

APPENDIX

CONTENTS

- 1.** Data Sources and Definitions
- 2.** Tests for Mean Reversion
- 3.** Tests for “Granger Causality”
- 4.** Generating “Confidence Intervals” for Future Stock Prices
- 5.** Confidence Intervals for Siegel’s Constant
- 6.** Historic Return Uncertainty Versus Random Returns

1. DATA SOURCES AND DEFINITIONS¹

DEFINITIONS OF q

The definition of q we have used throughout this book has been in line with the ratio published regularly in the corporate balance sheets produced alongside the Federal Reserve's flow of funds accounts.² This is defined, as we noted in Chapter 2, by the following formula:

$$q = \frac{\text{Market Value of Equities}}{\text{Corporate Net Worth}}$$

This formula differs somewhat from the definition originally used by Tobin, which is intended to capture the entire market value of the firm (including the market value of debt), and compares it with tangible assets; thus:

$$\text{"Tobin's } q\text{"} = \frac{\text{Market Value of Equities} + \text{Market Value of Net Debt}}{\text{Tangible Assets}}$$

It should be fairly evident that this definition has the same properties, with respect to the impact of changes in each of its components parts, as Tobin's original definition. Thus, since

$$\text{Net Worth} = \text{Tangible Assets} - \text{Net Debt}$$

corporate debt appears with a negative sign on the bottom of our ratio, while it appears with a positive sign on the top of the ratio in Tobin's definition. In both cases, therefore, a rise in the value

¹The information provided here draws heavily (as does much of the Appendix) on Robertson and Wright (1998) "The Good News and the Bad News About Long-Run Stock Market Returns," University of Cambridge Department of Applied Economics Working Paper No. 9822. Also available from www.econ.cam.ac.uk/faculty/wright.

²Board of Governors of the Federal Reserve System (1999). *Flow of Funds Accounts of the United States: Flows and Outstandings*, Fourth Quarter 1998.

of net debt raises the ratio, other things being equal. Note also that a value of 1 for our definition will also imply a value of 1 in Tobin's definition.

Robertson and Wright (*op. cit.*) show that, in terms of their statistical properties, the two ratios look extremely similar. We prefer our definition, however, for two reasons.

First, the original definition is predicated, implicitly, on a world in which the "Miller-Modigliani" theorem holds. This states that, under certain assumptions, the market value of a firm is unaffected by its method of financing. We would not wish to take this assumption for granted, especially given the link between the use of any valuation criterion, and an implicit rejection of perfectly efficient markets. Note also that the Miller-Modigliani theorem clearly breaks down when there is a positive probability of bankruptcy.

Our second reason is more pragmatic. Since we are using q to value the stock market, rather than a combination of the equity and bond markets, our preferred definition fits more easily into the framework of Part Six, since, as we have seen, it can be easily broken down into two terms, in price, and in the implicit "fundamental" of corporate net worth per share. Thus, if, other things being equal, the stock price falls by $x\%$, our measure of q also falls by $x\%$. Tobin's original definition will, in contrast, fall by a lesser percentage, depending on the relative magnitude of the stocks of bonds and equities.³

Note that, in order to ensure consistency of our definition throughout the sample period, we use a definition of tangible assets that excludes land and residential capital, since this component of total corporate assets is not available before 1945. In practice, however, even in the published figures for later periods, data for land are highly imperfect. Indeed, as we show below, there is a distinct discontinuity in the Fed's methodology before and after 1989, so that our methodology is arguably more robust even after 1945.

³As we shall see in Appendix Section 4, this property is also particularly useful when we come to embed in the loglinear cointegrating vector autoregression that we use to generate confidence intervals, since it implies a simple restriction on the cointegrating vectors, which would be impossible using the original definition.

DATA SOURCES AND CONSTRUCTION FOR q

Capital Stock and Tangible Assets

The fixed capital stock series (the nonresidential capital stock of the nonfinancial corporate sector) comes directly from BEA figures.⁴ This series, which makes up the primary element of corporate net worth, is available throughout the sample period, from 1925 to 1998. It forms the core of our earlier estimates of net worth in the period before 1945. From 1945 we have estimates of inventories from the Federal Reserve.⁵ Before then we assume that inventories move in line with fixed capital over the period 1925 to 1944.

