1}^{n}(T_i - \bar{T})(V_i - \bar{V}) + \sum_{i=1}^{n}(T_i - \bar{T})(\varepsilon_i - \bar{\varepsilon})} \tag{22}$$

If $p\lim\sum_{i=1}^{n}(T_i - \bar{T})(U_i - \bar{U})$ exists, then $\hat{\beta}\hat{b}$ is a consistent estimator of β because both ε_i and η_i are independent of T_i. Eq. (22) can be written in matrix form as follows:

$$\begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix} = (\mathbf{T}'\mathbf{X})^{-1}\mathbf{T}'\mathbf{Y}, \tag{23}$$

where $\mathbf{T}' = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ T_1 & T_2 & T_3 & \dots & T_n \end{bmatrix}$, $\mathbf{X}' = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ X_1 & X_2 & X_3 & \dots & X_n \end{bmatrix}$, and $\mathbf{Y}' = \begin{bmatrix} Y_1 & Y_2 & Y_3 & \dots & Y_n \end{bmatrix}$.

However, finding an instrumental variable uncorrelated with η_i and ε_i while highly correlated with V_i and U_i is difficult. Durbin (1954) suggests that if the order of U_i is the same as the order of X_i, then a better instrumental variable would be $T_i = i$, where X_i are ordered by magnitude. That is, $T' = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & 2 & 3 & \dots & n \end{bmatrix}$ and $X' = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ X_1 & X_2 & X_3 & \dots & X_n \end{bmatrix}$. This variable will lead to a more efficient estimate than that of the method of grouping. If we let $T_i = 1$ for X_i greater than its median, $T_i = 0$ for X_i equal to its median, and $T_i = -1$ for X_i smaller than its median, then the estimator of the instrumental variable method will be the same as the estimator of Wald's two-group grouping method. Therefore, Wald's two-group grouping method is a special case of the instrumental variable method. In other words, the instrumental variable method is more generalized than the grouping method.

Griliches and Hausman (1986) propose an instrumental variable approach to reduce the bias resulted from measurement error. In a penal data framework, they show that instrumental variables estimator is consistent if the measurement error $\varepsilon_{i,t}$ is *i.i.d* across i and t and unobserved independent variable is serially correlated. An instrumental variable using the lags or the difference of lags of the unobserved independent variable can result in a consistent estimator when T is finite and N approaches to infinite.

However, Griliches and Hausman's *i.i.d* assumption is too strong. Biorn (2000) further relaxes Griliches and Hausman's *i.i.d* assumption, and, instead, assumes that the regressor has τ period moving average process. Biorn (2000) shows that using the lags of the variables at least $\tau - 2$ periods as instruments can clear the memory of the moving average process and obtain the consistent estimator.

Lewbel (1997) show simple functions of the model data can be used as instruments for two staged least squares (TSLS) estimation. Such instruments can be used for identification and estimation when no other instruments are available or improve efficiency.

Given the standard linear regression model with measurement error:

$$Y_i = a + b'W_i + cX_i + e_i, \quad \text{and} \tag{24}$$

$$Z_i = d + X_i + v_i, \tag{25}$$

in which Y_i, W_i, and Z_i are observable for $i = 1, \ldots, n$, while X_i, e_i, and v_i are unobservable. Eqs. (34) and (35) imply that

$$Y_i = \alpha + b'W_i + cZ_i + \varepsilon_i. \tag{26}$$

However, since both Z_i and ε_i depend on v_i, estimators of b and c from OLS regression is inconsistent. Lewbel (1997) shows that the consistent estimators can be obtained by using TSLS with instruments 1, W_i, and q_i, where q_i is some vector of instruments that are correlated with X_i but not correlated with e_i and v_i.

Lewbel (1997) further empirically applies the instrumental variable method to testing elasticity of patent applications with respect to research and development (R&D) expenditures. He finds, using the TSLS instrumental variable model, the estimated elasticity yields very close to one. Therefore, the TSLS instrumental variable model can mitigate the effects of measurement error and confirm the relationship between patent and R&D.

In addition, Erickson and Whited (2000, 2002) propose a two-step generalized method of moments (GMM) estimators that exploit over-identifying information contained in the high-order moments of residuals obtained from perfectly measured regressors. Basing GMM estimation on residual moments of more than second order requires that the GMM covariance matrix be explicitly adjusted to account for the fact that estimated residuals are used instead of true residuals defined by population regressions. Erickson and Whited (2000) show that estimators obtained by using moments up to seventh order perform well in Monte Carlo simulations.

Almeida et al. (2010) use Monte Carlo simulations and empirically test investment models to compare the performance of the instrumental variables approach suggested by Biorn (2000) and generalized method of moments. They find that the instrumental variable method can obtain more consistent and efficient estimators than generalized method of moments when independent variables subject to measurement error.

However, it is difficult to obtain appropriate instrument variables, resulting in weak evidence in empirical research. Lewbel (2012) proposes a new method to deal with measurement error problems in regression model when instrumental variables are not available. Under the assumption of heteroscedastic errors, Lewbel (2012) shows that the regression model with measurement regressors can be identified and estimated by TSLS or GMM.

3.4. Mathematical method

3.4.1. Bivariate case

Deming (1943), York (1966) and Clutton-Brock (1967) have developed a weighted-regression-method-under-iteration approach. Deming (1943) proposed that the best straight line of Eq. (1) can be obtained by minimizing the sum in the following equation:

$$S = \sum_i \{w(X_i)(\hat{U}_i - X_i)^2 + w(Y_i)(\hat{V}_i - Y_i)^2\} \tag{27}$$

$\hat{U}_i$ and $\hat{V}_i$ are the adjusted value of X_i and Y_i which make the sum in Eq. (27) a minimum. Since we require $\hat{U}_i$ and $\hat{V}_i$ to lie on the best straight line, we must have

$$\hat{V}_i = \alpha + \beta \hat{U}_i, \quad (i = 1, \ldots, n) \tag{28}$$

Both $w(X_i)$ and $w(Y_i)$ are the weights of various observations. They are reciprocally proportional to the variance of their measurement error, respectively.

If these values of $\hat{U}_i$, $\hat{V}_i$, α, and β make S a minimum, we have

$$\beta^3 \sum_i \frac{k_i^2 x_i^2}{w(X_i)} - 2\beta^2 \sum_i \frac{k_i^2 x_i y_i}{w(X_i)} - \beta \left\{ \sum_i k_i x_i^2 - \sum_i \frac{k_i^2 y_i^2}{w(X_i)} \right\} + \sum_i k_i x_i y_i = 0, \tag{29}$$

where $x_i = X_i - \bar{X}$, $y_i = Y_i - \bar{Y}$, $\bar{X} = \frac{\sum_i k_i X_i}{\sum_i k_i}$, $\bar{Y} = \frac{\sum_i k_i Y_i}{\sum_i k_i}$, and $k_i = \frac{w(X_i)w(Y_i)}{\beta^2 w(Y_i) + w(X_i)}$.

Eq. (29) is the least-square cubic derived by York (1966). To solve Eq. (29), an initial value is assigned to β to estimate k_i. After obtaining the roots of Eq. (29), one of the legitimate solutions is assigned to estimate k_i and obtain new solutions for β again. A similar procedure is employed iteratively until a convergent solution is obtained.

The mathematical approach involves the estimation of the parameters of a function conditional on the maximum likelihood function adjusted for the true values. This method is different from the classical method in three ways. First, variances of measurement errors for every observation are different. Second, a weighted regression method is applied. Third, the iteration procedure is used to obtain a consistent estimator.

It can be proved that the mathematical programming method reduces to the classical method under three certain conditions.

(i) Only Y_i has an EIV problem

We can put more weight on X_i which has no EIV problem, $w(X_i) = \infty$, $k_i = w(Y_i)$. We can therefore solve the least square cubic

$$\beta = \frac{\sum_i w(Y_i) x_i y_i}{\sum_i w(Y_i) x_i^2}, \tag{30}$$

which is the estimated coefficient of weighted regression of Y_i on X_i.

(ii) Only X_i has an EIV problem

In this case, we put more weight on Y_i which has no EIV problem, then $w(Y_i) = \infty$, $k_i = \frac{w(X_i)}{\beta^2}$. We can solve the least square cubic

$$\beta = \frac{\sum_i w(X_i) y_i^2}{\sum_i w(X_i) x_i y_i}, \tag{31}$$

which is the inverse estimated coefficient of weighted regression of Y_i on X_i.

(iii) Both X_i and Y_i have EIV problem, and $w(X_i)/w(Y_i) = c$.

The least square cubic becomes

$$\beta^2 + \beta \frac{\{c\sum_i k_i x_i^2 - \sum_i k_i y_i^2\}}{\sum_i k_i x_i y_i} - c = 0 \tag{32}$$

3.4.2. Multivariate case

Lee (1973) extends the bivariate mathematical programming method, which was developed by Deming (1943), York (1966) and Clutton-Brock (1967), to a trivariate case. We define $w(Z_i)$, $w(Z_i)$, and $w(Y_i)$ which are the weights of the various observations of Z_i, X_i, and Y_i. It is assumed W, U, and V are functionally rather than structurally related. The mathematical programming procedure begins by minimizing[3]

$$\begin{aligned} S &= \textstyle\sum_i \{w(X_i)(x_i - X_i)^2 + w(Y_i)(y_i - Y_i)^2 + w(Z_i)(z_i - Z_i)^2\} \\ \text{s.t. } z_i &= \alpha + \beta x_i + \gamma y_i. \end{aligned} \tag{33}$$

This extension will reduce to Deming's (1943) weighted regression results when the quadratic term of equations are omitted, while Lee's (1973) result is more general than Deming's weighted multiple regression analysis.

3.5. Maximum likelihood method

In testing capital asset pricing model with dividend and tax, Litzenberger and Ramaswamy (1979) use maximum likelihood method to reduce the effect of errors-in-variables. Litzenberger and Ramaswamy (1979) show that, assuming that the variance of the measurement error in beta is known, the cross sectional variance of true betas can be replaced by the difference in the variation of the observed betas and the variance of the measurement error. Then the estimator in capital asset pricing model test, under such condition, is consistent by maximum likelihood method.

Kim (1995, 1997, 2010) further provides a maximum likelihood method to correct the EIV problem in testing the asset pricing model. Based upon two-pass capital asset pricing model, Kim (1995) shows that in a multifactor asset pricing model test the EIV leads to an underestimation of the independent variable with a measurement error and an overestimation of the independent variable without measurement error. To correct EIV biases, Kim (1995) extracts additional information about the relation between idiosyncratic error variance which can be obtained from the first step and the measurement error variance, and incorporates such additional information into the second step of the capital asset pricing model test. Assuming the homoscedasticity of the disturbance term of the market model, Kim (1995) shows that the corrected factors for the traditional least squares estimators of the cross-sectional regression coefficients can be obtained by the maximum likelihood method. The closed form estimators of the multifactor asset pricing model test can therefore be obtained. Assuming the first and second steps of the multifactor asset pricing model are

$$R_{i,t} = \alpha_i + \beta_i R_{m,t} + e_{i,t}, \tag{34}$$

and

$$R_{i,t} = \gamma_{0,t} + \gamma_{1,t}\hat{\beta}_{i,t-1} + \gamma_{2,t} V_{i,t-1} + e_{i,t}, \tag{35}$$

where $\beta_{i,t-1}$ is the market risk factor with measurement error, and $V_{i,t-1}$ is a risk factor with no measurement error for security i at time $t-1$. The adjusted estimators in the second step can be written as follows:

$$\begin{aligned} \hat{\gamma}_{1t} &= \frac{M + \left[M^2 + 4\delta_t m_{R\hat{\beta}}^2 (1 - (\hat{\rho}_{RV}\hat{\rho}_{\hat{\beta}V}/\hat{\rho}_{R\hat{\beta}}))^2\right]^{1/2}}{2m_{R\hat{\beta}}(1 - (\hat{\rho}_{RV}\hat{\rho}_{\hat{\beta}V}/\hat{\rho}_{R\hat{\beta}}))} \\ \hat{\gamma}_{2t} &= (m_{RV} - \hat{\gamma}_{1t} m_{\hat{\beta}V})/m_{VV} \\ \hat{\gamma}_{0t} &= \tilde{R}_t - \hat{\gamma}_{1t}\tilde{\hat{\beta}}_{t-1} - \hat{\gamma}_{2t}\tilde{V}_{t-1} \end{aligned} \tag{36}$$

where $M = m_{RR}(1 - \hat{\rho}_{RV}^2) - \delta_t m_{\hat{\beta}\hat{\beta}}(1 - \hat{\rho}_{\hat{\beta}V}^2)$, $m_{xy} = (1/N)\sum_{i=1}^N \sum_{j=1}^N w_{ij}(x_i - \bar{x})(y_i - \bar{y})/\sum_{i=1}^N \sum_{j=1}^N w_{ij}$, $\bar{x} = \sum_{i=1}^N \sum_{j=1}^N w_{ij} x_i / \sum_{i=1}^N \sum_{j=1}^N w_{ij}$, $\bar{y} = \sum_{i=1}^N \sum_{j=1}^N w_{ij}\bar{x}_i / \sum_{i=1}^N \sum_{j=1}^N w_{ij}$, w_{ij} is the (i, j) element of inverse matrix of residual variance in the first-step, $\hat{\Sigma}_\varepsilon^{-1}$, and $\hat{\rho}_{xy}^2 = m_{xy}/(m_{xx}m_{yy})^{1/2}$.

As a result, the maxima likelihood method can correct the problem on exaggerating the estimated coefficient associated to the variable without measurement error. Moreover, the absolute value of estimated intercept by maxima likelihood method is generally smaller than the absolute value of estimated intercept by traditional least squares.

3.6. LISREL and MIMIC methods

Goldberger (1972) conceptually described the LISREL model as a combination of factor analysis and econometrics model. In addition, Anderson (1963) has shown that factor analysis is a generalized version of errors-in-variables (EIV) methods. In this section, we will review and discuss how LISREL and MIMIC methods can be used to deal with EIV in finance research.

