(numerically) inverse-Fourier transforming the spectral density of returns. To find the univariate, invertible moving average representation from the spectral density, you have to factor the spectral density $S_{rr}(z) = a(z)a(z^{-1})$, where $a(z)$ is a polynomial with roots outside the unit circle, $a(z) = (1 - \gamma_1 z)(1 - \gamma_2 z) \cdots \gamma_i < 1$. Then, since $a(L)$ is invertible, $r_t = a(L)\varepsilon_t$ $\sigma_\varepsilon^2 = 1$ is the univariate representation of the return process.

The autocorrelations and spectral densities are directly revealing: a string of small negative autocorrelations or a dip in the spectral density near frequency zero correspond to mean-reversion; positive autocorrelations or a spectral density higher at frequency zero than elsewhere corresponds to momentum.

Multivariate Mean-Reversion

> I calculate the responses to multivariate rather than univariate shocks. In a multivariate system you can isolate expected return shocks and dividend growth shocks. The price response to expected return shocks is *entirely* stationary.

We are left with a troubling set of facts: high price/dividend ratios strongly forecast low returns, yet high past returns do not seem to forecast low subsequent returns. Surely, there must be some sense in which "high prices" forecast lower subsequent returns?

The resolution must involve dividends (or earnings, book value, or a similar divisor for prices). A price rise with no change in dividends results in lower subsequent returns. A price rise that comes with a dividend rise does not result in lower subsequent returns. A high return combines dividend news and price/dividend news, and so obscures the lower expected return message. In a more time-series language, instead of looking at the response to a univariate return shock, a return that was unanticipated based on lagged returns, let us look at the responses to multivariate shocks, a return that was unanticipated based on lagged returns and dividends.

This is easy to do in our simple VAR. We can simulate (20.17)–(20.20) forward and trace the responses to a dividend growth shock and an expected return (d/p ratio) shock. Figures 20.4 and 20.5 present the results of this calculation. (Cochrane [1994a] presents a corresponding calculation using an unrestricted VAR, and the results are very similar.)

Start with Figure 20.4. The negative expected return shock raises prices and the p-d ratio immediately. We can identify such a shock in the data as a return shock with no contemporaneous movement in dividends. The p-d ratio then reverts to its mean. Dividends are not forecastable, so they show

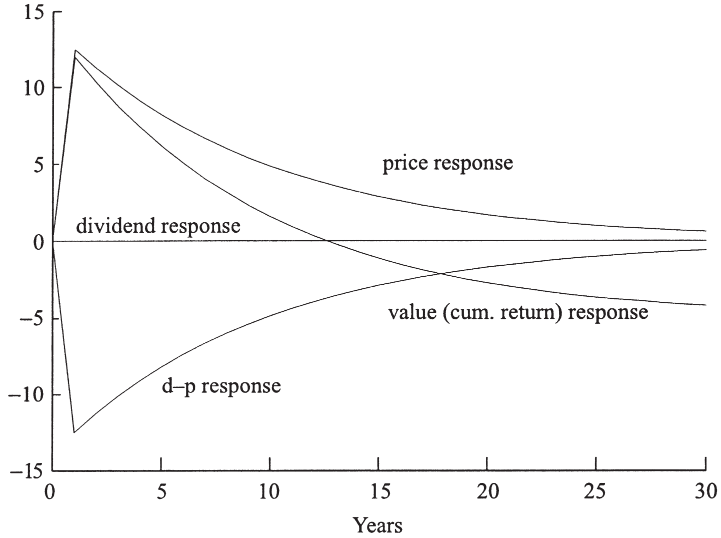

Figure 20.4. *Responses to a one-standard-deviation (1.7%) negative expected return shock in the simple VAR.*

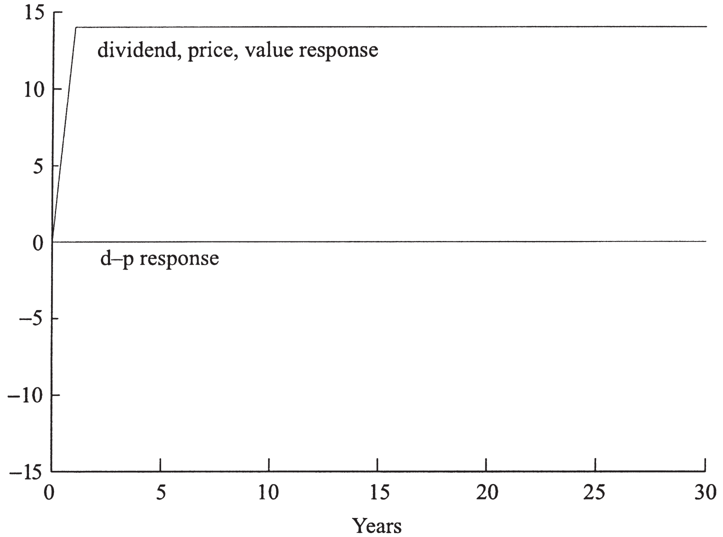

Figure 20.5. *Responses to a one-standard-deviation (14%) dividend growth shock in the simple VAR.*