From 1945 onward there is an alternative source of data on tangible assets published in the Federal Reserve Balance Sheets, which at least in principle provide a fuller coverage, since they include estimates of land values. These figures feed into net worth calculations as published by the Federal Reserve. However, quite apart from the clear break that would arise if this series were used after 1945, but not before, there are discontinuities in the Fed's treatment of land in the balance sheets, even after 1945.⁶

From 1956 to 1989, estimates of corporate land values, which are smooth interpolations of five-year observations derived from census data on taxable property values, are added to BEA capital stock figures to arrive at total tangible assets. Before 1956 land values are a pure extrapolation, as a fixed share of tangible assets.

Beyond 1989, the methodology changes radically. The stock of real estate is assumed to evolve according to the following equation:

$$\text{Stock (t)} = \text{Stock (t-1)} \text{ REP (t)/REP (t-1)} + \text{BEA Net Investment in Structures}$$

where REP is an index of corporate real estate prices. Hence the stock of land is implicitly assumed constant from 1989 onward.

⁴Bureau of Economic Analysis (1997). "Revisions to Tangible Assets, 1926–1997," *Survey of Current Business*, May 1997, Department of Commerce, Washington, D.C.

⁵Board of Governors of the Federal Reserve System (*op. cit.*).

⁶We are grateful to Rochelle Antonimowicz and Elizabeth Fogler of the Federal Reserve's Flow of Funds section for advising us on their methodology.

Thus until 1989 land data are independently derived, and simply added to BEA capital stock figures; thereafter they are a residual. Since there is no overlap between the real estate price index and the previous source for land values, it is impossible to check the impact of this change in methodology, but there is a clear impact on the relative magnitudes of the two stocks—indeed, in some years since 1989 implicit land values have turned negative.

Nor should this be entirely surprising. The explicit methodology of the BEA approach is to value the capital stock, including structures, at replacement cost—that is, to evaluate the cost of building structures (suitably depreciated) from new. In contrast, valuing structures using real estate prices implicitly asks what these same structures could be sold for in the open market—this estimate of value is likely to be, and since 1989 has been, considerably more volatile.

Arguably, quite apart from problems with data consistency if land values are used, the build-or-buy arbitrage that was the original rationale for q is more consistent with the BEA approach than with the Fed's approach since 1989. By excluding land values entirely, we clearly impart a (modest) upward bias to the average value of q (offset in practice, as we have previously noted, by other downward biases); but our methodology is otherwise equivalent to assuming land to be a constant fraction of tangible assets, which, as noted, is indeed in effect Fed practice in data between 1945 and 1956, and would, perforce, have been necessary before then.

Corporate Financial Assets and Liabilities

From 1945 onward, we use Federal Reserve figures from the flow of funds.⁷ Before 1945 an estimate is derived indirectly from grossed-up corporate interest payments, available from the national accounts statistics⁸ from 1929 onward and corporate bond yields.⁹ For the period 1925 to 1928, the source of data for net

⁷Board of Governors of the Federal Reserve System (*op. cit.*), Table L102.

⁸Bureau of Economic Analysis (1992) *National Income and Product Accounts of the United States*, Vol. 2, 1929–1958, Department of Commerce, Washington D.C.

⁹Ibbotson Associates (1994), *Stocks, Bonds, Bills and Inflation*, Chicago, Ibbotson Associates.

debt figures is even more restricted, since the data for net interest payments are available only from 1929 onward. Over this period, therefore, we assume, for lack of better information, that net liabilities were unchanged as a proportion of net worth. In effect, therefore, over this period, changes in net worth are driven solely by changes in fixed capital.

Market Value of Equities

From 1945 onward, the numerator of q (the market value of corporate equities) comes direct from the same source as financial assets and liabilities. This however needs to be extrapolated backward before 1945. We have two alternative sources of data that are reassuringly similar in their implications. Using an approach similar to that applied to the construction of net liabilities data, national accounts data on dividend payments can be divided by the reported dividend yield (on the S&P Industrials). Since the coverage of the balance sheet figures is considerably wider than that of the S&P Index, this approach is not perfect. Instead, we use an alternative data source provided to us by Jim Bianco of Arbor Trading in Chicago, to whom we are extremely grateful. This series is for the market capitalization of all listed corporations in the USA, since the start of 1926. While we use the Federal Reserve figures where available, we assume that the market value of equities moves in line with the Arbor Trading series before 1945.¹⁰

Construction of q for 1900 to 1925

For this period we have considerably more limited data, and simply use proportional extrapolation of q back in time using market value and net worth data derived from data constructed by Professor Olivier Blanchard and colleagues, to whom we are most grateful.¹¹

¹⁰Chart 2 of Robertson and Wright (*op. cit.*) shows that the grossing-up method yields very similar results.