The linear simultaneous equation system is widely used in finance and accounting related research. However, a serious limitation of the simultaneous equation approach is an EIV problem. For example, the theoretical determinants of capital structure in corporate finance can be attributed to unobservable constructs that are usually measured in empirical studies by a variety of observable indicators or proxies. These observable indicators or proxies can then be viewed as measures of latent variables with measurement errors. Maddala and Nimalendran (1996) show that the use of these indicators as theoretical explanatory variables may cause EIV problems. Bentler (1983) also emphasizes the estimated results of the traditional simultaneous equation model has no meaning when variables have measurement errors. Therefore, the latent variable covariance structure model is provided and applied in corporate finance. Titman and Wessels (1988), Chang et al. (2009) and Yang et al. (2009), mitigate the measurement problems of proxy variables, and apply structure equation models (e.g. LISREL model and MIMIC model) to determine capital structure decision. Maddala and Nimalendran (1996) use the structure equation model to examine the effect of earnings surprises on stock prices, trading volumes, and bid-ask spreads.

Goldberger (1972) and Jöreskog and Goldberger (1975) developed a structure equation model with multiple indicators and multiple causes of a single latent variable, MIMIC mode, and obtained maximum likelihood estimates of parameters. Fig. 1 shows the path diagram that depicts a simplified MIMIC model in which variables in a rectangular box denote observable variables, while variables in an oval box are latent constructs. In this diagram, observable variables X_1, X_2, and X_3 are causes of the latent variable η, while Y_1, Y_2, and Y_3 are indicators of η. In our study, X's are determinants of capital structure (η), which are then measured by Y's.

Jöreskog and Sörbom (1989) show that the full structural equation (LIEREL model) can be restricted to a MIMIC model. We here discuss the structural model and show how structural model can be restricted to a MIMIC model.

3.6.1. Structural model (LISEREL model)

A structural equation model is composed of two sub-models – structural sub-model and measurement sub-model. The structural model can be defined as

$$\eta = \Gamma X + \zeta, \tag{37}$$

[3] Please see Lee (1973) for the solution of Eq. (33).

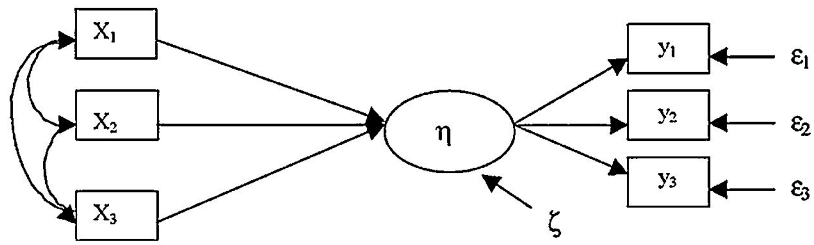


Fig. 1. Path diagram of a simplified MIMIC model.

$$\mathrm{Y} = \Lambda_{\mathrm{y}}\eta + \varepsilon, \tag{38}$$

where Y is a vector of indicators of the latent variable η, and X is a vector of causes of η.

The latent variable η is linearly determined by a set of observable exogenous causes, $X = (x_1, x_2, \ldots, x_q)'$, and a disturbance ζ. The latent variable η, in turn, linearly determines a set of observable endogenous indicators, $Y = (y_1, y_2, \ldots, y_p)'$ and a corresponding set of disturbance, $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_p)'$.[4]

3.6.2. MIMIC model

Substituting Eq. (48) into Eq. (49), we obtain a reduced form:

$$Y = \Lambda_y\eta + \varepsilon = \Lambda_y(\gamma' X + \zeta) + \varepsilon = \Pi' X + z \tag{39}$$

In structural equation modeling, the total effect of a cause variable on an indicator can be measured as the sum of the direct effect and the indirect effect. Since a MIMIC model is a reduced form of a structural equation model, the total effect of MIMIC model, denoted as Π' in Eq. (39), comes merely from the indirect effect.

Since the scale of the latent variable is unknown, the factor indeterminacy is a common problem in the MIMIC model, as in other structure equation models. We can obtain infinite parameter estimates from the reduced form by arbitrarily changing the scale of the latent variables. However, by fixing the scales of latent variables, one can solve the indeterminacy problem. Two methods are usually adopted to fix the scale of latent variables. One method is normalization in which a unit variance is assigned to each latent variable, while another method is to fix a non-zero coefficient at unity for each latent variable.

In terms of estimation of the parameters, Jöreskog and Goldberger (1975) adopt the normalization method to deal with the factor indeterminacy problem and use maximum likelihood estimation method in structural equation modeling to estimate parameters. The maximum likelihood estimates for the parameters of the model are obtained at the minimization of the fit function as follows:

$$F = \log||\Sigma|| + \mathrm{tr}(S\Sigma^{-1}) - \log||\mathrm{S}|| - (p-q), \tag{40}$$

where Σ is the population covariance matrix; S is the model-implied covariance matrix; p is the number of exogenous observable variables; and q is the number of endogenous observable variables. Minimization of the fit function can be done by the LISREL program provided by Jöreskog and Sörbom (1981).

3.7. Bayesian approach

Zellner (1970) uses the Bayesian approach to deal with measurement problems in the estimation of regression relationships containing unobservable independent variables. Zellner (1970) shows that the Bayesian approach can obtain optimal estimates under a finite sample. Several studies use Bayesian approaches to examine cost of capital (e.g. Lee & Wu, 1989) and asset pricing models (e.g. Ang & Chen, 2007; Davis, 2010; Geweke & Zhou, 1996; McCulloch & Rossi, 1991).

Davis (2010) develops a Bayesian approach and uses U.S. firm level data to reexamine the capital asset pricing model. The Bayesian approach can estimate all parameters simultaneously in one step and effectively avoid the errors-in-variables problem on the estimators induced from two-pass capital asset pricing test.

Davis (2010) uses a Bayesian approach to simultaneously estimate coefficients of the following three equation system.

$$r_{i,t,y} = \alpha_{i,y} + \gamma_{i,y} r_{m,t,y} + \delta_{i,y} r_{m,t-1,y} + \varepsilon_{i,t,y},$$
$$\text{where } \varepsilon_{i,t,y} \sim N(0, \sigma^2_{r_{i,y}}); \tag{41}$$

$$\overline{r_{i,y} - r_{f,y}} = c_{0,y} + c_{m,y}\beta_{i,y} + \eta_{i,y}, \quad \text{where } \eta_{i,y} \sim N(0, \sigma^2_y) \text{ and}$$
$$\beta_{i,y} \sim \gamma_{i,y} + \delta_{i,y}; \tag{42}$$

$$\mathbf{c}_y = \begin{bmatrix} c_{0,y} \\ c_{m,y} \end{bmatrix} \sim N(\mathbf{c}, \mathbf{V}_c), \quad \text{where } \mathbf{c} = \begin{bmatrix} c_0 \\ c_m \end{bmatrix} \quad \text{and}$$
$$\mathbf{V}_c = \begin{bmatrix} \sigma_0^2 & \sigma_{0m}^2 \\ \sigma_{0m}^2 & \sigma_m^2 \end{bmatrix}; \tag{43}$$

where $r_{i,t,y}$ is firm i's return in month t during the time period y, and $\overline{r_{i,y} - r_{f,y}}$ denotes the average monthly excess return for firm i during the time period y. The model allows firm-level βs to vary over each time period, y. The model also assumes that the joint normality of stock returns and market returns, contemporaneously estimated βs, and average excess returns are statistically independent.

In addition to deal with errors-in-variables problem, there are several advantages using the Bayesian approach to test the capital asset pricing model. First, the Bayesian approach allows βs to vary over time periods and firms and controls the inherent uncertainty associated with firm-level βs. Second, the Bayesian approach can modify the distribution assumptions in stock returns and market returns. Third, the Bayesian inference is free from the use of asymptotic approximations and therefore can be used under finite sample. Fourth, the Bayesian approach takes parameter uncertainty associated with all the model parameters into account.

4. Applications of errors-in-variables models in finance research

For the last four and a half decades, EIV models have been used to correct estimation bias associated with empirical results in various finance-related research issues. We here review four kinds of research, cost of capital, asset pricing models, capital structure, and investment equation, and discuss how EIV models can remedy measurement error problems induced from finance-related

[4] Stapleton (1978) further develops MIMIC with more latent variables.

research. It is therefore useful to understand the statistical properties of these EIV models in situations resembling real research question.

The main focus of this section is to discuss the measurement error problems on various empirical studies related to finance research and investigate how EIV models can remedy such problems. However, in empirical studies, it is impossible to observe variables without measurement error. We cannot evaluate EIV models and suggest a best EIV model for a certain circumstances. Instead, we here provide Table 1 to summarize the application of EIV models in finance-related research. Research topics, EIV models, specialties of EIV model, and results for each study are included in Table 1.

4.1. Cost of capital

Miller and Modigliani (1966) developed a theoretical expression for the value of a firm from which the firm's cost of capital could be derived. They assume a perpetual stream of earnings from real assets, and a constant capitalization rate (ρ), at which the market discounts the uncertain pure (unlevered) equity stream of earnings for some risk classes and perfect markets. It is thus possible to estimate the market capitalization rate (and thus the cost of capital) of a group of firms by performing a cross-sectional regression of the market value of the firm's equity on the expected average earnings of the firm, the market value of debt, and the growth resulting from the above-average investment opportunities. The above analysis suggests a cross-sectional regression:

$$(V - \tau_c D) = a_0 + a_1\overline{X}(1 - \tau_c) + a_2(\text{growth potential}) + \varepsilon, \tag{44}$$

where V is sum of the market value of all securities issued by the firm, τ_c is the corporate tax rate, D is the market value of a firm's debt, and $\bar{X}$ is the expected level of average annual earnings generated by current assets.

To avoid heteroscedasticity of regression residuals, the equation must be adjusted to compensate for the dominance of the large companies. Miller and Modigliani (1966) use weighted least square to adjust the standard deviation of the error term to firm size (deflating each variable by the book value of total assets). Therefore Eq. (51) can be adjusted to:

$$\frac{(V - \tau_c D)}{A} = \frac{a_0}{A} + a_1\bar{X}\frac{(1 - \tau_c)}{A} + a_2\frac{\Delta\bar{A}}{A} + u, \tag{45}$$

where $u = \varepsilon/A$. With this reformulation, the regression equation is expected to be homogeneous, that is, to have no constant term, and the term A, total assets, is used to avoid heteroscedasticity.

An additional problem beyond that of heteroscedasticity is the possible error of measurement associated with the earnings term. Since anticipated average earnings are essentially unobservable, accounting-statement estimates of earnings must be used instead. Therefore, the true relation between value and anticipated earnings, when replaced by the observable estimates, implies a simultaneous system of relationships:

$$V_i^* = \alpha X_i^* + \sum_j \beta_j Z_{ij} + u_i, \tag{46}$$

$$X_i = X_i^* + v_i, \tag{47}$$

$$X_i^* = \sum_j \delta_j Z_{ij} + w_i, \tag{48}$$

where $V_i^* = (V_i - \tau_c D_i)/A_i$, $X_i^* = (\bar{X}(1 - \tau_c))/A_i$ (the true anticipated earnings); v_i = measurement errors associated with current earnings; X_i = observable estimate of earnings derived from the accounting statements; and Z_{ij} = other relevant variables determining earnings. Equations (46)–(48) are related to anticipated earnings and a set of explanatory variables which may also be correlated with the firm's anticipated earnings.

In addition, the earnings variable used in the regression only approximates the true value of anticipated earnings, varying by the error of measurement, v_i. The system represents the simultaneous determination of two endogenous variables, V^* and X, by the Z_j exogenous variables. In regressing:

$$V_i^* = \alpha X_i + \sum \beta_j Z_{ij} + U' \tag{49}$$

the coefficients will be biased. The coefficient for earnings, α, will have a downward bias.

In an attempt to remedy the simultaneous-equation bias, Miller and Modigliani (1966) use an instrumental-variable approach. In this approach, the endogenous variable X is first regressed against all the instrumental variables, Z_j, to obtain estimates of the various coefficients. These estimates are then used to develop a new variable, X, which is

$$\hat{X}_i^* = \sum \hat{\delta}_j Z_{ij} \tag{50}$$

Depending on the choice of Z_j, the new estimate of earnings, $\hat{X}_i^*$, should be relatively free of the error measurement. It can then be used in the second-stage regression as the earnings variable. The resulting estimates of α and β can be shown to be consistent.

Miller and Modigliani (1966) hypothesized that the constant term was really zero. The reduction of bias on the estimates through the use of the two-stage process also seems to support the hypothesis that the constant term is zero. Miller and Modigliani (1966) state that the reason the constant term was significantly different from zero for the direct least-squares cases was that the error of measurement for earnings was large. This error is reduced by the two-stage process.

Higgins (1974) derives and tests a finite-growth model for the estimation of the cost of capital and share price of electric utility industry between 1960 and 1968. He suggests that the market value of equity is related to the trend of earnings and the trend of population in utility's service area. Assuming that observations of a variable consist of a true component and a random element, if such random elements have zero mean and are serially uncorrelated, the smoothing procedure can reduce potential errors in measurement. Empirical results show that the extrapolation of historical population trends is superior to the conventional use of change of capital, and share prices are not a positive function of dividends as often suggested.

Zellner (1970) proposes a least squares regression method to deal with potential errors-in-variables problems. He shows that his methods utilize more information than traditional instrumental variables methods do in dealing with an errors-in-variables problem. Lee and Wu (1989) further apply Zellner's method to reexamine Miller and Modigliani's (1966) cost of capital estimation for utility industry and obtain better cost of capital estimates than OLS methods and instrumental variable method.