no immediate or eventual response to the expected return shock. Prices show a long and *complete* reversion back to the level of dividends. This shock looks a lot like a negative yield shock to bonds: such a shock raises prices now so that bonds end up at the same maturity value despite a smaller expected return.

The cumulative return "mean-reverts" even more than prices. For given prices, dividends are now smaller (smaller d-p) so returns deviate from their mean by more than price growth. The cumulative return ends up *below* its previously expected value. Compare this value response to the univariate value response, which we calculated above, and ends up at about 0.8 of its initial response.

The dividend shock shown in Figure 20.5 raises prices and cumulative returns immediately and proportionally to dividends, so the price/dividend ratio does not change. Expected returns or the discount rate, reflected in any slope of the value line, do not change. If the world were i.i.d., this is the only kind of shock we would see, and dividend/price ratios would always be constant.

Figures 20.4 and 20.5 plot the responses to "typical," one-standard-deviation shocks. Thus you can see that actual returns are typically about half dividend shocks and half expected return shocks. That is why returns alone are a poor indicator of expected returns.

In sum, at last we can see some rather dramatic "mean-reversion." Good past returns by themselves are not a reliable signal of lower subsequent returns, because they contain substantial dividend growth noise. Good returns that do not include good dividends isolate an expected return shock. This does signal low subsequent returns. It sets off a *completely* transitory variation in prices.

Cointegration and Short- vs. Long-Run Volatility

> If $d-p$, Δp, and Δd are stationary, then the long-run variance of Δd and Δp must be the same, long-run movements in d and p must be perfectly correlated, and d and p must end up in the same place after any shock. Thus, the patterns of predictability, volatility, and univariate, and multivariate mean-reversion really all just stem from these facts, the persistence of $d-p$ and the near-unforecastability of Δd.

You might think that the facts about predictability depend on the exact structure of the VAR, including parameter estimates. In fact, most of what we have learned about predictability and mean-reversion comes down to a few facts: the dividend/price ratio, returns, and dividend growth are all stationary; dividend growth is not (or is at best weakly) forecastable, and dividend growth varies less than returns.

These facts imply that the dividend and price responses to each shock are eventually equal in Figures 20.4 and 20.5. If $d - p$, Δp, and Δd are stationary, then d and p must end up in the same place following a shock. The responses of a stationary variable $(d - p)$ must die out. If dividends are not forecastable, then it must be the case that prices do all the adjustment following a price shock that does not affect dividends.

Stationary $d - p$, Δp, and Δd also implies that the variance of long-horizon Δp must equal the variance of long-horizon Δd:

$$\lim_{k\to\infty} \frac{1}{k}\mathrm{var}(p_{t+k} - p_t) = \lim_{k\to\infty} \frac{1}{k}\mathrm{var}(d_{t+k} - d_t), \tag{20.41}$$

and the correlation of long-run price and dividend growth must approach one. These facts follow from the fact that the variance ratio of a stationary variable must approach zero, and $d - p$ is stationary. Intuitively, long-run price growth cannot be more volatile than long-run dividend growth, or the long-run $p - d$ ratio would not be stationary.

Now, if dividend growth is not forecastable, its long-run volatility is the same as its short-run volatility—its variance ratio is one. Short-run price growth is more volatile than short-run dividend growth, so we conclude that prices must be mean-reverting; their variance ratio must be below one.

Quantitatively, this observation supports the magnitude of univariate mean-reversion that we have found so far. Dividend growth has a short run, and thus long-run, standard deviation of about 10% per year, while returns and prices have a standard deviation of about 15% per year. Thus, prices must have a long-run variance ratio of about $(2/3)^2$, or a long-run response to univariate shocks of 2/3.

The work of Lettau and Ludvigson (2001b) suggests that we may get much more dramatic implications by including consumption data. The ratio of stock market values to consumption should also be stationary; if wealth were to explode people would surely consume more and vice versa. The ratio of dividends to aggregate consumption should also be stationary. Consumption growth seems independent at all horizons, and consumption growth is very stable, with roughly 1% annual standard deviation. For example, Lettau and Ludvigson (2001b) find that none of the variables that forecast returns in Table 20.2—including $d - p$ and a consumption to wealth ratio—forecast consumption growth at any horizon.