¹¹These data were originally used to produce q estimates (on Tobin's definition) for a paper (which we also refer to in Chapter 28) by Olivier Blanchard, Chanyong Rhee, and Lawrence Summers, "The Stock Market, Profit, and Investment" (*Quarterly Journal of Economics*, 1990). For a comparison of the two data sets, which are reassuringly similar when they overlap, see Robertson and Wright (*op. cit.*).

OTHER DATA DEFINITIONS AND SOURCES

Dividend Yields and Price-Earnings Multiples

We use series for the S&P Industrials (for greater consistency with our q figures, which relate to the nonfinancial sector only) where available (from 1926 onward). Before 1926 we use historical series kindly supplied by Professor Jeremy Siegel.

Stock, Bond, and Bill Returns

Our figures are from the data set used by Jeremy Siegel in *Stocks for the Long Run*. We are, again, most grateful to Professor Siegel for allowing us to use these figures.

2. TESTS FOR MEAN REVERSION

As we noted in our discussion of this issue in the main text, a lot can be learned about whether a given series mean-reverts simply by visual inspection, coupled with understanding of the mechanisms that may or may not bring about mean reversion. However, more formal statistical tests can give an additional element of objectivity. We present here the most commonly used test, the “ADF” Test, which essentially tests whether changes in a given series are systematically related to the level of the series itself.¹² Thus, the test identifies whether high values of the series in one year are correlated with falls in the series in the following year, and vice versa. The hypothesis being tested is that no such relationship exists. Mean reversion is therefore inferred, in the rather convoluted way that statisticians like, by *rejecting* the hypothesis that the series does *not* mean-revert.¹³ Table A1 shows the results of this test for q , the dividend yield, and the P/E multiple, over two samples, 1900 to 1998, and 1900 to 1991.

¹²The original test procedure was described in David Dickey and Wayne Fuller, “Distribution of the Estimators for Autoregressive Time Series with a Unit Root,” *Journal of the American Statistical Association* 74(366), Part 1 (June 1979), pp. 427–31. This test, the “DF” test, is now more generally carried out using techniques which allow for more complex dynamic models; hence the acronym stands for “Augmented Dickey-Fuller” test.

¹³A weakness of this approach is that the test does not discriminate very well between mean-reverting series that do so slowly, and series that do not mean-revert at all.

TABLE A1

"ADF" Tests

Indicator	"ADF" Test Statistics	
	1900–1998	1900–1991
q	-2.88**+	-3.32**++
Dividend yield	-2.22**	-3.84****
P/E	-3.16****	-3.75****

**(*) Rejects hypothesis at 95% (90%) level, assuming asymptotic normality.

**(*) Rejects hypothesis at 95% (90%) level, assuming Dickey-Fuller distribution.

Two different samples are used because, as the table shows, the impact of the last few years on the tests has been quite marked. In a sample period truncated to include the period in which all three indicators imply that the market was overvalued, all three reject the hypothesis very strongly—hence indicating mean reversion. In the full sample, only the P/E multiple entirely conclusively rejects; q rejects strongly if the test statistic is asymptotically normally distributed, but is just on the margin of rejection at the 95% level in the full sample, using the more conservative Dickey-Fuller distribution.¹⁴ In effect this means that the test in this form implies a probability of almost precisely 5% that q does *not* mean-revert.¹⁵ Robertson and Wright (*op. cit.*) show, however, that if q really did not mean-revert, the probability that it would appear to do so, *and* would have predictive power for stock prices to the extent it does (see next section), is even lower, at around 3%.

The test results for the dividend yield are interesting. Over the truncated sample it appears to mean-revert most strongly of all, whereas including more recent years very much weakens this result. This is not too surprising, given the extent of the fall in recent years (see Chart 21.1). Inclusion of data for the nineteenth century also very much weakens the evidence for mean reversion; our table, however, restricts itself to data for the twentieth century to give comparability among the three different indicators.

¹⁴For discussion of the issue of which distribution is more appropriate, see Robertson and Wright (*op. cit.*). The Dickey-Fuller, although it is the correct distribution under the null, has the tendency not to reject a "unit root" too often—that is, it has "low power."

¹⁵The 95% critical value for this sample size is 2.89.

3. TESTS FOR “GRANGER CAUSALITY”

In this section we discuss evidence underlying Test No. 4: that a valuation indicator should have some (but not too much) predictive power for stock prices and returns. Statistical tests of this property are tests of “Granger Causality”: in essence, whether the indicator in question is a “leading indicator” of prices.