More recently, Pastor, Sinha, and Swaminathan (2008) propose an implied cost of capital which is calculated by earnings forecasts and argue that the implied cost of capital can capture time variation in expected stock returns. Ortiz-Molina and Phillips (2014) adopt Pastor et al. (2008) method to investigate the relationship between real asset liquidity and the cost of capital, and find the implied cost of capital can mitigate measurement error problem on determine the cost of capital. Guay, Kothari, and Shu (2011) further propose the implied cost of capital corrected sluggish analyst forecast to

Table 1
Applications of errors-in-variables models in finance research.

Study	Issue	Method	Specialties/conditions	Results
Miller and Modigliani (1966)	Determinants of cost of capital	Instrumental variable method	Have to find exogenous variables	– Measurement error problem matters
Black et al. (1972)	CAPM test	Grouping (10 groups)	Panel data; 2-pass estimation	– Reject both the CAPM and the zero-beta CAPM
Blume and Friend (1973)	CAPM test	Grouping (12 groups)	Panel data; 2-pass estimation	– Linear model is better than quadratic model in explaining expected return – Reject both the CAPM and the zero-beta CAPM
Fama and MacBeth (1973)	CAPM test	Grouping (20 groups), period by period	Panel data; 2-pass estimation	– Find a linear relationship between the expected return and beta risk, beta is the only risk measure in explaining expected return, and risk premium is greater than zero – CAMP and efficient capital market hold
Higgins (1974)	Determinants of cost of capital	Smoothing procedure	Assume that measurement error has zero mean and serially uncorrelated	– The extrapolation of historical population trends is superior to the conventional use of change of capital, and share prices are not a positive function of dividends as often suggested
Lee (1977)	CAPM test	Wald's Grouping/Instrumental Variable	Adjust for measurement error of market return in first-step	– Estimated risk premium is larger than realized risk premium – Reject CAPM
Litzenberger and Ramaswamy (1979)	CAPM test	MLE, OLS, GLS	For individual stocks	– Before-tax expected rates of return are linearly related to systematic risk and dividend yield – MLE can obtain consistent estimators without losing efficiency – CAPM is rejected because of non-zero $\hat{\gamma}_0$
Cheng and Grauer (1980)	CAPM test	Grouping (20 groups)	Price-level testing (Invariance Law)	– Neither framework of Invariance Law or security market line can accommodate the possibility that the CAPM may hold for each period – Reject CAPM
Gibbons (1982)	CAPM test	One-step Gauss-Norman Procedure (40 groups)	One-step estimation	– Gauss-Norman procedure can increase the precision of estimated risk premium – Reject CAPM
Titman and Wessels (1988)	Determinants of capital structure	LISREL model	Deal with EIV problem due to the imperfect representation of proxy variables for interested attributes	– Do not support for four of eight propositions on the determinants of capital structure – A firm's capital structure is not significantly related to its non-debt tax shields, volatility of earnings, collateral value of assets, and future growth
Lee and Wu (1989)	Determinants of cost of capital	Zellner's EV method	Use sample information to estimate the ratio of error variances and construct an operational estimator	– Obtain better cost of capital estimates
MacKinlay and Richardson (1991)	CAPM test	GMM	Release the assumption of the normality of asset returns	– Conclusions of mean-variance efficiency vary by settings
Shanken (1992)	CAPM test	MLE	For individual stocks; deal with small-sample bias in the second-step cross-sectional regression estimates	– The adjustment does not have much effect on Fama and MacBeth's (1973) conclusion – Support CAPM
Fama and French (1992)	CAPM test	2-way grouping (10×10 groups)	Take size and book-to-market ratio into account	– The market capitalization and the book-to-market ratio can replace beta altogether – Reject CAPM
Jagannathan and Wang (1996)	CAPM test	Multifactor Asset Pricing Model	Test conditional CAPM	– Including human capital and business cycle can increase explanatory power of expected return – Support CAPM

Table 1 (*Continued*)

Study	Issue	Method	Specialties/conditions	Results
Kim (1995, 2010)	CAPM test	MLE	Closed form adjustment; for individual stocks or groups	– MLE method can effectively adjust the errors-in-variables bias and CAPM holds – Support CAPM
Kim (1997)	CAPM test	MLE	Closed form adjustment; for individual stocks or groups; for multifactor estimation	– Linear relationship between beta and expected return – Book-to-market ratio has significant explanatory power for expected return, but size has not
Lewbel (1997)	Elasticity of patent applications to R&D expenses	Instrumental variable method	Applicable if no outside data available for use as instruments	– TSLS estimator can mitigate the effects of measurement error – The estimated elasticity of patent applications with respect to R&D expenditures yields very close to one
Erickson and Whited (2000)	Test *q* theory	GMM	For balanced panel data	– Cash flow does not affect firms' financial decision, even for financially constrained firms – Support the *q* theory if measurement error is taken into account
Pastor et al. (2008)	Determinants of cost of capital	Implied cost of capital	Use earnings forecasts to compute implied cost of capital	– The implied cost of capital can capture time variation in expected stock returns
Chang et al. (2009)	Determinants of capital structure	MIMIC model	Allow several observable variables as indicators without multicollinearity problem	– Seven constructs, growth, profitability, collateral value, volatility, non-debt tax shields, uniqueness, and industry, as determinants of capital structure have significant effects on capital structure decision
Yang et al. (2009)	Determinants of capital structure	LISREL model	Jointly determine capital structure and return	– Stock returns, expected growth, uniqueness, asset structure, profitability, and industry classification are main determinants of capital structure – Leverage, expected growth, profitability, firm value, and liquidity can explain stock returns – The capital structure and stock return, in addition, are mutually determined by each other
Almeida et al. (2010)	Test *q* theory	GMM and instrumental variables method	Simple instrumental variable – lagged variable	– Estimators from GMM are unstable across different specifications and not economically meaningful – Estimators from a simple instrumental method are robust and conform to *q* theory
Davis (2010)	CAPM test	Bayesian approach	One-step estimation	– Positive relationship between excess return and market risk – Support CAPM
Jagannathan et al. (2010)	CAPM tests	Three-stage cross-sectional regression	Adjust for rolling window betas	– Dealing with measurement error in rolling window betas – Support CAPM
Guay et al. (2011)	Determinants of cost of capital	Corrected implied cost of capital	Sluggish analyst forecasts may result measurement error on implied cost of capital	– The corrected implied cost of capital can improve the ability to explain cross-sectional variation in future stock returns
Da et al. (2012)	CAPM tests	Three-stage cross-sectional regression	Adjust for rolling window betas	– Dealing with measurement error in rolling window betas – Support CAPM
Erickson and Whited (2012)	Test *q* theory	GMM	For unbalanced panel data	– Instrumental variables, dynamic panel estimators, and high-order moment estimators can perform well under correct specification – Developing a minimum distance technique allowing high-order moment estimators be used in unbalanced panel data

Table 1 (*Continued*)

Study	Issue	Method	Specialties/conditions	Results
Ortiz-Molina and Phillips (2014)	Determinants of cost of capital	Implied cost of capital	Use earnings forecasts to compute implied cost of capital	– The implied cost of capital can mitigate measurement error problem on determine the cost of capital
Lee and Tai (2014)	Determinants of capital structure	SEM with CFA approach	Jointly determine capital structure and return	– SEM with CFA approach outperforms MIMIC model and 2SLS method in terms of the joint determinants of capital structure and stock return

improve the ability to explain cross-sectional variation in future stock returns.

4.2. *Capital asset pricing model*

The capital asset pricing model (CAPM) developed by Sharpe (1964), Lintner (1965) and Mossin (1966) implies that the expected returns on securities and their market risks (β) are positively and linearly correlated and that market risks have sufficient power to explain expected returns of securities. Black et al. (1972), Fama and MacBeth's (1973), and others use the two-step method to test CAPM. In the first step, estimated betas are obtained by time-series market model for each security. In the second step, the estimated betas are used in testing the linear relationship between betas and expected returns on securities. Because estimated betas are subjected to a measurement error (estimation error) problem, there exists an EIV problem in the second step. The EIV problem will result in estimating the explanatory power of beta and the estimated rate of return on beta risk. More specifically, the EIV problem leads to an underestimation of the coefficient associated with beta risk. Although the EIV problem exists in the two-step method of the asset pricing test, most researchers do not carefully use sophisticated econometric and statistical methods to deal with this kind of problem. Roll (1969, 1977) shows that the testing asset pricing model suffers an EIV problem, concluding that (i) no correct and unambiguous test of the theory has appeared in the literature, and (ii) practically no possibility exists that such a test can be accomplished in the future. Roll and Ross (1994) show that the measurement error problem of market rate of return can bias the empirical test of CAPM.

Several studies focus on beta estimation in the first step to solve the EIV problem in testing CAPM. Brennan (1970), Lee and Jen (1978) and Roll (1977) show that the possible measurement error in market beta risk is the unobserved market rate of return and risk-free rate of return.[5] To improve the beta estimator, Fabozzi and Francis (1978) and Lee and Chen (1979) use the random coefficient procedure to estimate random coefficient betas. Brennan and Schwartz (1977), Brennan (1979) and Brown and Warner (1980) also provide different types of market models which can produce different results in predicting rates of return, testing efficient market hypotheses, and measuring security price performance.

To reduce the impact of the measurement error problem in the second step, Black et al. (1972) use the grouping method in the capital asset pricing model test. Although the results support the linear relationship between the systematic risk and expected return, CAPM cannot hold because of the non-zero intercept term and the lower market premium in the cross-sectional regression. Blume and Friend (1973) and Fama and MacBeth (1973) also use the grouping method in testing CAPM and show that the CAPM is valid. However, Jagannathan, Skoulakis, and Wang (2009) show that the time series average of the cross-sectional estimators converges in probability to the true value of the estimator. Although their results support CAPM indicating a linear relationship between the systematic risk and the expected return, the lower value of time-series average of the market premium shows that the measurement error problem still exists after Fama and MacBeth's (1973) grouping method.

Considering the measurement errors of the market rate of return and risk-free rate of return, Lee (1977) uses two EIV estimation methods, Wald's two-group grouping method and Durbin's instrumental variable method, to adjust the estimated beta risk in the first step of the capital asset pricing test. Although correcting the measurement errors induced from the unobservable market rate of return, Lee (1977) finds that the predictive ability of the capital asset pricing model is still poor.

Litzenberger and Ramaswamy (1979) derive an after-tax version of CAPM and show that, in the equilibrium, the before-tax expected return on a security is linearly related to its systematic risk and its dividend yield. Litzenberger and Ramaswamy (1979) further empirically test both the before tax and the after-tax versions of CAPM. Instead of grouping method, Litzenberger and Ramaswamy (1979) use maximum likelihood estimation in the second-step regression to test the before-tax and the after-tax versions of capital asset pricing model. Although maximum likelihood estimators are consistent, the average risk premium is small and not significantly different from zero.

Given the EIV bias in the two-step CAPM test, Gibbons (1982) introduces a one-step Gauss-Newton procedure and uses the maximum likelihood method to obtain the estimated price of systematic risk. Because the one-step Gauss-Newton procedure does not use estimated beta as an explanatory variable in a regression model, the measurement errors problem of estimated beta can be avoided. Shanken (1992) shows that using generalized least square (GLS) on the second step of CAPM test can yield an estimator identical to the Gauss-Newton estimator obtained by Gibbons' (1982) maximum likelihood method. Gibbons (1982) shows that the Gauss-Newton procedure increases the precision of estimated risk premium, but rejects the mean-variance efficiency of the market portfolio.

Shanken (1992), who provides a modified version of the two-step estimator by using maximum likelihood estimation, finds that Fama and MacBeth's two-step procedure overstates the precision of the estimator in the second-step and therefore provides an adjusted standard error for the estimator in the second-step regression. Jagannathan and Wang (1998a, 1998b) further

[5] Roll (1969), Roll (1977) and Lee and Jen (1978) show that the observed market rate returns in terms of stock market index are measured with errors since the stock market index does not include all assets which investors can invest. Lee and Jen (1978) have theoretically shown how beta estimate and Jensen performance measures can be affected by both constant and random measurement errors of R_m and R_f. Diacogiannis and Feldman (2011), Green (1986), Roll and Ross (1994) and Gibbons and Ferson (1985) have argued that market portfolio measure with errors is an inefficient portfolio and show how the inefficient benchmark can affect theoretical CAPM derivation. Diacogiannis and Feldman (2011) provide a pricing model that uses inefficient benchmarks, a two beta model, one induced by the benchmark, and one adjusting for its inefficiency.

release Shanken's assumption that asset returns are conditional homoscedasticity to derive a more general standard error for the second-step estimator by generalized least squares.

Fama and French (1992) use a two-way sort grouping method to control for size effect, and find a weak relationship between beta risk and expected return. Before reaching the conclusion that the capital asset pricing has not been valid in the recent years, one possible reason that may be considered is that the measurement error problem cannot be fully eliminated by the grouping method, and results of CAPM test may vary depending on the portfolio formation technique (e.g. Ahn et al., 2009).

To deal with the problem of EIV in testing CAPM, Kim (1995) provides a maximum likelihood method, extracting information associated with the relationship between the measurement error variance and idiosyncratic error variance and incorporating such information into the maximum likelihood estimation in the second step of the capital asset pricing model test. Given the assumption that the disturbance term of the market model is homoscedasticity, the corrected factors for the traditional least squares estimators of the cross-sectional regression coefficients can be obtained. Although Kim's (1995) maximum likelihood method can only deal with the EIV problem of the estimated beta in the first pass, the maximum likelihood method can test, besides the capital asset pricing model, the multifactor asset pricing models. Kim (1995) uses maxima likelihood method to reexamine CAPM and multi-factor asset pricing model and finds more support for the role of market beta risk and less support for the role of firm size. His results show that the prominent risk factors (e.g. size, book-to-market ratio, and momentum factors) might result a different explaining power for cross-sectional stock returns after correcting the EIV problem.