These facts suggest that aggregate dividends *are* forecastable, by the consumption/dividend ratio, and strongly so—the long-run volatility of aggregate dividend growth must be the 1% volatility of consumption growth, not the 10% short-run volatility of dividend growth.

These facts also suggest that almost all of the 15% or more variation in annual stock market wealth must be transitory—the long-run volatility of stock market value must be no more than the 1% consumption growth volatility!

However, total market value is not the same thing as price, price is not the same thing as cumulated return, and aggregate dividends are not the same thing as the dividend concept we have used so far (dividends paid to a dollar investment with dividends consumed), or dividends paid to a dollar investment with dividends reinvested. Lettau and Ludvigson show that the consumption/wealth ratio does forecast returns, but no one has yet worked out the mean-reversion implications of this fact.

My statements about the implications of stationary $d - p, \Delta d, \Delta p, r$ are developed in detail in Cochrane (1994b). They are special cases of the representation theorems for *cointegrated* variables developed by Engle and Granger (1987). A regression of a difference like Δp on a ratio like $p - d$ is called the *error-correction* representation of a cointegrated system. Error-correction regressions have subtly and dramatically changed almost all empirical work in finance and macroeconomics. The vast majority of the successful return forecasting regressions in this section, both time-series and cross-section, are error-correction regressions of one sort or another. Corporate finance is being redone with regressions of growth rates on ratios, as is macroeconomic forecasting. For example, the consumption/GDP ratio is a powerful forecaster of GDP growth.

Bonds

> The expectations model of the term structure works well on average and for horizons of four years or greater. At the one-year horizon, however, a forward rate one percentage point higher than the spot rate seems entirely to indicate a one percentage point higher expected excess return rather than a one percentage point rise in future interest rates.

The venerable expectations model of the term structure specifies that long-term bond yields are equal to the average of expected future short-term bond yields. As with the CAPM and random walk, the expectations model was the workhorse of empirical finance for a generation. And as with those other views, a new round of research has significantly modified the traditional view.

Table 20.8 calculates the average return on bonds of different maturities. The expectations hypothesis seems to do pretty well. Average holding period returns do not seem very different across bond maturities, despite

Table 20.8. *Average continuously compounded* (log) *one-year holding period returns on zero-coupon bonds of varying maturity*

Maturity N	Avg. Return $E\left(\mathrm{hpr}_{t+1}^{(N)}\right)$	Std. error	Std. dev. $\sigma\left(\mathrm{hpr}_{t+1}^{(N)}\right)$
1	5.83	0.42	2.83
2	6.15	0.54	3.65
3	6.40	0.69	4.66
4	6.40	0.85	5.71
5	6.36	0.98	6.58

Annual data from CRSP 1953–1997.

the increasing standard deviation of bond returns as maturity rises. The small increase in returns for long-term bonds, equivalent to a slight average upward slope in the yield curve, is usually excused as a small "liquidity premium." In fact, the curious pattern in Table 20.8 is that bonds do *not* share the high Sharpe ratios of stocks. Whatever factors account for the volatility of bond returns, they seem to have very small risk prices.

Table 20.8 is again a tip of an iceberg of an illustrious career for the expectations hypothesis. Especially in times of great inflation and exchange rate instability, the expectations hypothesis does a very good first-order job.

However, one can ask a more subtle question. Perhaps there are *times* when long-term bonds can be forecast to do better, and other times when short-term bonds are expected to do better. If the times even out, the unconditional averages in Table 20.8 will show no pattern. Equivalently, we might want to check whether a forward rate that is *unusually high* forecasts an unusual *increase* in spot rates.

Table 20.9 gets at these issues, updating Fama and Bliss' (1987) classic regression tests. (Campbell and Shiller [1991] and Campbell [1995] make the same point with regressions of yield changes on yield spreads.) The left-hand panel presents a regression of the change in yields on the forward-spot spread. The expectations hypothesis predicts a coefficient of 1.0, since the forward rate should equal the expected future spot rate. At a one-year horizon we see instead coefficients near zero and a negative adjusted R^2. Forward rates one year out seem to have no predictive power whatsoever for changes in the spot rate one year from now. On the other hand, by four years out, we see coefficients within one standard error of 1.0. Thus, the expectations hypothesis seems to do poorly at short (1 year) horizons, but much better at longer horizons and on average (Table 20.8).