The tests are carried out in a vector autoregressive (VAR) framework, which captures the nature of historic correlations among different series. Thus, for example, for our purposes, statistical techniques allow us to capture the extent to which stock prices can be predicted from their own history, and from the history of the “fundamental” in question; and vice versa for the fundamental itself. The three fundamentals examined are net worth per share (for q), real dividends per share (for the dividend yield), and real earnings per share (for the P/E multiple).

We present two different sets of results, both of which are complementary. In the first, summarized in Table A2, we test the statistical significance of each of the three fundamentals in predicting stock prices, and vice versa. Again, the test is rather convoluted: The hypothesis is that a given fundamental does *not* help in predicting stock prices (and vice versa). Hence rejection of the hypothesis implies that there is some predictive power.

TABLE A2

Tests of “Granger Causality”¹⁶

Fundamental	Probability That Fundamental Does <i>Not</i> Have Predictive Power for Real Stock Prices	Probability That Real Stock Prices Do <i>Not</i> Have Predictive Power for Fundamental
Real net worth per share (fundamental for q)	2.0%	9.8%
Real dividends per share (fundamental for dividend yield)	15.1%	0.00%
Real earnings per share (fundamental for P/E multiple)	32.4%	0.02%

¹⁶Carried out in an unrestricted second-order vector autoregression over the sample 1902 to 1998, testing for block deletion of all lags of the variable in question.

The table shows a very sharp distinction on this test between q and the other two indicators. Although in all three cases there is some evidence of mean reversion, the *mechanism* by which mean reversion takes place is distinctly different. Thus, in the case of both dividends per share and earnings per share, there is a reasonably high probability that they have *no* predictive power at all for stock prices. The probability of this for q is, however, extremely low, at only 2%. Conversely, stock prices help predict both earnings and dividends, but have only rather weak predictive power for real net worth.

A limitation of the above sets of results is that no restriction is imposed on the nature of the relationship between the fundamental and stock prices. In principle, for example, we might find a “fundamental” which appeared to help predict stock prices, but in a perverse way. A rise in the fundamental might, for example, imply a *fall* in the stock price; or a rise in the fundamental of 1% might imply a rise in the stock price of less, or more, than 1%. Such a relationship would not make sense in economic terms, and would therefore on these grounds be suspect. To combat this problem, therefore, all three fundamentals were examined within a similar framework, but one in which the predictive power was constrained to come from the relevant valuation *ratio*.¹⁷

Table A3 presents results of two tests. The first column shows whether the data accept the restriction that the fundamental should have an impact only via the relevant ratio.¹⁸ In the case of q this is very easily accepted, whereas it is very far from being accepted in the case of both dividends and earnings.¹⁹ In the second column, the results shown are very much analogous to those in Table A2: here a low probability implies (given the convoluted way that the test is set up) that there *is* predictive power from the relevant ratio. Again, there is a sharp division between q and the other two candidates for the fundamental. q has predictive power; the other two ratios in effect have none. The fact that the ratios nonetheless have a tendency to mean-revert arises (as Table A2 shows) from the reactions of the relevant fundamentals, not from the stock price.

¹⁷In econometric terms, the tests were carried out in a cointegrating vector autoregression, where the single cointegrating relationship was the (log of) the valuation ratio in question.

¹⁸Formally this is a test of the restriction that the cointegrating vector is (1, -1 in logs).

¹⁹Conventionally you would reject this restriction only if the probability was less than 5%.

TABLE A3

Granger Causality Tests Using Valuation Ratios

Fundamental	Probability That Fundamental Only Has Impact Via Relevant Valuation Ratio	Probability That Ratio Has No Predictive Power for Stock Prices
Real net worth per share (fundamental for q)	46.7%	1.4%
Real dividends per share (fundamental for dividend yield)	0.01%	43.8%
Real earnings per share (fundamental for P/E multiple)	2.1%	88.6%

We can therefore conclude that, while in terms of mean reversion q does not obviously dominate the other indicators (most notably the P/E multiple), an examination of the nature of the process of mean reversion singles q out much more clearly.