MacKinlay and Richardson (1991) use generalized method of moments (GMM) to test the mean–variance efficiency. They theoretically show that the estimator from GMM and the estimator from maximum likelihood method are equivalent when stock returns are conditionally homoscedasticity, but GMM can avoid the EIV problem by estimating coefficients in one step. Empirical and simulation results show that the conclusion mean–variance efficiency of market indexes is sensitive to the model settings.

Chen (2011) offers an empirical examination of various EIV estimation methods in the testing of CAPM, including the grouping method, the instrumental variable method, and the maximum likelihood method. Both potential measurement error problems of market return in the first pass and estimated beta in the second pass are corrected by either the grouping method or the instrumental variable method. Chen (2011) shows that empirical results support the role of market beta in the capital asset pricing model after correcting the EIV problem.

To deal with the measurement error problem associated with testing both CAPM and APT, Lee and Wei (1984) and Wei (1984) use the MIMC model to test whether APT outperformed CAPM. Betas are obtained from simultaneous equation system, and a cross-sectional regression of the security return against its β will be used to test the CAPM. They conclude that the beta estimated from the MIMC model by allowing measurement error on the market portfolio does not significantly improve the OLS beta estimate, and MLE estimator does a better job than the OLS and GLS estimators in the cross-sectional regressions because the MLE estimator takes care of the measurement error in beta.

More recently, Jagannathan, Kim, and Skoulakis (2010) and Da, Guo, and Jagannathan (2012) use three-stage cross-sectional regression to correct the errors-in-variables problem from the rolling-window betas. Their empirical findings support CAPM in explaining option-adjusted stock returns at the individual stock level.

4.3. Capital structure

Titman and Wessels (1988), Chang et al. (2009) and Yang et al. (2009) use structure equation models (e.g. LISREL model and MIMIC model) to mitigate the measurement problems of proxy variables when working on capital structure theory. Titman and Wessels (1988) use LISREL method to investigate determinants of capital structure. In the structure equation model, they use 15 indicators associated with eight latent variables and set 105 restrictions on the coefficient matrix. Empirical results, however, do not support four of eight propositions on the determinants of capital structure. Specifically, their results show that a firm's capital structure is not significantly related to its non-debt tax shields, volatility of earnings, collateral value of assets, and future growth. One possible reason for the poor results is that the indicators used in the empirical study do not adequately reflect the nature of the attributes suggested by financial theory.

Chang et al. (2009) apply a Multiple Indicators and Multiple Causes model (MIMIC) with refined indicators to reexamine Titman and Wessels (1988) work on determinants of capital structure. Chang et al. (2009) examine the seven indicator factors as follows: growth, profitability, collateral value, volatility, non-debt tax shields, uniqueness, and industry. Their empirical results show that the growth is the most influential determinant on capital structure, followed by profitability, and then collateral value. Under a simultaneous cause–effect framework, their seven constructs as determinants of capital structure have significant effects on capital structure decision.

Yang et al. (2009) apply a LISREL model to find determinants of capital structure and stock returns, and estimate the impact of unobservable attributes on capital structure decisions and stock returns. Using leverage ratios and stock returns as two endogenous variables and 11 latent factors as exogenous variables, Yang et al. (2009) find that stock returns, expected growth, uniqueness, asset structure, profitability, and industry classification are main determinants of capital structure, while leverage, expected growth, profitability, firm value, and liquidity can explain stock returns. In addition, the capital structure and stock return are mutually determined by each other.

Lee and Tai (2014) develop a simultaneous determination model to identify the joint determinants of capital structure and stock returns. The structural equation model with confirmatory factor analysis shows that stock return, asset structure, growth rate, industry classification, uniqueness, volatility and financial rating, profitability, government financial policy, and managerial entrenchment are key factors in determining a firm's capital structure. Such results are robust in the MIMIC and two-stage least square methods.

4.4. Measurement error in investment equation

Modern q theory, developed by Lucas and Prescott (1971) and Mussa (1977), shows that the shadow value of capital, marginal q, is the firm manager's expectation of the marginal contribution of new capital goods to future profits. Marginal q, therefore, should summarize the effects of all factors relevant to the investment decision. Lucas and Prescott (1971) and Hayashi (1982) show that the equality of marginal q with average q is under the assumptions of constant returns to scale and perfect competition. Because the marginal q is unobservable in the real world, most of empirical studies adopt Lucas and Prescott (1971) and Hayashi's (1982) assumption and use average q instead of marginal q to test q theory. In addition, if financial markets' valuation of the capital will be equal to the manager's valuation, average q, should equal an observable value, Tobin's q, defined as the ratio of the market value to the replacement value. Most empirical studies use Tobin's q as a proxy for marginal q to test the q theory of investment. However, their empirical results are inconsistent to the q theory (e.g. Blundell, Bond, Devereux, and Schiantarelli, 1992; Fazzari, Hubbard, and Petersen, 1988; Gilchrist & Himmelberg, 1995; Schaller, 1990).[6]

The model introduced by Fazzari et al. (1988) is

$$\frac{I_{it}}{K_{it}} = \eta_i + \beta q_{it}^* + \alpha \frac{CF_{it}}{K_{it}} + u_{it}, \tag{51}$$

where I_{it} represents the investments of firm i at time t, K_{it} is capital stock of firm i at time t, q_{it}^* is the marginal q_{it}, CF_{it} is cash flow of firm i at time t, η_i is the firm-specific effect, and u_{it} is the innovation term.

Almeida et al. (2010) show that OLS estimated coefficient of independent variable with measurement error, q_{it}^*, will be biased downward, and OLS estimated coefficient of the independent variable without measurement error, CF_{it}/K_{it}, will be biased upward. Following Eq. (51), if one of independent variables has measurement error and the other independent variables has no measurement error; the asymptotic biases of estimated coefficients can be defined as:

$$\text{p}\lim \hat{\beta} - \beta = \frac{-\beta \sigma_\varepsilon^2}{\sigma_{CF/K}^2 - b_{CF/K,q^*} + \sigma_\varepsilon^2}, \tag{52}$$

and

$$\text{p}\lim \hat{\alpha} - \alpha = \beta b_{CF/K,q^*} \left(\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + \sigma_{CF/K}^2 (1 - R_{CF/K,q^*}^2)} \right), \tag{53}$$

in which $\sigma_{CF/K}^2$ is the variance of CF/K, σ_ε^2 is the variance of error term between unobserved marginal q and observable average q^*, $b_{CF/K,q^*}$ is the auxiliary regression coefficient of a regressing q^* on CF/K, and $R_{CF/K,q^*}^2$ is the correlation coefficient between q^* and CF/K. We know that $\sigma_{CF/K}^2 - b_{CF/K,q^*} + \sigma_\varepsilon^2$ is generally positive, so the estimated coefficient of q^* is downward biased. In addition, the direction of bias of estimated coefficient of CF/K will depend on the signs of β and $b_{CF/K,q^*}$. Given that q and cash flow are positively correlated, we can get the conclusion of Almeida et al. (2010) that $\hat{\beta}$ is downward biased and $\hat{\alpha}$ is upward biased.

Erickson and Whited (2000) argue that the measurement error of marginal q can result in different implications in empirical q models. They incorporate an EIV model to reexamine the empirical work done by Fazzari et al. (1988). By using generalized method of moments (GMM), Erickson and Whited (2000) obtain consistent estimators that the information contained in the third- and higher-order moments of the joint distribution of the observed regression variables. The estimator precision and consistency can be increased by exploit the information afforded by an excess of moment equations over parameters. Results show that cash flow does not affect a firms' financial decision, even for financially constrained firms, and the q theory is held if measurement error is taken into account.

Almeida et al. (2010) use Monte Carlo simulations and real data to compare the performance of generalized method of moments and instrumental variables approach dealing with measurement error problems in investment equations. In Monte Carlo simulations, they find estimators of GMM proposed by Erickson and Whited (2000) are biased for both mismeasured and well-measured regressors when the data have individual-fixed effects, heteroscedasticity, or no high degree of skewness. In contrast, the instrumental variable method results fairly unbiased estimators under those same conditions. Almeida et al. (2010) further empirically examine the investment equation introduced by Fazzari et al. (1988) by using GMM and instrumental variable method. Almeida et al. (2010) adopt Biorn's (2000) method using the lags of the variable as instruments in testing the investment equation. Empirical results show that estimators from generalized method of moments are unstable across different specifications and not economically meaningful, while estimators from a simple instrumental method are robust and conform to q theory. Almeida et al. (2010) conclude that instrumental method yields more consistent estimators and support the q theory in the investment equation.

In contrast, Erickson and Whited (2012) compare the ability of three errors-in-variables models, instrumental variables, dynamic panel estimators, and the high-order moment estimators in investment equation. They conclude that all of three models can perform well under correct specification, while the high-order moment estimators often outperform the instrumental variables and dynamic panel estimators in terms of bias and dispersion. Erickson and Whited (2012) also demonstrate a minimum distance technique to extend the high-order moment estimators used on unbalanced panel data.

5. Conclusion

In this paper, we investigate theoretical issues related to errors-in-variables (EIV) problem, and review how existing EIV estimation methods deal with measurement error problem. We first show how EIV problems affect the coefficients of independent variables in the regression model. We then discuss how classical method, mathematical programming method, grouping method, instrumental variable method, maximum likelihood method, and LISREL method deal with EIV problems. We further investigate how alternative EIV models have been used in empirical finance research. We find that the empirical research of cost of capital, asset pricing, capital structure, and investment equation have used alternative EIV methods to improve the empirical results. Not only can the reader of this paper understand the important research topics in finance, but also can the reader realize how measurement error problems affect the results of empirical work in such research topics. Finally, we suggest that future empirical studies on finance related issues should pay more efforts to deal with EIV problems and obtain more robust empirical results.

References

Ahn, D., Conrad, J., & Dittmar, R. F. (2009). Basis asset. *Review of Financial Studies*, *22*, 5122–5174.

Almeida, H., Campello, M., & Galvao, A. F., Jr. (2010). Measurement errors in investment equations. *Review of Financial Studies*, *23*, 3279–3328.

Anderson, T. W. (1963). The use of factor analysis in the statistical analysis of multiple time series. *Psychometrika*, *28*, 1–25.

Ang, A., & Chen, J. (2007). CAPM over the long run: 1926–2001. *Journal of Empirical Finance*, *14*, 1–40.

Barnett, V. D. (1967). A not on linear Structural relationship when both residual variances are known. *Biometrika*, *54*, 670–672.

Bentler, P. M. (1983). Some contributions to efficient statistics in structural models: Specification and estimation of moment structures. *Psychometrika*, *48*, 493–517.

Biorn, E. (2000). Panel data with measurement errors: Instrumental variables and GMM procedures combining levels and differences. *Econometric Reviews*, *19*, 391–424.

Black, F., Jensen, M. C., & Scholes, M. (1972). The capital asset pricing model: Some empirical tests. In *Studies in the theory of capital markets*. New York: Praeger.

Blume, M. E., & Friend, I. (1973). A new look at the capital asset pricing model. *Journal of Finance*, *28*, 19–33.

Blundell, R., Bond, S., Devereux, M., & Schiantarelli, F. (1992). Investment and Tobin's Q: Evidence from company panel data. *Journal of Econometrics*, *51*, 233–257.

Brennan, M. J. (1970). Taxes, market valuation and corporate financial policy. *National Tax Journal*, *23*, 417–427.

[6] Empirical work in testing association between the investment decision and cash flow shows that cash flow has poor explanation in determining investment decision. In addition to cash flow, output, sales, and internal funds have significant explanation in determining investment decision.

Brennan, M. J. (1979). The pricing of contingent claims in discrete time models. *Journal of Finance*, *34*, 53–68.

Brennan, M. J., & Schwartz, E. S. (1977). Convertible bonds: Valuation and optimal strategies for call and conversion. *Journal of Finance*, *32*, 1699–1715.

Brown, S. J., & Warner, J. B. (1980). Measuring security price performance. *Journal of Financial Economics*, *8*, 205–258.

Chang, C., Lee, A. C., & Lee, C. F. (2009). Determinants of capital structure choice: A structural equation modeling approach. *Quarterly Review of Economics and Finance*, *49*, 197–213.

Chen, H. Y. (2011). *Momentum strategies, dividend policy, and asset pricing test* (Ph.D. dissertation). Rutgers: State University of New Jersey.

Cheng, P. L., & Grauer, R. R. (1980). An alternative test of the capital asset pricing model. *American Economic Review*, *70*, 660–671.

Clutton-Brock, M. (1967). Likelihood distributions for estimating functions when both variables are subject to error. *Technometrics*, *9*, 261–269.

Cochran, W. G. (1970). Some effects of errors of measurement on multiple correlation. *Journal of the American Statistical Association*, *65*, 22–34.

Da, Z., Guo, R. J., & Jagannathan, R. (2012). CAPM for estimating the cost of equity capital: Interpreting the empirical evidence. *Journal of Financial Economics*, *103*, 204–220.

Davis, P. (2010). *A firm-level test of the CAPM. Working paper*. Rutgers University.

Deming, W. E. (1943). *Statistical adjustment of data*. New York: John Wiley and Sons.

Diacogiannis, G., & Feldman, D. (2011). *Linear beta pricing with inefficient benchmarks. Working paper*. University of Piraeus.

Durbin, J. (1954). Errors in variables. *Review of the International Statistical Institute*, *22*, 23–32.

Erickson, T., & Whited, T. (2000). Measurement error and the relationship between investment and *q*. *Journal of Political Economy*, *108*, 1027–1057.

Erickson, T., & Whited, T. (2002). Two-step GMM Estimation of the errors-in-variables model using higher-order moments. *Econometric Theory*, *18*, 776–799.

Erickson, T., & Whited, T. (2012). Treating measurement error in Tobin's *Q*. *Review of Financial Studies*, *25*, 1286–1329.

Fama, E. F., & French, K. R. (1992). The cross-section of expected stock returns. *Journal of Finance*, *47*, 427–466.