If the yield expression of the expectations hypothesis does not work at one-year horizons, then the expected return expression of the expectations

Table 20.9. Forecasts based on forward-spot spread

	Change in yields $y^{(1)}_{t+N} - y^{(1)}_t = a + b\left(f_t^{(N\to N+1)} - y_t^{(1)}\right) + \varepsilon_{t+N}$					Holding period returns $\mathrm{hpr}^{(N+1)}_{t+1} - y^{(1)}_t = a + b\left(f_t^{(N\to N+1)} - y_t^{(1)}\right) + \varepsilon_{t+1}$				
N	a	$\sigma(a)$	b	$\sigma(b)$	$\overline{R}^2$	a	$\sigma(a)$	b	$\sigma(b)$	$\overline{R}^2$
1	0.1	0.3	−0.10	0.36	−0.02	−0.1	0.3	1.10	0.36	0.16
2	−0.01	0.4	0.37	0.33	0.005	−0.5	0.5	1.46	0.44	0.19
3	−0.04	0.5	0.41	0.33	0.013	−0.4	0.8	1.30	0.54	0.10
4	−0.3	0.5	0.77	0.31	0.11	−0.5	1.0	1.31	0.63	0.07

OLS regressions 1953–1997 annual data. Yields and returns in annual percentages. The left-hand panel runs the change in the one-year yield on the forward-spot spread. The right-hand panel runs the one-period excess return on the forward-spot spread.

hypothesis must not hold either—one must be able to forecast one-year bond returns. To check this fact, the right-hand panel of Table 20.9 runs regressions of the one-year excess return on long-term bonds on the forward-spot spread. Here, the expectations hypothesis predicts a coefficient of zero: no signal (including the forward-spot spread) should be able to tell you that this is a particularly good time for long bonds versus short bonds. As you can see, the coefficients in the right-hand panel of Table 20.9 are all about 1.0. A high forward rate does not indicate that interest rates will be higher one year from now; it seems entirely to indicate that you will earn that much more holding long-term bonds. (The coefficients in yield and return regressions are linked. For example in the first row $1.10 + (-0.10) = 1.0$, and this holds as an identity. Fama and Bliss call them "complementary regressions.")

Figures 20.6 and 20.7 provide a pictorial version of the results in Table 20.9. Suppose that the yield curve is upward sloping as in the top panel. What does this mean? A naive investor might think this pattern indicates that long-term bonds give a higher return than short-term bonds. The expectations hypothesis denies this conclusion. If the expectations hypothesis were true, the forward rates plotted against maturity in the top panel would translate one-for-one to the forecast of future spot rates in the bottom panel, as plotted in the line marked "Expectations model." Rises in future short rates should lower bond prices, cutting off the one-period advantage of long-term bonds. The rising short rates would directly raise the multiyear advantage of short-term bonds.

We can calculate the actual forecast of future spot rates from the estimates in the left-hand panel of Table 20.9, and these are given by the line marked "Estimates" in Figure 20.7. The essence of the phenomenon is sluggish adjustment of the short rates. The short rates do eventually rise to meet