4. GENERATING “CONFIDENCE INTERVALS” FOR FUTURE STOCK PRICES

In Charts 15.1 and 15.3 we showed the range of uncertainty for q and for future stock prices in a “90% confidence interval.” The methodology involved is rather technical, so we restrict ourselves here to a summary discussion.²⁰

The case of Chart 15.1 is relatively more simple. In predicting future values of q we have some help, in the near term, from the fact that we know the *current* value. While “noise” in the share price makes q hard to predict even one period ahead, we also know that the range of values it might take will not be too far from the current level. However, the further ahead we look, the less help the current value will be in helping us to forecast; hence the band of uncertainty inevitably gets wider. Once we look far enough out

²⁰Those wishing for a more precise, but of necessity more technical discussion of the procedure will find it fully described in Robertson and Wright (*op. cit.*). The approach used to generate Chart 15.3 follows the methodology outlined there, with a dataset extended backward to run from 1900, and forward to 1998.

(in practice, around ten years), the current value of q is essentially no help to us at all, so all we can do is assess the range of uncertainty that would face us if we knew nothing at all about the recent past. This problem is compounded by “parameter uncertainty,” which arises from the fact that we can only estimate the true mean value of q .

The case of the stock price, in Chart 15.3, is more complicated. Just as with q , knowing the current stock price helps us in the short term. But, unlike q , the stock price has no tendency to mean-revert, so we cannot get any help at all from past average values of the stock price. What *does* help us, however, is the fact that q mean-reverts, since this implies that the stock price cannot get too far away from net worth per share. This makes prediction easier, but also more complicated in technical terms, since if we want to predict stock prices, we need also to predict net worth. Hence, we need again to use the “vector autoregressive” framework already discussed in the previous section,²¹ which captures historic correlations between stock prices and net worth, and essentially extrapolates these into the future. Since neither stock prices nor net worth mean-reverts, the uncertainty about their future values grows forever, with parameter uncertainty becoming ever more important, the further out we look. However, the *rate* at which that uncertainty grows is very much reduced by exploiting the predictive power of q , as compared with the range of uncertainty if we simply had to predict stock prices from their own history: This is, in essence, a statistical representation of the good news about q .²²

5. CONFIDENCE INTERVALS FOR SIEGEL'S CONSTANT

If stock returns were random, it would be relatively straightforward to derive a confidence interval for the average real stock return (which we refer to as Siegel's constant) over a given historical period. The formula for the standard deviation of the mean

²¹Strictly speaking, it is a “cointegrating vector autoregression,” with q as a cointegrating relationship.

²²Robertson and Wright (*op. cit.*) includes a chart showing that the confidence interval from the VAR is very much narrower than from a random walk. The relative magnitude of these two intervals very much resembles that shown in Chart 17.3.

of a random variable is given by $\sigma / T^{1/2}$, where σ is the sample standard deviation, and T is the number of observations. However, this formula is appropriate only if the variable in question is random; and the evidence of q 's predictive power is, as we have seen, that returns are not quite random. One way to combat this is to look at longer-period returns, since these are more likely to be independent of each other; on the other hand, in order to ensure independence, only nonoverlapping returns should be used. This cuts down on the number of observations, T . Table A4 shows that using returns over different horizons can nonetheless result in some reduction in the degree of uncertainty about the true value of the average return. The confidence interval used in Chart 17.1 is derived from the standard deviation of the mean (nonoverlapping) 30-year return. The table makes clear that if we assumed that annual returns were random, the range of uncertainty as to the true value of Siegel's constant would be distinctly greater.

TABLE A4Confidence Intervals for Siegel's Constant²³

	95% Lower Bound	95% Upper Bound
From annual data	0.0426	0.0924
From nonoverlapping 10-year returns	0.045	0.086
From nonoverlapping 20-year returns	0.038	0.088
From nonoverlapping 30-year returns	0.049	0.077

6. HISTORIC RETURN UNCERTAINTY VERSUS RANDOM RETURNS

In Chart 17.3 we showed the historic range of variability in cumulative real returns on investments in stocks, depending on the investment horizon. The longer the investment horizon, the

²³The mean (log) real return on stocks.

greater the degree of variability, for the same reasons as those described in the previous section in relation to forecast uncertainty.

The range of variability for actual historic returns was derived by simply finding the range of values between which actual historic returns lay 90% of the time. The range of variability that would have prevailed if stock returns had in fact been random was derived by computer simulation, on the assumption that returns were independently normally distributed, with mean and sample variance as in the historic data. Since computing power is cheap these days, we produced 10,000 different alternative histories of 200 years, and derived the properties of the long horizons returns in each of them. Since, as was made clear in the previous section, we do not know the true value of the mean return, each of these 10,000 alternative histories had a different mean return, with the value taken being drawn in turn from a normal distribution, with appropriate variance, given the observed uncertainty about the true mean. The range shown is simply the values between which returns at different horizons lay 90% of the time.