Fama, E. F., & MacBeth, J. D. (1973). Risk, return and equilibrium: Empirical tests. *Journal of Political Economy*, *81*, 607–636.

Fabozzi, F. J., & Francis, J. C. (1978). Beta as a random coefficient. *Journal of Financial and Quantitative Analysis*, *13*, 101–116.

Fazzari, S., Hubbard, R. G., & Petersen, B. (1988). Financing constrains and corporate investment. *Brookings Papers on Economic Activity*, *1*, 141–195.

Geweke, J., & Zhou, G. (1996). Measuring the pricing error of the arbitrage pricing theory. *Review of Financial Studies*, *9*, 557–587.

Gibbons, M. R. (1982). Multivariate tests of financial models: A new approach. *Journal of Financial Economics*, *10*, 3–27.

Gibbons, M. R., & Ferson, W. (1985). Testing asset pricing models with changing expectations and an unobservable market portfolio. *Journal of Financial Economics*, *14*, 217–236.

Gilchrist, S., & Himmelberg, C. P. (1995). Evidence on the role of cash flow for investment. *Journal of Monetary Economics*, *36*, 541–572.

Goldberger, A. S. (1972). Structural equation methods in the social sciences. *Econometrica*, *40*, 979–1001.

Green, R. (1986). Benchmark portfolio inefficiency and deviations from the security market line. *Journal of Finance*, *41*, 295–312.

Griliches, Z., & Hausman, J. A. (1986). Errors in variables in panel data. *Journal of Econometrics*, *31*, 93–118.

Guay, W., Kothari, S. P., & Shu, S. (2011). Properties of implied cost of capital using analysts' forecasts. *Australian Journal of Management*, *36*, 125–149.

Hayashi, F. (1982). Tobin's marginal *q* and average *q*: A neoclassical interpretation. *Econometrica*, *50*, 213–224.

Higgins, R. C. (1974). Growth, dividend policy and capital cost in the electric utility industry. *Journal of Finance*, *29*, 1189–1201.

Jöreskog, K. G., & Goldberger, A. S. (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association*, *70*, 631–639.

Jöreskog, K. G., & Sörbom, D. (1981). *LISREL V: Analysis of Linear Structural Relationships by the Method of Maximizing Likelihood, User's Guide*. IL: National Educational Resources, Inc.

Jöreskog, K. G., & Sörbom, D. (1989). *LISREL 7: A Guide to the Program and Applications*. Chicago, IL: SPSS Inc.

Jagannathan, R., Kim, S., & Skoulakis, G. (2010). *Revisiting the errors in variables problem in studying the cross section of stock returns. Working paper*. Northwestern University.

Jagannathan, R., Skoulakis, G., & Wang, Z. (2009). The analysis of the cross section of security returns. In Y. Aït-Sahalia, & L. Hansen (Eds.), *Handbook of Financial Econometrics* (Vol. 2) (pp. 73–134). Amsterdam: North-Holland.

Jagannathan, R., & Wang, Z. (1998a). The conditional CAPM and the cross-section of expected returns. *Journal of Finance*, *51*, 3–53.

Jagannathan, R., & Wang, Z. (1998b). An asymptotic theory for estimating beta-pricing models using cross-sectional regression. *Journal of Finance*, *53*, 1285–1309.

Jagannathan, R., & Wang, Z. (1996). The conditional CAPM and the cross-section of expected returns. *Journal of Finance*, *51*, 3–53.

Kendall, M. G., & Stuart, A. (1961). . *The advanced theory of statistics* (Vol. II) London: Griffin.

Kiefer, J. (1964). Review of Kendall and Stuart's advanced theory of statistics – II. *Annals of Mathematical Statistics*, *35*, 1371–1380.

Kim, D. (1995). The errors in the variables problem in the cross-section of expected stock returns. *Journal of Finance*, *50*, 1605–1634.

Kim, D. (1997). A reexamination of firm size, book-to-market, and earnings price in the cross-section of expected stock returns. *Journal of Financial and Quantitative Analysis*, *32*, 463–489.

Kim, D. (2010). Issues related to the errors-in-variables problems in asset pricing tests. In C. F. Lee, A. C. Lee, & J. Lee (Eds.), *Handbook of quantitative finance and risk management*. Springer.

Lee, C. F. (1973). *Errors-in-variables estimation procedures with applications to a capital asset pricing model* (Ph.D. dissertation). State University of New York at Buffalo.

Lee, C. F. (1977). Performance measure, systematic risk and errors-in-variable estimation method. *Journal of Economics and Business*, *29*, 122–127.

Lee, C. F. (1984). Random coefficient and errors-in-variables models for beta estimates: Methods and application. *Journal of Business Research*, *12*, 505–516.

Lee, C. F., & Chen, S. N. (1979). A random coefficient model for reexaming risk decomposition method and risk-return relationship test. *Financial Review*, *14*, 65–65.

Lee, C. F., & Jen, F. C. (1978). Effects of measurement errors on systematic risk and performance measure of a portfolio. *Journal of Financial and Quantitative Analysis*, *13*, 299–312.

Lee, C. F., & Tai, T. (2014). *The Joint determinants of capital structure and stock rate of return: A LISREL model approach. Working paper*. Rutgers University.

Lee, C. F., & Wei, K. C. (1984). *Multi-factor, multi-indicator approach to asset pricing: Methods and empirical evidence. Working paper*. University of Illinois at Urbana-Champaign.

Lee, C. F., & Wu, C. C. (1989). Using Zellner's errors-in-variables model to reexamine MM's valuation model for the electric utility industry. *Advances in Financial Planning and Forecasting*, *3*, 63–73.

Lewbel, A. (1997). Constructing instruments for regressions with measurement error when no additional data are available, with an application to patents and R&D. *Econometrica*, *65*, 1201–1213.

Lewbel, A. (2012). Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models. *Journal of Business and Economic Statistics*, *30*, 67–80.

Lintner, J. (1965). The valuation of risky assets and the selection of risky investments in stock portfolio and capital budgets. *Review of Economics and Statistics*, *47*, 13–37.

Litzenberger, R., & Ramaswamy, K. (1979). The effects of personal taxes and dividends on capital asset prices: Theory and empirical evidence. *Journal of Financial Economics*, *7*, 163–195.

Lucas, R. E., Jr., & Prescott, E. C. (1971). Investment under uncertainty. *Econometrica*, *39*, 659–681.

MacKinlay, A. C., & Richardson, M. P. (1991). Using generalized method of moments to test mean–variance efficiency. *Journal of Finance*, *46*, 511–527.

Maddala, G. S., & Nimalendran, M. (1996). Error-in-variables problems in financial models. In G. S. Maddala, & C. R. Rao (Eds.), *Handbook of statistics*. New York: Elsevier Science.

McCulloch, R., & Rossi, P. E. (1991). A Bayesian approach to testing the arbitrage pricing theory. *Journal of Econometrics*, *49*, 141–168.

Miller, M. H., & Modigliani, F. (1966). Some estimates of the cost of capital to the electric utility industry, 1954–57. *American Economic Review*, *56*, 333–391.

Mossin, J. (1966). Equilibrium in a capital asset market. *Econometrica*, *34*, 768–783.

Mussa, M. (1977). External and internal adjustment costs and the theory of aggregate and firm investment. *Economica*, *44*, 163–178.

Ortiz-Molina, H., & Phillips, G. M. (2014). Real asset illiquidity and the cost of capital. *Journal of Financial and Quantitative Analysis*, *49*, 1–32.

Pastor, L., Sinha, M., & Swaminathan, B. (2008). Estimating the intertemporal risk-return tradeoff using the implied cost of capital. *Journal of Finance*, *63*, 2859–2897.

Roll, R. (1969). Bias in fitting the Sharpe model to time-series data. *Journal of Financial and Quantitative Analysis*, *4*, 271–289.

Roll, R. (1977). A critique of the asset pricing theory's tests: Part 1. On past and potential testability of the theory. *Journal of Financial Economics*, *4*, 129–176.

Roll, R., & Ross, S. A. (1994). On the cross-sectional relation between expected returns and betas. *Journal of Finance*, *49*, 101–121.

Schaller, H. (1990). A re-examination of the q theory of investment using US firm data. *Journal of Applied Econometrics*, *5*, 309–325.

Shanken, J. (1992). On the estimation of beta pricing models. *Review of Financial Studies*, *5*, 1–33.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, *19*, 425–442.

Stapleton, D. C. (1978). Analyzing political participation data with a MIMIC model. In K. F. Schuessler (Ed.), *Sociological methodology*. San Francisco, CA: Jossey-Bass.

Titman, S., & Wessels, R. (1988). The determinants of capital structure choice. *Journal of Finance*, *43*, 1–19.

Theil, H. (1971). *Principles of econometrics*. Toronto, NY: John Wiley & Sons.

Wald, A. (1940). The fitting of straight lines if both variables are subject to error. *Annals of Mathematical Statistics*, *11*, 284–300.
Wei, K. C. (1984). *The arbitrage pricing theory versus the generalized intertemporal capital asset pricing model: Theory and empirical evidence* (Ph.D. dissertation). University of Illinois at Urbana-Champaign.
Yang, C. C., Lee, C. F., Gu, Y. X., & Lee, Y. W. (2009). Co-determination of capital structure and stock returns – A LISREL approach: An empirical test of Taiwan stock markets. *Quarterly Review of Economics and Finance*, *50*, 222–233.
York, D. (1966). Least-squares fitting of a straight line. *Canadian Journal of Physics*, *44*, 1079–1086.
Zellner, A. (1970). Estimation of regression relationships containing unobservable variables. *International Economic Review*, *11*, 441–454.

Decomposing the Size Premium

Zhiyao Chen*, Jun Li†, Huijun Wang‡

January 15, 2017

Abstract

We decompose firm size into four components: the lagged 5-year component that represents size five years ago, and the long-run, intermediate-run, and short-run components that capture changes in size in each horizon. Our analyses indicate that while the lagged 5-year component explains about 80% of the cross-sectional variation in size, it has little return predictability. In contrast, the long-run change in size component explains only 18% of size, but it completely captures the size premium. Our decomposition also sheds light on the January effect, the disappearance of the size premium since early 1980s, and the return behaviors of new entrants.

*Department of Finance, The Chinese University of Hong Kong, 12 Chak Cheung Street, Shatin, N.T. Hong Kong. e-mail: nicholaschen@baf.cuhk.edu.hk

†Department of Finance and Managerial Economics, University of Texas at Dallas, 800 West Campbell Road, SM 31, Richardson, TX 75080. e-mail: Jun.Li3@utdallas.edu.

‡Lerner College of Business and Economics, University of Delaware, 305 Purnell Hall, Newark, DE 19716. e-mail: wangh@udel.edu.

1 Introduction

Size premium, the empirical finding that small stocks (measured by market capitalization) outperform big stocks on average, is one of the most well-known investment strategies in the stock market. In the asset pricing literature, besides the market factor, the small-minus-big (SMB) size factor is the only factor that is included in all leading multi-factor asset pricing models, including Fama and French (1993) three-factor model, Carhart (1997) four-factor model, and more recent Hou, Xue, and Zhang (2015) four-factor model based on q theory of investment and Fama and French (2015) five-factor model. In this paper, we examine the size premium from a different perspective from many existing studies in the literature. We decompose the firm size into four components and study how each component contributes to current firm size and to the size premium.

Our decomposition is motivated by the well-known cross-sectional stock return patterns at various horizons, namely, long-term contrarian (De Bondt and Thaler (1985)), intermediate-term momentum (Jegadeesh and Titman (1993)), and short-term reversal (Jegadeesh (1990)). The first component in our size decomposition is the (log) firm size five years ago, capturing the extremely persistent component of firm size. The second component, the long-run component, measures the cumulative change in (log) firm size during prior 13-60 months, following the timing of long-term contrarian strategy. The third component, the intermediate-run component, captures the cumulative change in (log) firm size during prior 2-12 months, corresponding to the timing of the price momentum strategy. The last component is the short-run component, defined as the prior 1-month change in (log) firm size, consistent with the timing of the short-term reversal.

Our empirical analyses suggest that despite being the most important determinant of current firm size, the lagged 5-year component has little predictive power for future stock returns. Compared with the 4.93% annual size premium, the premium based on the lagged 5-year component is only 1.53% per year with a t-statistic of 0.63. Controlling for the market factor further reduces the magnitude of the lagged 5-year size premium to -0.59% per year. Given that firm size is highly persistent, this result indicates that the size premium does not originate from the *level* of firm size; instead, it is the *changes* in firm size during past years that possess the predictive power for future stock returns. The second component is the long-run change in size, which explains an average of 18.4% of the cross-sectional variation in firm size. However, this component strongly predicts stock returns. In decile portfolios sorted by this long-run component, the difference in stock returns between firms with the most decrease in market value and firms with most increase in market value in the prior 13-60 months is 7.33% per year with an annual Sharpe ratio of 0.5. As a comparison, the Sharpe ratio for the size premium based on the same sample is only 0.3. The intermediate-run component is also a strong return predictor, with stocks with most increase in market value in the prior 2-12 months outperforming stocks with most decrease in market value by 8.64% per year, which has a similar magnitude to the momentum profit. All else being equal, small firms tend to have worse past stock performance than big firms, so the size strategy contains a short position in momentum which contributes negatively to the size premium. Despite its large premium, the intermediate-run component explains only about 2.6% of firm size, so its overall effect on the size premium is small. For the same reason, the short-run component only explains less than 1% of the cross-sectional variation in size.