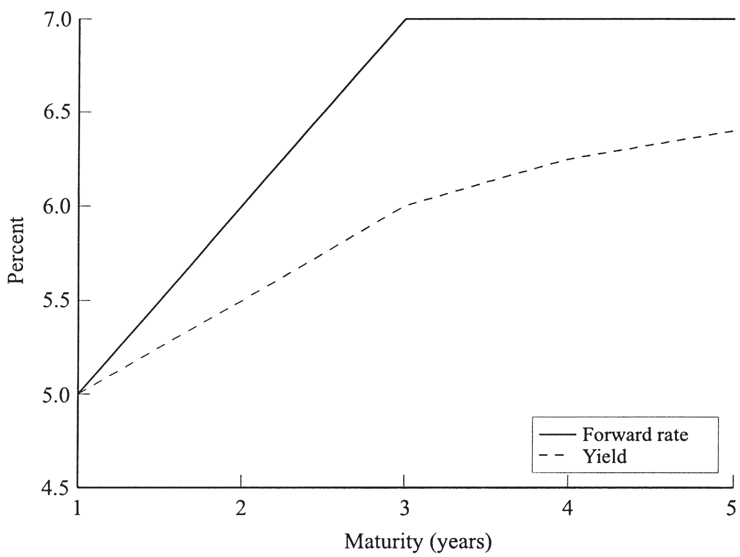

Figure 20.6. *If the current yield curve is as plotted here. . . .*

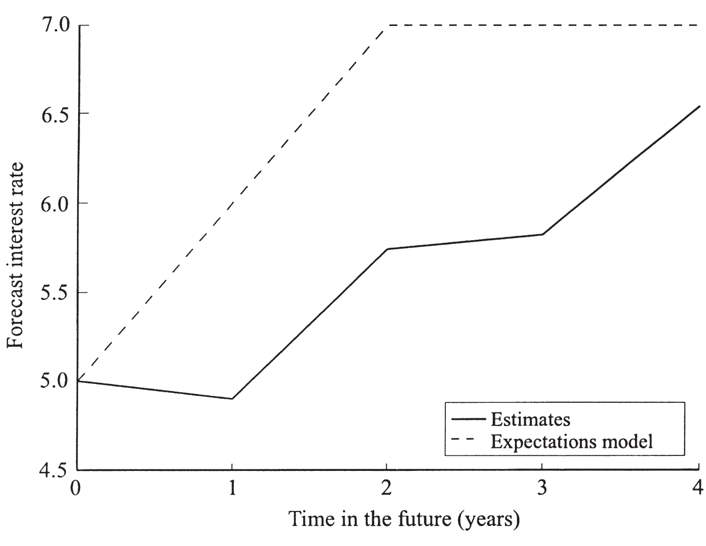

Figure 20.7. *. . . this is the forecast of future one-year interest rates. The dashed line gives the forecast from the expectations hypothesis. The solid line is constructed from the estimates in Table 20.8.*

the forward rate forecasts, but not as quickly as the forward rates predict that they should.

As dividend growth should be forecastable so that returns are not forecastable, short-term yields *should* be forecastable so that returns are *not* forecastable. In fact, yield changes are almost unforecastable at a one-year horizon, so, mechanically, bond returns are. We see this directly in the first row of the left-hand panel of Table 20.9 for the one-period yield. It is an implication of the right-hand panel as well. If

$$\mathrm{hpr}_{t+1}^{(N+1)} - y_t^{(1)} = 0 + 1\left(f_t^{(N\rightarrow N+1)} - y_t^{(1)}\right) + \varepsilon_{t+1}, \tag{20.42}$$

then, writing out the definition of holding period return and forward rate,

$$\begin{aligned} p_{t+1}^{(N)} - p_t^{(N+1)} + p_t^{(1)} &= 0 + 1\left(p_t^{(N)} - p_t^{(N+1)} + p_t^{(1)}\right) + \varepsilon_{t+1}, \\ p_{t+1}^{(N)} &= 0 + 1\left(p_t^{(N)}\right) + \varepsilon_{t+1}, \\ y_{t+1}^{(N)} &= 0 + 1\left(y_t^{(N)}\right) - \varepsilon_{t+1}/N. \end{aligned} \tag{20.43}$$

A coefficient of 1.0 in (20.42) is equivalent to yields or bond prices that follow random walks: yield changes that are completely unpredictable.

Of course yields are stationary and not totally unpredictable. However, they move slowly. Thus, yield changes are very unpredictable at short horizons but much more predictable at long horizons. That is why the coefficients in the right-hand panel of Table 20.9 build with horizon. If we did holding period return regressions at longer horizons, they would gradually approach the expectations hypothesis result.

The roughly 1.0 coefficients in the right-hand panel of Table 20.9 mean that a one percentage point increase in forward rate translates into a one percentage point increase in expected return. It seems that old fallacy of confusing bond yields with their expected returns also contains a grain of truth, at least for the first year. However, the one-for-one variation of expected returns with forward rates does not imply a one-for-one variation of expected returns with yield spreads. Forward rates are related to the slope of the yield curve,

$$\begin{aligned} f_t^{(N\rightarrow N+1)} - y_t^{(1)} &= p_t^{(N)} - p_t^{(N+1)} - y_t^{(1)} \\ &= -Ny_t^{(N)} + (N+1)y_t^{(N+1)} - y_t^{(1)} \\ &= N\left(y_t^{(N+1)} - y_t^{(N)}\right) + \left(y_t^{(N+1)} - y_t^{(1)}\right). \end{aligned}$$

Thus, the forward-spot spread varies more than the yield spread, so regression coefficients of holding period yields on yield spreads give coefficients greater than one. Expected returns move *more* than one-for-one with yield spreads. Campbell (1995) reports coefficients of excess returns on yield spreads that rise from one at a two-month horizon to 5 at a five-year horizon.

The facts are analogous to the dividend/price regression. There, dividends should be forecastable so that returns are not forecastable. But dividends were essentially unforecastable and the dividend yield was persistent. These facts implied that a one percentage point change in dividend yield implied a 3–5 percentage point change in expected excess returns.

Of course, there is risk: the R^2 are all about 0.1–0.2, about the same values as the R^2 from the dividend/price regression at a one-year horizon, so this strategy will often go wrong. Still, 0.1–0.2 is not zero, so the strategy does pay off more often than not, in violation of the expectations hypothesis. Furthermore, the forward-spot spread is a slow-moving variable, typically reversing sign once per business cycle. Thus, the R^2 build with horizon as with the D/P regression, peaking in the 30% range (Fama and French [1989]). (Also, Cochrane and Piazzesi (2003) extend these regressions to more maturities on the right-hand side, and find R^2 as high as 44%.)

The fact that the regressions in Table 20.9 run the *change* in yield and the excess *return* on the forward-spot *spread* is very important. The overall level of interest rates moves up and down a great deal but slowly over time. Thus, if you run $y_{t+j}^{(N)} = a + bf_t^{(N+1)} + \varepsilon_{t+N}$, you will get a coefficient b almost exactly equal to 1.0 and a stupendous R^2, seemingly a stunning validation of the expectations hypothesis. If you run a regression of tomorrow's temperature on today's temperature, the regression coefficient will be near 1.0 with a huge R^2 as well, since the temperature varies a lot over the year. But today's temperature is not a useful temperature forecast. To measure a temperature forecast we want to know if the forecast can predict the *change* in temperature. Is (forecast − today's temperature) a good measure of (tomorrow's temperature − today's temperature)? Table 20.9 runs this regression.

The decomposition in (20.43) warns us of one of several econometric traps in this kind of regression. Notice that two of the three right-hand variables are the same. Thus any measurement error in $p_t^{(N+1)}$ and $p_t^{(1)}$ will induce a spurious common movement in left- and right-hand variables. In addition, since the variables are a triple difference, the difference may eliminate a common signal and isolate measurement error or noise. There are pure measurement errors in the bond data, and we seldom observe pure discount bonds of the exactly desired maturity. In addition, various liquidity and microstructure effects can influence the yields of particular bonds in ways that are not exploitable for typical investors.

As an example of what this sort of "measurement error" can do, suppose all bond yields are 5%, but there is one "error" in the two-period bond price

Table 20.10. Numerical example of the effect of measurement error in yields on yield regressions

t	0	1	2	3
$p_t^{(1)}$	−5	−5	−5	−5
$p_t^{(2)}$	−10	−15	−10	−10
$p_t^{(3)}$	−15	−15	−15	−15
$y_t^{(i)}, i \neq 2$	5	5	5	5
$y_t^{(2)}$	5	7.5	5	5
$f_t^{(1\to 2)}$	5	10	5	5
$f_t^{(1\to 2)} - y_t^{(1)}$	0	5	0	0
$\text{hpr}_t^{(2\to 1)} - y_t^{(1)}$	0	0	5	0

at time 1: rather than being −10 it is −15. Table 20.10 tracks the effects of this error. It implies a blip of the one-year forward rate in year one, and then a blip in the return from holding this bond from year one to year two. The price and forward rate "error" automatically turns into a subsequent return when the "error" is corrected. If the price is real, of course, this is just the kind of event we want the regression to tell us about—the forward rate did not correspond to a change in future spot rate, so there was a large return; it was a price that was "out of line" and if you could trade on it, you should. But the regression will also pounce on measurement error in prices and indicate spuriously forecastable returns.

Foreign Exchange

> The expectations model works well on average. However, a foreign interest rate one percentage point higher than its usual differential with the U.S. rate (equivalently, a one percentage point higher forward-spot spread) seems to indicate even more than one percentage point expected excess return; a further appreciation of the foreign currency.

Suppose interest rates are higher in Germany than in the United States. Does this mean that one can earn more money by investing in German bonds? There are several reasons that the answer might be no. First, of course, is default risk. While not a big problem for German government bonds, Russia and other governments have defaulted on bonds in the past and may do so again. Second, and more important, is the risk of devaluation. If German interest rates are 10%, U.S. interest rates are 5%, but the Euro falls 5% relative to the dollar during the year, you make no more money holding the German bonds despite their attractive interest rate. Since lots of investors are making this calculation, it is natural to conclude that an interest rate differential across countries on bonds of similar credit risk should reveal an expectation of currency devaluation. The logic is exactly the same as the "expectations hypothesis" in the term structure. Initially attractive yield or interest rate differentials should be met by an offsetting event so that you make no more money on average in one country or another, or in one maturity versus another. As with bonds, the expectations hypothesis is slightly different from pure risk neutrality since the expectation of the log is not the log of the expectation. Again, the size of the phenomena we study usually swamps this distinction.