The relative performances of the strategies based on size components suggest that the size premium is mainly driven by the long-run component, which we further confirm in several ways. First, in an independent double sort by firm size and its long-run component, we find that conditional on firm size, the average premium based on the long-run component is 5.42% with a t-statistic of 3.84, whereas the average size premium conditional on the long-run component is only 2.65% per year with a t-statistic of 1.31. Controlling for the market factor further amplifies the difference, generating a capital asset pricing model (CAPM) alpha of 6.11% and 1.41%, respectively. Second, we conduct linear factor model time series regressions tests. When size portfolio returns are regressed on the long-run size component factor (together with the market factor), none of these portfolios has a statistically significant abnormal return, including the long-short size portfolio. On the other hand, when returns of portfolios sorted by the long-run component are regressed on the size premium factor (together with the market factor), we find a significant abnormal return of more than 3 standard errors from zero for the long-short spread portfolio. Our last test is Fama-MacBeth regressions. Although firm size and its long-run component are both significant predictors for the future stock returns in univariate regressions, the coefficient on firm size becomes insignificant once controlling for its long-run component. Taken together, our results suggest that for size premium investors, a strategy that is based on its long-run component consistently dominates the traditional size strategy in terms of risk-return tradeoff.

Our decomposition is simple and straightforward. It also sheds lights on several other aspects of the size premium. For instance, the close link between changes in the firm size and stock returns provides a natural explanation for the negative (positive) correlation between momentum (long-term contrarian) profits and size premium. More interestingly, our decomposition uncovers a novel seasonality of the size premium in its exposure to the momentum factor due to the standard Fama and French (1992) timing. In Fama and French (1992), size portfolios are rebalanced at the end of every June, and firm size in June of year t is used to create size portfolios from July of year t to June of year $t+1$. This timing implies that the relative weight of the intermediate-run component decreases monotonically from July of current year to June of next year. If a large portion of the change in market equity is due to stock returns, we expect a similar seasonality in the momentum factor exposure of the Fama and French (1992) size premium. Indeed, in the time series regressions of the long-short Fama and French (1992) size portfolio returns on the market and momentum factors, we find the negative momentum factor loading peaks in the third quarter (-0.17 with a t-statistic of -2.72) and bottoms in the second quarter (0.05 but statistically insignificant). As a comparison, when we repeat the same analysis using the size portfolios sorted by the market value from the previous month, the seasonality in momentum betas disappears.

Our decomposition also provides insights into the January effect, the empirical finding that the size premium is concentrated in January (e.g., Keim (1983)). Two leading explanations for the January effect in the literature are the tax-loss selling hypothesis and institutional investor window dressing hypothesis, both of which posit that shortly before year-end, investors sell stocks that have had losses during the year. Since the stock performance within the past year is closely related to the intermediate-run and short-run components, our decomposition provides a quantitative evaluation of these two hypotheses. Our empirical analyses suggest that although all size components positively contribute to the strong performance of size strategies in January, the intermediate-run and short-run components combined only explain less than 20% of the January effect. In contrast, we find a surprisingly large January effect based on the lagged 5-year and the long-run components, which contribute about 60% and 30%, respectively, to the overall January effect. The results for the long-run component, and especially for the lagged 5-year component, pose a challenge to both leading hypotheses of the January effect, as neither interpretation traces firm performance for such a long horizon. Therefore, our analysis indicates that a large portion of January effect remains puzzling.

Size premium is found to have disappeared since its discovery in early 1980s. In a review paper on anomalies and market efficiency, Schwert (2003) writes that "it seems that the small-firm anomaly has disappeared since the initial publication of the papers that discovered it". Indeed, the average size premium between 1982 and 2002 is only 1.55% per year (t-statistic = 0.39) in our sample. However, we find that the long-run size component remains a significant return predictor during the same sample period. When firms are sorted into decile portfolios based on the long-run component, the long-short portfolio generates an average return of 8.22% per year (t-statistic = 2.16). Therefore, although the traditional size premium indeed disappears between early 1980s and early 2000s, the premium based on the component that is driving the size premium (i.e., the long-run component) was still alive and remains quite strong. Our analysis also suggests that the disappearance of the size premium is primarily due to the bad performance of the lagged 5-year component, which produces an average annual excess return of -2.19% in that sample period.

Lastly, we apply our decomposition to uncover a novel phenomenon among new entrants, which are excluded from our benchmark analyses. We find a positive relation between the size premium and firm age for firms that enter the CRSP dataset within the past 5 years. In the Fama-MacBeth univariate regressions of one-month ahead stock returns on the log firm size, the size coefficient decreases in magnitude from -0.22 for stocks of 4-5 years old to -0.03 (statistically insignificant) for stocks younger than one year old. Our decomposition provides a natural explanation for this interesting pattern. Since young firms do not have a long history, the long-run component, the component that drives the size premium, becomes less important in explaining the cross-sectional size variation, whereas the intermediate-run component that drags down the size premium becomes relatively more important. Our analysis indicates that although the premiums based on each size component remain quantitatively similar and stable across age groups, the change in the size composition with firm age implies a smaller size premium among younger firms.

The paper adds to the large literature on the firm size effect. Beginning with Banz (1981), size premium has been studied extensively in the past three decades. Fama and French (1995) find that size premium can be related to financial distress. Fama and French (1996) use the size premium factor to mimic the underlying risk factor that size premium represents. Studies most closely related to us are Berk (1995, 1996). Berk (1995) argues that size-related regularities should not be regarded as anomalies if size is measured by market value. All else being equal, a firm with higher discount rate has a smaller firm value and higher expected return, so even without specifying the underlying data generating process for stock returns, the negative relation between firm size and future stock returns should always be observed. Berk (1996) uses alternative non-market based measures of firm size, including book value of assets, book value of un-depreciated property, plant, and equipment, total value of annual sales, and total number of employees, but finds no return predictive power. The result of our analyses is consistent with Berk (1995, 1996). Instead of studying non-market based size measures, we find the lagged 5-year market value also doesn't predict stock returns, which suggests that it is not the level of market value, but its changes in recent years that predict stock returns. In addition, because past changes in market value have no direct relation with the current level of book asset, annual sale, or number of employees, the lack of return predictability of these variables that is documented in Berk (1996) should not be surprising.

Our decomposition and its implication for return predictability are motivated by the cross-sectional stock return regularities at various horizons, including long-term contrarian, intermediate-term momentum, short-term reversal, and equity issuance. The objective of this paper is not to explain these patterns.[1] Instead, we take these phenomena as given and study how the composition of these size components quantitatively affects the overall size premium. In terms of the methodology of variable decompositions, our paper is similar to Gerakos and Linnainmaa (2016) who decompose the book-to-market ratio to understand the value premium.

The paper proceeds as follows. Section 2 describes the data. In Section 3, we provide detailed discussions on how to decompose firm size into four components. Section 4 explores the return predictability of each size component. We document that the size premium is mainly driven by the component that captures the change in firm size in prior 13-60 months. In Section 5, we apply our size decomposition to other aspects of the size premium, including a novel seasonality in the momentum factor exposure, the January effect, the disappearance of size premium since early 1980s, and the behaviors of new entrants. Section 6 concludes.

[1]There is large literature on these phenomena in the cross section. For long-term contrarian and value premium, see, for instance, De Bondt and Thaler (1985), De Bondt and Thaler (1987), Lakonishok, Shleifer, and Vishny (1994), Zhang (2005), Lettau and Wachter (2007), Da (2009), Ai, Croce, and Li (2013), Ai and Kiku (2015), Kogan and Papanikolaou (2014). For intermediate-term momentum, see Jegadeesh and Titman (1993), Jegadeesh and Titman (2001), Johnson (2002), Liu and Zhang (2008), Liu and Zhang (2014). For short-term reversal, see Lehmann (1990), Jegadeesh (1990), Jegadeesh and Titman (1995), Nagel (2012), Da, Liu, and Schaumburg (2013). For equity issuance, see, Daniel and Titman (2006), Pontiff and Woodgate (2008), Lyandres, Sun, and Zhang (2008). A small line of research focuses on a joint explanation for these phenomena, especially for intermediate-term momentum, long-term contrarian, and value premium. See, for example, Barberis, Shleifer, and Vishny (1998), Daniel, Hirshleifer, and Subrahmanyam (1998), Hong and Stein (1999), Sagi and Seasholes (2007), Li and Zhang (2017), and Li (2016). See Fama and French (2008) and Nagel (2013) for excellent literature reviews on the cross-sectional stock returns.

2 Data and Summary Statistics

Our data come from several sources. Stock data are from the monthly CRSP database. Accounting data are from Compustat Annually database. The Fama and French factors are from the Fama/French data library. Our sample include NYSE/AMEX/NASDAQ common stocks (with a share code of 10 or 11) with nonmissing market value at the end of June. Our sample period covers 630 months from July 1963 to December 2015. We use size and market value interchangeably unless specified otherwise. In addition, we follow Shumway (1997) to correct for the delisting bias.

Panel A of Table 1 presents the mean (R^{e}), standard deviation (Std), Sharpe ratio (SR), Skewness (Skew), and Kurtosis (Kurt) of the value-weighted excess returns, as well as the CAPM alpha (α^{CAPM}), of size decile portfolios and the spread portfolio that buys small firms and short-sells big firms (S-B). Panel B reports the firm characteristics of these portfolios. Following Fama and French (1992), at the end of each June, we form ten portfolios based on the market equity at June using NYSE breakpoints. The portfolios are then held for one year. Panel A of Table 1 confirms the finding in previous literature that small firms have higher average returns than big firms. The return difference between small and big firms is 3.28% per year, with an annualized Sharpe ratio of 0.2. The long-short portfolio return has large positive skewness and kurtosis, indicating that the size strategy has a small chance of gaining large positive returns. CAPM captures a portion of the size premium. After controlling for the market factor, the average abnormal return of the S-B portfolio becomes only 2.21% per year with a t-statistic of 0.89. The latter finding is consistent with Israel and Moskowitz (2013), who document that CAPM captures a sizable portion of the size premium.

[Insert Table 1 Here]

Panel B reports the characteristics of a typical firm in each size decile. These characteristics include the log market cap (ME), the book-to-market equity ratio (BM), prior 2-12 month stock returns (MOM), and the prior 13-60 month stock return (LTCON). This panel shows that there is a large cross-sectional dispersion in firm size. For instance, the average market value for a typical firm in the small size decile (Decile S) is only 20 million dollars, compared with 6.6 billion dollars in the big size decile (Decile B). Small (big) firms tend to be value (growth) firms, and the book-to-market ratio decreases monotonically from the small decile to the big decile. More importantly, big firms tend to have better past stock performance than small firms. Specifically, the average prior 2-12 month return (MOM) is 0.44% for small firms and 11.54% for big firms, and the average prior 13-60 month return (LTCON) is 6.66% for small firm and 69.72% for big firms.

The pattern above suggests that the size strategy contains a long position in long-term contrarian strategy that buys long-term losers and short-sells long-term winners, and a short position in the momentum strategy that buys momentum winners and short-sells momentum losers. The large profitability of the long-term contrarian and intermediate-term momentum strategies documented in the literature motivates us to decompose firm size into components over corresponding horizons.

3 Decomposing firm size

Based on the results from the previous section, we decompose the log firm size into four components. The first component, ΔME(SR), captures the change in firm size during the past 1-month, consistent with the timing of short-term reversal effect. As we will see later in this paper, although this short-run component may not be crucial for the Fama and French (1992) size strategy, it is an important component for the size strategy based the market value of the previous month, especially for its unique role in January effect. The second component, ΔME(IR), captures the change in firm size during prior 2-12 months, consistent with the timing of intermediate-term momentum effect. The third component, ΔME(LR), captures the changes in firm size during prior 13-60 months, consistent with the timing of the long-term contrarian effect. The last component, ME(lag5), measures the firm size five years ago and captures the extremely persistent component of firm size. For the benchmark analysis, we focus on the Fama and French (1992) timing in creating size portfolios using the market value at the end of June.[2] We also restrict our benchmark sample to only include firms that have non-missing market value in previous June and June 5 years ago. In Section 5.4, we apply our decomposition to understanding the behaviors of these new entrants.

With the Fama and French (1992) timing, the size portfolios are created at the end of June of year t, and the firm size at June and the resulting size decile is assigned to every month from July of year t to June of year $t+1$. In contrast, the short-run, intermediate-run, and long-run windows that correspond to the timing of short-term reversal, momentum, and long-term contrarian strategies are moving with the calendar month even within the twelve months following the rebalancing of size portfolios at the end of June. With this difference taken into consideration, each month, we decompose market equity as follows. For firms in July of year t, the (log) firm size contains all four components, because the size change from the beginning to the end of June of year t represents the short-run component. As a result, our decomposition is based on the following three cross-sectional regressions:

$$ME_{t,6} = a_{0t} + b_{0t} \times ME_{t,5} + \epsilon_{0t} \equiv \widehat{ME}_{0t} + \Delta\text{ME(SR)}_t \tag{1.1}$$
$$\widehat{ME}_{0t} = a_{1t} + b_{1t} \times ME_{t-1,6} + \epsilon_{1t} \equiv \widehat{ME}_{1t} + \Delta\text{ME(IR)}_t \tag{1.2}$$
$$\widehat{ME}_{1t} = a_{2t} + b_{2t} \times ME_{t-5,6} + \epsilon_{2t} \equiv \text{ME(lag5)} + \Delta\text{ME(LR)}_t, \tag{1.3}$$

where $ME_{t,6}$, $ME_{t,5}$, $ME_{t-1,6}$, $ME_{t-5,6}$, are the log market equity (in million dollars) at the end of June in year t, the end of May in year t, the end of June in year $t-1$, and the end of June in year $t-5$, respectively. In the cross-sectional regression equation (1.1), we regress the log size at the end of June of year t on the log size at the end of May of year t, and the residual component $\Delta\text{ME(SR)}_t$ is the short-run component for the size in July of year t. The predicted values $\widehat{ME}_{0t}$ from (1.1) are then used as the dependent variables in equation (1.2) to extract the intermediate-run component. Specifically, we regress $\widehat{ME}_{0t}$ on the log size at the end of June of year $t-1$, and

[2] In some analyses of this paper, we also consider an alternative size strategy that is based on the market value from the end of previous month.

the residual $\Delta\text{ME(IR)}_t$ is the intermediate-run component. Lastly, we regress the predicted values from equation (1.2), $\widehat{ME}_{1t}$, on the log size at the end of June of year $t-5$ in equation (1.3). The residual is the long-run component $\Delta\text{ME(LR)}_t$, whereas the predicted value is our lagged 5-year component ME(lag5).