As with the expectations hypothesis in the term structure, the expected depreciation view ruled for many years, and still constitutes an important first-order understanding of interest rate differentials and exchange rates. For example, interest rates in east Asian currencies were very high on the eve of the currency collapses of 1997, and many banks were making tidy sums borrowing at 5% in dollars to lend at 20% in local currencies. This situation should lead one to suspect that traders expect a 15% devaluation, or a small chance of a larger devaluation. That is, in this case, exactly what happened. Many observers and policy analysts who ought to know better often attribute high nominal interest rates in troubled countries to "tight monetary policy" that is "strangling the economy" to "defend the currency." In fact, one's first-order guess should be that such high nominal rates reflect a large probability of devaluation—loose monetary and fiscal policy—and that they correspond to much lower real rates.

Still, does a 5% interest rate differential correspond to an exactly 5% expected depreciation, or does some of it still represent a high expected return from holding debt in that country's currency? Furthermore, while expected depreciation is clearly a large part of the story for high interest rates in countries that have constant high inflation or that may suffer spectacular depreciation of a pegged exchange rate, how does the story work for, say, the United States versus Germany, where inflation rates diverge little, yet exchange rates fluctuate a surprisingly large amount?

Table 20.11 presents the facts, as summarized by Hodrick (forthcoming) and Engel (1996). The first row of Table 20.11 presents the average appreciation of the dollar against the indicated currency over the sample period. The dollar fell against DM, yen, and Swiss Franc, but appreciated against the pound. The second row gives the average interest rate differential—the amount by which the foreign interest rate exceeds the U.S. interest rate. According to the expectations hypothesis, these two numbers should

Table 20.11.

	DM	£	¥	SF
Mean appreciation	−1.8	3.6	−5.0	−3.0
Mean interest differential	−3.9	2.1	−3.7	−5.9
b, 1975–1989	−3.1	−2.0	−2.1	−2.6
R^2	.026	.033	.034	.033
b, 1976–1996	−0.7	−1.8	−2.4	−1.3

The first row gives the average appreciation of the dollar against the indicated currency, in percent per year. The second row gives the average interest differential—foreign interest rate less domestic interest rate, measured as the forward premium—the 30-day forward rate less the spot exchange rate. The third through fifth rows give the coefficients and R^2 in a regression of exchange rate changes on the interest differential = forward premium,

$$s_{t+1} - s_t = a + b(f_t - s_t) + \varepsilon_{t+1} = a + b\big(r_t^f - r_t^d\big) + \varepsilon_{t+1},$$

where s = log spot exchange rate, f = forward rate, r^f = foreign interest rate, r^d = domestic interest rate.
Source: Hodrick (forthcoming) and Engel (1996).

be equal—interest rates should be higher in countries whose currencies depreciate against the dollar.

The second row shows roughly the right pattern. Countries with steady long-term inflation have steadily higher interest rates, and steady depreciation. The numbers in the first and second rows are not exactly the same, but exchange rates are notoriously volatile so these averages are not well measured. Hodrick shows that the difference between the first and second rows is not statistically different from zero. This fact is exactly analogous to the fact of Table 20.8 that the expectations hypothesis works well "on average" for U.S. bonds and is the tip of an iceberg of empirical successes for the expectations hypothesis as applied to currencies.

As in the case of bonds, however, we can also ask whether times of *temporarily* higher or lower interest rate differentials correspond to times of above and below average depreciation as they should. The third and fifth rows of Table 20.11 address this question, updating Hansen and Hodrick's (1980) and Fama's (1984) regression tests. The number here should be +1.0 in each case—an extra percentage point interest differential should correspond to one extra percentage point expected depreciation. As you can see, we have exactly the opposite pattern: a higher than usual interest rate abroad seems to lead, if anything, to further *ap*preciation. It seems that the old fallacy of confusing interest rate differentials across countries with expected returns, forgetting about depreciation, also contains a grain of truth. This is the "forward discount puzzle," and takes its place alongside the forecastability of stock and bond returns. Of course it has produced a similar avalanche of academic work dissecting whether it is really there and if so, why. Hodrick (1987), Engel (1996), and Lewis (1995) provide surveys.

The R^2 shown in Table 20.11 are quite low. However, like D/P, the interest differential is a slow-moving forecasting variable, so the return forecast R^2 build with horizon. Bekaert and Hodrick (1992) report that the R^2 rise to the 30–40% range at six-month horizons and then decline again. Still, taking advantage of this predictability, like the bond strategies described above, is quite risky.

The puzzle does *not* say that one earns more by holding bonds from countries with higher interest rates than others. Average inflation, depreciation, and interest rate differentials line up as they should. If you just buy bonds with high interest rates, you end up with debt from Turkey and Brazil, whose currencies inflate and depreciate steadily. The puzzle *does* say that one earns more by holding bonds from countries whose interest rates are *higher than usual* relative to U.S. interest rates.