For all other months from August of year t to June of year $t+1$, there are only three components, because the information about the change in size from the previous month is absent in the size at the end of June.[3] For each of these months, we first regress the log firm size onto the log size twelve months ago. The residual is the intermediate-run component $\Delta\text{ME(IR)}_t$, and the predicted values from the first step are then regressed onto the log size at the end of June of year $t-5$. For example, for size in August of year t, we perform the following decomposition:

$$ME_{t,6} = a_{1t} + b_{1t} \times ME_{t-1,7} + \epsilon_{1t} \equiv \widehat{ME}_{1t} + \Delta\text{ME(IR)}_t \tag{2.1}$$

$$\widehat{ME}_{1t} = a_{2t} + b_{2t} \times ME_{t-5,6} + \epsilon_{2t} \equiv \text{ME(lag5)} + \Delta\text{ME(LR)}_t, \tag{2.2}$$

The residuals from equations (2.1) and (2.2), $\Delta\text{ME(IR)}_t$ and $\Delta\text{ME(LR)}_t$, are the intermediate-run and long-run size components for firms in August year t, whereas the predicted value from equation (2.2), ME(lag5), represents the lagged 5-year component.

The decomposition implies that even though the firm size is a constant (fixed to be the market value at the end of previous June) within the twelve months following size portfolio rebalancing, its components do change from one month to the next. The change in the composition has novel implications on the factor loadings and the performance of new entrants, which we discuss in later sections. We also realize that there are other ways to decompose firm size in a similar spirit. We choose the current procedure because, by construction, it guarantees that: 1) the components add up to the (log) firm size in June; and 2) these components are orthogonal to each other.[4]

Figure 1 shows the time series variation of the relative importance of these size components in explaining the cross-sectional variance of firm size. Each month, we run cross-sectional univariate regressions of log size in June on each of these four components (or three components if not in July) and collect the adjusted R^2. The R^2 for year t is then calculated as the average R^2 from July year $t-1$ to June year t to remove the seasonality of the size components. The figure shows that among all four size components, the lagged 5-year component, ME(lag5), is the most important determinant that explains about 80.5% of the cross-sectional variance of firm size.[5] This result is expected because firm size is highly persistent over time – a big firm today is very likely to remain a big firm five years later. The next important component is the long-run component ΔME(LR), which explains an average of 18.4% of current size. Between late 1980s and early 2000s, we observe

[3]However, this short-run component is present every month for the size measure based on the market value at the end of the previous month.

[4]We repeat our analysis based on alternative decomposition procedures and find very similar results. For instance, when we construct size components by directly taking the difference in log size between the beginning and end of each horizon, instead of running regressions, the main finding is quantitatively similar. These results are available upon requests.

[5]These R^2s are reported in Panel A of Table 7 under Group 0.

an increase in the stock return idiosyncratic volatility, which could drive the increase in the R^2 of the long-run component relative to the lagged 5-year component. The last two components, the intermediate-run ΔME(IR) and short-run ΔME(SR), explain an average of 2.6% and 0.4%, respectively, of the cross-sectional size distribution. Given that the short-run component is only available in July, we ignore it in most discussions on the Fama and French (1992) size strategy.

[Insert Figure 1 Here]

Table 2 reports the firm characteristics of deciles sorted by the size components. Since our benchmark sample now imposes the restriction of non-missing size components, which differs from Table 1, we also report the characteristics of size deciles using this benchmark sample in Panel A. Besides properties of log size (ME), book-to-market (BM), prior 2-12 month returns (MOM), and prior 13-60 month returns (LTCON), we also report the results for size components– ME(lag5), ΔME(IR), and ΔME(LR). Panel A shows that the components display an increasing pattern from the small decile to the big decile size portfolios. For ΔME(IR), it is -0.06 for small firms and 0.05 for big firms. For ΔME(LR), it increases from -0.36 for small firms to 0.43 for big firms. Interestingly, the dispersion in lagged 5-year firm size is only slightly smaller than that in the current firm size, again indicating that firm size is highly persistent. This finding is also consistent with the large explanatory power of the lagged 5-year component for the cross-sectional variation in firm size plotted in Figure 1.

[Insert Table 2 Here]

Panels B, C, and D of Table 2 report the characteristics of the decile portfolios sorted by the size components. Since our decomposition procedure enforces an orthogonal condition among these components, sorting by one component does not create dispersions in other components, as shown in the last three rows of each panel. In Panel B, the intermediate-run component sorts create a large spread in the prior 2-12 month return. Firms with high ΔME(IR) have an average momentum (MOM) of 57.03%, in contrast with -27.19% for firms with low ΔME(IR). In Panel C, firms with high ΔME(LR) have a large long-term contrarian (LTCON) of 271.58%, compared with that among firms with low ΔME(LR) (-45.78%). Therefore, the strategies based on intermediate-run and long-run components are closely related to momentum and long-term contrarian strategies, respectively.[6] Similar to the patterns for the size portfolio in Panel A, Panel D also shows that firms with large lagged 5-year size have higher current size and lower book-to-market ratio than those with small lagged 5-year size.

[6] Changes in firm size can be due to both stock returns and net issuance. The existing studies document both variables predict future stock returns. We could have further decomposed the change in firm size at each horizon into the change in price and change in number of shares outstanding. We choose not to do this for the sake of parsimony.

4 Decomposing the Size Premium

Based on the size decomposition from the previous section, we study the return predictability of these components and quantitatively estimate their contributions to the overall size premium.

Table 3 reports the mean (R^e), standard deviation (Std), and Sharpe ratio (SR) of the value-weighted excess return, as well as CAPM alpha (α^{CAPM}), of decile portfolios sorted by size (Panel A), ΔME(IR) (Panel B), ΔME(LR) (Panel C), and ME(lag5) (Panel D). By restricting non-missing size components, the size premium becomes stronger: the average size premium is 4.93% per year (t-statistic = 2.02) with a Sharpe ratio of 0.3. However, controlling for the market factor reduces the size premium to 4.01% and the corresponding t-statistic becomes 1.68.

[Insert Table 3 Here]

Panel B reports the results for the intermediate-run component ΔME(IR). Stocks with the largest increase in firm size in the intermediate run (Decile Hi) have an average excess return of 10.32% per year (t-statistic = 3.36), compared with only 1.69% for the firms with the largest decrease in size (Decile 1) in the same horizon. The difference in average returns between the two extreme decile portfolios is 8.64%, which is more than 3.6 standard errors from zero. A long-short investment strategy that buys high ΔME(IR) firms and short-sells low ΔME(IR) generates a Sharpe ratio of 0.51. In addition, CAPM fails to explain the strategy returns; controlling for market exposures creates an abnormal return of 9.1% per year with a t-statistic of 3.85. This strategy performance is consistent with the momentum strategy that past intermediate-term winners have higher future returns than intermediate-term losers. Unfortunately, size premium investors do not benefit from its good performance at all, because the size strategy effectively takes a short position in it. In fact, this exposure consistently drags down the profitability of the size strategy over time.

Panel C reports the stock performance of the long-run component ΔME(LR). Opposite to the intermediate-run component, firms with the largest increase in size in the long run underperform firms with the largest decrease in firm size by 7.33% per year (t-statistic = 3.23). The long-short investment strategy based on ΔME(LR) generates a Sharpe ratio of 0.5, and this strong profitability is not captured by CAPM. The CAPM abnormal return is 7.76% with a t-statistic of 3.41. The result for the portfolios sorted by the lagged 5-year component is reported in Panel D. In contrast to the other two components from Panel B and Panel C, the return displays a hump shape from the low ME(lag5) decile to the high ME(lag5) decile. The long-short portfolio generates an insignificant average return of only 1.53% with a Sharpe ratio of 0.1. Controlling for the market factor further reduces this premium to a negative value (-0.59% per year).

The result in Table 3 indicates that among all components of firm size from our decomposition, only the long-run component, ΔME(LR), contributes positively to the overall size premium in an statistically and economically significant way. In other words, the size premium is likely to be mainly driven by this long-run component. To test this conjecture, we perform three different analyses. In the first analysis, we compare the performance of portfolios double sorted by size and its long-run component. In particular, we create 5-by-5 portfolios double-sorted independently by size and ΔME(LR). Panel A.1 of Table 4 reports the conditional size premium within each ΔME(LR) quintile and the average conditional size premium across ΔME(LR) quintiles. Among all ΔME(LR) quintiles, the conditional size premium is only significant in ΔME(LR) quintile 2. In the high ΔME(LR) quintile, the conditional size premium is negative at -1.24% per year, even though it is not statistically significant from zero. The average conditional size premium across ΔME(LR) quintiles is insignificant at 2.65%. The unconditional CAPM further reduces the abnormal conditional size premium to 1.41% per year. In sharp contrast, the conditional ΔME(LR) premium is significant in 4 out of 5 size quintiles. It ranges from 8.92% (t-statistic = 6.37) among small firms to 3.65% (t-statistic = 1.50) among big firms. The average conditional ΔME(LR) premium is 5.42% per year, which is more than 3.85 standard errors from zero. The CAPM alpha for the conditional ΔME(LR) premium is even higher at 6.11% per year, with a t-statistic of 4.33.

[Insert Table 4 Here]

The second test is a linear factor model test between size premium and ΔME(LR) premium. In Panel B.1 of Table 4, we test a two-factor model on size decile portfolios with the market factor and the ΔME(LR) premium factor as the factors. The ΔME(LR) premium factor is calculated as the return difference between the low ΔME(LR) decile and the high ΔME(LR) decile. Compared with the CAPM result from Panel A of Table 3, none of the ΔME(LR) deciles has a significant abnormal return in the two-factor model, and the long-short portfolio (L-H) has an abnormal return of -0.54% per year (t-statistic = -0.29). The reduction in abnormal returns is mainly due to the exposure to the ΔME(LR) factor, which decreases monotonically from 0.49 for small firms to -0.09 for big firms, and the difference is 12.6 standard errors from zero.

When we switch the order and regress the ΔME(LR) decile excess returns on a two-factor model with the market factor and size premium factor as the factors, the result looks quite different. In Panel B.2, we find that despite the strong decreasing pattern of the size factor exposures across ΔME(LR) portfolios, the abnormal return remains large in many portfolios. In addition, the long-short ΔME(LR) portfolio (L-H) has an abnormal return of 5.8% per year, which is more than 3.18 standard errors from zero. Adding the ΔME(IR) premium factor does not alter the result in a significant way (Panel C). If anything, the abnormal return of the long-short ΔME(LR) portfolio becomes even bigger (7.72% per year with a t-statistic = 4.38).

Our third test is Fama-MacBeth regressions. Compared to the value-weighted portfolio approach in the first two tests, Fama-MacBeth regressions put relatively more weights on small firms. Each month, we run a cross-sectional regression of one-month ahead stock returns on log size and its components ΔME(IR), ΔME(LR), and ME(lag5), and the time series average of these coefficients are reported in Table 5. Columns (1)-(4) report the univariate regression results. Consistent with the results from Table 3, we find that although size (ME) is a strong predictor for future stock returns, the coefficient on ME(lag5) is only -0.05 with a t-statistic of -1.43. In contrast, the other two components ΔME(IR) and ΔME(LR) have much stronger predictive power. The coefficient on ΔME(IR) is 0.73, which is more than 4 standard errors from zero. The coefficient on ΔME(LR) is

-0.41 with an even stronger t-statistic of -5.52. These results confirm the relative performance of the corresponding long-short portfolios reported in Table 3.

[Insert Table 5 Here]

Columns (5)-(7) present the horse race results from the Fama-MacBeth regressions using log size and one of its components as the return predictors. In Column (5), when both size and size 5 years ago are included into the same regression, the coefficient on size becomes more significant at -0.32 with a t-statistic of -4.4, compared with the univariate specification. On the other hand, the coefficient on ME(lag5) is now positive. This finding is intuitive from our decomposition: if the lagged 5-year component has no predictive power for stock return and is adding noise to the size premium, controlling this component would make the size premium stronger. We find a similar pattern when we control for the ΔME(IR) component (Column (7)). As we discussed earlier, the size premium strategy contains a short position in the ΔME(IR) premium, so controlling ΔME(IR) would enhance the performance of size strategies. In column (6), the horse race between size and its long-run component ΔME(LR), controlling for ΔME(LR), firm size has little predictive power for returns. Its coefficient decreases from -0.11 in the univariate regression in Column (1) to an insignificant value of -0.04 (t-statistic $= -1.12$). Interestingly, the coefficient on the ΔME(LR) becomes more statistically significant.

One advantage of the Fama-MacBeth regressions over the portfolio approach is that we can quantify the contribution of each size component to the overall size premium. If one component explains an average of X (or $100X$ percent) of the cross-sectional variation in size, and the Fama-MacBeth regression coefficient on this component is Y, its contribution to the coefficient on size in the Fama-MacBeth regression would be $X \times Y$. We use the coefficient estimates from Specifications (2)-(4) and the explanatory power of each component for the cross-sectional variation in firm size in Section 3 to estimate their contributions. For the lagged 5-year component, its percentage contribution is approximately 39.9% ($0.054 \times 80.5\%/0.109$), and this is compared with 69.4% ($0.411 \times 18.4\%/0.109$) for the long-run component and -17.5% ($-0.734 \times 2.6\%/0.109$) for the intermediate-run component.[7] This result indicates that although the 5-year component captures more than 80% of the firm size, it only contributes less than 40% of the size premium.[8] In contrast, the long-run component ΔME(LR) captures only 18% of firm size but contributes almost 70% of the size premium. In untabulated analyses, we find the ΔME(LR) premium is not driven by extremely small and illiquid firms, which is a criticism for the implementability of size strategies (Horowitz, Loughran, and Savin (2000)). For example, when we exclude from our sample firms with market value of less than 5 million dollars, or firms with end-of-June price lower than \$5 per

[7]The fact that the contributions from these components do not exactly add up to one can be due to: 1) we did not include the short-run component in this calculation; 2) there are seasonality in the size components within a year; and 3) the panel data is not balanced; we have more observations in later years than earlier years.

[8]The contribution from the lagged 5-year component is driven by its correlation with the premium based on the long-run component. In an untabulated analysis, we find that the ME(lag5) premium changes sign and becomes positive after controlling for the market factor and a ΔME(LR) premium factor.

share, the long-short ΔME(LR) portfolio still produces an average return of more than 7% per year with a t-statistic greater than 3.

Taken together, our analyses suggest that the size premium is driven by its long-run component ΔME(LR). Once controlling for this long-run component, firm size does not have significant predictive power for future stock returns. These findings are consistent with Berk (1995, 1996). Berk (1995) argues that size-related regularities should not be regarded as anomalies if size is measured by market value. All else being equal, a firm with a higher discount rate has a smaller firm value and higher expected return, so even without specifying the data generating process of stock returns of a firm, the negative relation between firm size and future stock returns should always be observed. Our findings can be consistent with this argument: firms that experience a large decrease in market value in the prior 13-60 months (i.e., in the long run) could have experienced positive shocks to discount rate (either rationally or irrationally). Their realized returns are negative but expected returns increase. However, we find the similar argument does not hold for the horizon of prior 2-12 months. Instead, this intermediate-run change in firm size, which is related to the momentum strategies, positively predicts future stock returns. Berk (1996) studies the size premium using alternative measures of firm size including book value of assets, book value of un-depreciated property, plant, and equipment, total value of annual sales, and total number of employees, but find no return predictive power. Our decomposition provides a natural interpretation for his findings. Particularly, it is not the level of market value, but its recent change, that predicts stock returns. Because the recent changes in market value have no direct relation with the level of these alternative size measures, it is not surprising these variables have no return predictive power. For size premium investors, our findings also suggest that investing in its long-run component is far better than investing in firm size itself from the perspective of risk-return tradeoff.

5 Further Implications

In this section, we explore additional implications of our size decomposition. In Section 5.1, we uncover an interesting seasonality in the momentum factor loading of size portfolios that is due to the Fama and French (1992) timing. In Section 5.2, we use our decomposition to evaluate leading explanations for the January effect in the existing literature. In Section 5.3, we discuss the disappearance of the size premium between early 1980s and early 2000s. We study how our decomposition can be applied to new entrants in Section 5.4.

5.1 Seasonality in momentum beta

Our size decomposition is performed at each month, so these components change from one month to the next. The rolling horizons of the intermediate-run and long-run components indicate that there is a seasonality of the intermediate-run and long-run components in the twelve months following the portfolio rebalancing at the end of each June. For instance, in July of year t, the intermediate-run component is based on the change in log market value from July of year $t-1$ to May of year t.

As time moves forward by one month, the horizon shrinks by one month, and the intermediate-run component in August of year t is based on the change in log market value from August of year $t-1$ to May of year t. In June of year $t+1$, although the firm size still corresponds to the market value at the end of June of year t, its intermediate-run component is only based on the change in size from May of year t to June of year t. Since the intermediate-run change in size is highly correlated with the price momentum, the size premium should also show a seasonality in its momentum factor exposure. Figure 2 presents this seasonality.

[Insert Figure 2 Here]

In Panel A of Figure 2, we plot the average quarterly momentum factor beta of the size premium following portfolio rebalancing at the end of June according to the Fama and French (1992) timing. For each quarter, we estimate the momentum beta by running time series regressions of the monthly long-short size portfolio excess return on the market excess return and the winner-minus-loser portfolio return from momentum deciles. We test this at the quarterly frequency to avoid even higher frequency seasonality such as the January effect. Panel A shows that for Quarter 3 (Q3) from July to September, i.e., the first quarter following portfolio rebalancing, the size premium has a large negative momentum factor loading of -0.174 (t-statistic $= -2.72$). This negative sign is consistent with the short position of the size premium in the momentum strategies. The momentum beta increases monotonically over time, and by Quarter 2 (Q2) from April to June, i.e., the last quarter of the 12-month holding period, it becomes positive at 0.05 but statistically insignificant (t-statistic $= 0.56$). As a comparison, we also estimate the momentum beta across quarters for the size premium based on the market value from the end of the previous month, and the result is plotted in Panel B of Figure 2. With this alternative timing, size deciles are rebalanced every month, and the horizon for the intermediate-run component is constant. Therefore, we do not expect a strong seasonality in momentum beta for this size strategy. Indeed, from Panel B of Figure 2, the size premium has a negative momentum exposure in all quarters: the momentum beta is -0.176 for Q3, -0.114 for Q4, -0.194 for Q1, and -0.13 for Q2. This confirms that the seasonality of momentum beta from Panel A is due to the Fama and French (1992) timing.

5.2 The January effect

The size premium itself is also highly seasonal. Banz (1981) and Reinganum (1983) document that the good stock market performance in January is mainly driven by small stocks. Keim (1983) finds that half of the size premium over the 1963 to 1979 period occurs during January, whereas Blume and Stambaugh (1983) show that all of the size effect occurs in January after adjusting for the "bid-ask spread" bias. Among alternative hypotheses explaining the puzzling January effect, two leading explanations are tax-loss selling hypothesis (see, e.g., Branch (1977), Dyl (1977), Givoly and Ovadia (1983), Starks, Yong, and Zheng (2006)) and the institutional investor window dressing hypothesis (see, for instance, Haugen and Lakonishok (1988), Musto (1997), and Ritter and Chopra (1989)). Both hypotheses argue that investors tend to sell stocks that have had bad performance, and this selling pressure depresses year-end stock prices which rebound in January. In the tax-loss selling hypothesis, investors sell losing stocks in order to lower taxes on net capital gains. In the institutional investor window dressing hypothesis, portfolio managers sell losing stocks to avoid revealing that they have held poorly performing stocks.

To evaluate these two hypotheses, we decompose the size premium in January in the same way as we decompose the overall size premium. Table 6 reports the result. We consider both the Fama and French (1992) size strategy, and the size strategy based on the previous-month-end market value to fully capture the stock performance during the previous year. Panel A presents the average explanatory power of each size component for the cross-sectional variation in firm size. For the Fama and French (1992) size strategy, there are three size components because the short-run component only exists in July. The lagged 5-year, long-run, and intermediate-run components explain 80.6%, 18.6%, and 2.3% of firm size, respectively. The composition looks very similar for the size premium based on the previous-month market value, except for the intermediate-run component, which doubles its explanatory power to 4.7%. Furthermore, the short-run component now shows up and explains about 0.6% of firm size.

[Insert Table 6 Here]

Panel B of Table 6 reports the results from Fama-MacBeth regressions. For the Fama and French (1992) timing, the coefficient on the log firm size in January is -1.62, which is 14.9 times greater than the estimate for all months (Column (1) of Table 5), confirming the January effect in our sample. In addition, these values suggest that the average size premium estimated from the Fama-MacBeth regression for non-January months is negative. The three size components all contribute positively to January effect. The estimated coefficient on ME(lag5) is -1.29, which is 23.9 times greater than that from Table 5, suggesting that even the lagged 5-year component displays a strong January effect. The coefficient on ΔME(LR) is -3.04, which is 7.4 times greater than that from Table 5. This result indicates that although the ΔME(LR) premium is significantly higher in January, it also exists in other months.[9] The coefficient on that ΔME(IR) is -3.27, which is 4.5 times greater than that for all months in magnitude but with an opposite sign. The latter pattern is consistent with the momentum literature that find the average momentum profit to be negative in January (e.g., Jegadeesh and Titman (1993)). The predictive power of the short-run and intermediate-run components for the January return is also consistent with Branch (1977)'s observation that stocks that had negative returns during the prior year also have high returns in January, a finding that motivates the tax-loss selling hypothesis.

In order to quantify the contribution of each size component to the January effect, we multiply the estimated R^2 from Panel A of Table 6 with the corresponding Fama-MacBeth regressions coefficients from Panel B. In the case of the Fama and French (1992) timing, the contributions from the lagged 5-year, long-run, and intermediate-run components are 64.0%, 34.8%, and 4.6%,

[9] Indeed, in an untabulated analysis, we find that the average value-weighted return of the long-short ΔME(LR) portfolio for non-January months is 4.5% with a t-statistic of 2.06.

respectively. In the case of the size strategy based on the market value from previous month, the intermediate-run component becomes more important and these three components contribute to 58.0%, 29.2%, and 11.5%, respectively. In addition, the short-run component appears in January and has a large negative predictability for the January return. The estimated coefficient of ΔME(SR) is -17.22 with a t-statistic of -7.23, indicating that stocks that perform poorly in previous December rebounds strongly in January. This pattern is consistent with the selling pressure on losing stocks in the tax-loss selling hypothesis and institutional investor window dressing hypothesis. Despite its significance, this component only explains about 5.8% of the overall January effect, due to its low explanatory power for firm size from Panel A.

Our quantitative results pose a challenge for both leading hypotheses for the January effect. Both hypotheses rely on investors' behaviors in reaction to the stock performance from the previous year. Our estimates suggest that the contribution from the stock performance in the previous year is well below 20%. The significant coefficient on the long-run component is consistent with De Bondt and Thaler (1985) and Chan (1986), and suggests that investors may wait for years before realizing losses. Still, there is about 60% the January size premium that comes from the lagged 5-year component. Therefore, a large portion of the January effect remains puzzling.[10]

5.3 The disappearance of size premium

It has been documented that the size premium has disappeared since its discovery. For example, Schwert (2003) reports an average CAPM alpha of 0.2% per month with a t-statistic of 0.67 between 1982 and 2002 for the Dimensional Fund Advisors (DFA) US 9-10 Small Company Portfolio, which closely mimics the size strategy described in Banz (1981). Several studies have proposed potential explanations for this disappearance. Hou and van Dijk (2014) argue that it is the large negative profitability shocks that drives the poor performance of small firms after early 1980s. Asness, Frazzini, Israel, Moskowitz, and Pedersen (2015) document that size premium is robust after controlling for quality. Shi and Xu (2015) emphasize the importance of the delisting bias. They document that there is a positive size premium for firms close to be delisted, and once excluding these observations, the size premium reappears. Ahn, Min, and Yoon (2016) find that the size effect is significantly positive at the bottom of the business cycles.

Our decomposition provides an alternative explanation for this phenomenon. Figure 3 plots the cumulative returns of the long-short portfolio based on firm size and its components. The figure shows that the strategy based on the long-run component outperforms the size strategy, whereas the intermediate-run and the lagged 5-year components perform poorly. For the full sample period from July 1963 to December 2015, the cumulative return is about 190.5% for the size premium, which is smaller than 327.3% for the ΔME(LR) premium. On the other hand, the cumulative return is only 14.4% for the lagged 5-year component, and -531.1% for the intermediate-run component.

[10] In untabulated analyses, we extend the horizon back further and find that the premium based on the lag 10-year or even 20-year market values still displays a strong January effect. This pattern is unlikely to be explained by the delayed realization of long-run losses by investors.

The large negative loss for the intermediate-run component is consistent with its strong return predictability documented in Section 4.

[Insert Figure 3 Here]

Narrowing the sample period down to 1982-2002 during which Schwert (2003) documents the disappearance of the size effect, we find that the average size premium is indeed only 1.55% per year, which is about 0.39 standard errors from zero. However, the premium based on the long-run component, ΔME(LR), is 8.22% per year with a t-statistic of 2.16. These results suggest that although the size premium has disappeared between early 1980s and early 2000s, the premium based on the component that drives the size premium (i.e., the long-run component) was still alive and remained quite strong. But what makes the overall size premium disappeared? Our analysis indicates that one main reason is the poor performance of the lagged 5-year component. When focusing on the pattern of the cumulative returns of the size premium and the lagged 5-year size premium between 1982 and early 2000s in Figure 3, we can see a strong comovement between these two time series. More importantly, the average premium of this lagged 5-year component is -2.19% per year during this sample period. This bad performance, together with its large explanatory power for the cross-sectional variation in size (Figure 1), drags down the average size premium.[11,12]

5.4 New entrants

Our analysis in previous sections focuses solely on firms that have non-missing size components from the decomposition. In particular, we exclude firms that entered the CRSP database within the previous 5 years. Fama and French (2004) document that firms that obtain public equity financing expands dramatically in the 1980s and 1990s. The cross section of the profitability of these firms are highly left skewed but their growth rates are highly right skewed. Therefore, the behavior of the size premium among these new entrants could potentially be different from those in our benchmark sample.

In this subsection, we apply our size decomposition to these new entrants. Different from a relatively mature firm in our benchmark sample, the composition of a newly entered young firm depends on the number of years since its entry. For example, a firm that enters the CRSP database 4 years ago has both the long-run and intermediate-run components. In contrast, for a firm that enters 11 months ago, the long-run component is absent. To control this cohort effect, we separate these new entrants into five groups. Group 1 includes stocks younger than 1 year, Group 2 includes

[11] One possible explanation for the negative ME(lag5) premium is the increased idiosyncratic volatility. In untabulated analyses, we notice that the level of common idiosyncratic volatility (CIV) (Herskovic, Kelly, Lustig, and Van Nieuwerburgh (2016)) doubled from early 1980s to early 2000s. Furthermore, we find a strong negative exposure of the ME(lag5) premium to the CIV shock, so unexpected increases in CIV lower the ME(lag5) premium in this sample period. A comprehensive exploration of these historically small firms can be interesting for future studies.

[12] Another explanation for the disappearance of size premium since early 1980s is the bad performance of the newly entered firms. See, for example, Fama and French (2004), Hou and van Dijk (2014). We discuss the implication of our decomposition on the performance of these firms in Section 5.4.