However, the fact that the "usual" rate of depreciation and "usual" interest differential varies through time, if they are well-defined concepts at all, may diminish if not eliminate the out-of-sample performance of trading rules based on these regressions.

The foreign exchange regressions offer a particularly clear-cut case in which "Peso problems" can skew forecasting regressions. Lewis (1995) credits Milton Friedman for coining the term to explain why Mexican interest rates were persistently higher than U.S. interest rates in the early 1970s even though the currency had been pegged for more than a decade. A small probability of a huge devaluation each period can correspond to a substantial interest differential. You will see long stretches of data in which the expectations hypothesis seems not to be satisfied, because the collapse does not occur in sample. The Peso subsequently collapsed, giving substantial weight to this view. Since then, "Peso problems" have become a generic term for the effects of small probabilities of large events on empirical work. Rietz (1988) offered a Peso problem explanation for the equity premium that investors are afraid of another great depression which has not happened in sample. Selling out-of-the-money put options and earthquake insurance in Los Angeles are similar strategies whose average returns in a sample will be severely affected by rare events that may not be seen in surprisingly long samples.

20.2 The Cross Section: CAPM and Multifactor Models

Having studied how average returns change over time, now we study how average returns change across different stocks or portfolios.

The CAPM

> For a generation, portfolios with high average returns also had high betas. I illustrate with the size-based portfolios.

The first tests of the CAPM such as Lintner (1965b) were not a great success. If you plot or regress the average returns versus betas of individual stocks, you find a lot of dispersion, and the slope of the line is much too flat—it does not go through any plausible risk-free rate.

Miller and Scholes (1972) diagnosed the problem. Betas are measured with error, and measurement error in right-hand variables biases down regression coefficients. Fama and MacBeth (1973) and Black, Jensen, and Scholes (1972) addressed the problem by grouping stocks into portfolios. Portfolio betas are better measured because the portfolio has lower residual variance. Also, individual stock betas vary over time as the size, leverage, and risks of the business change. Portfolio betas may be more stable over time, and hence easier to measure accurately.

There is a second reason for portfolios. Individual stock returns are so volatile that you cannot reject the hypothesis that all average returns are the same. $\sigma/\sqrt{T}$ is big when $\sigma = 40$–80%. By grouping stocks into portfolios based on some characteristic (other than firm name) related to average returns, you reduce the portfolio variance and thus make it possible to see average return deferences. Finally, I think much of the attachment to portfolios comes from a desire to more closely mimic what actual investors would do rather than simply form a statistical test.

Fama and MacBeth and Black, Jensen, and Scholes formed their portfolios on betas. They found individual stock betas, formed stocks into portfolios based on their betas, and then estimated the portfolio's beta in the following period. More recently, size, book/market, industry, and many other characteristics have been used to form portfolios.

Ever since, the business of testing asset pricing models has been conducted in a simple loop:

1. Find a characteristic that you think is associated with average returns. Sort stocks into portfolios based on the characteristic, and check that there is a difference in average returns between portfolios. Worry here about measurement, survival bias, fishing bias, and all the other things that can ruin a pretty picture out of sample.
2. Compute betas for the portfolios, and check whether the average return spread is accounted for by the spread in betas.
3. If not, you have an anomaly. Consider multiple betas.

This is the traditional procedure, but econometrics textbooks urge you not to group data in this way. They urge you to use the characteristic as an instrument for the poorly measured right-hand variable instead. It is an interesting and unexplored idea whether this instrumental variables approach could fruitfully bring us back to the examination of individual securities rather than portfolios.

The CAPM proved stunningly successful in empirical work. Time after time, every strategy or characteristic that seemed to give high average returns turned out to also have high betas. Strategies that one might have thought gave high average returns (such as holding very volatile stocks) turned out not to have high average returns when they did not have high betas.

To give some sense of that empirical work, Figure 20.8 presents a typical evaluation of the Capital Asset Pricing Model. (Chapter 15 presented some of the methodological issues surrounding this evaluation; here I focus on the facts.) I examine 10 portfolios of NYSE stocks sorted by size (total market capitalization), along with a portfolio of corporate bonds and long-term government bonds. As the spread along the vertical axis shows, there is a sizeable spread in average returns between large stocks (lower average return) and small stocks (higher average return), and also a large spread between stocks and bonds. The figure plots these average returns against market betas. You can see how the CAPM prediction fits: portfolios with higher average returns have higher betas. In particular, notice that the

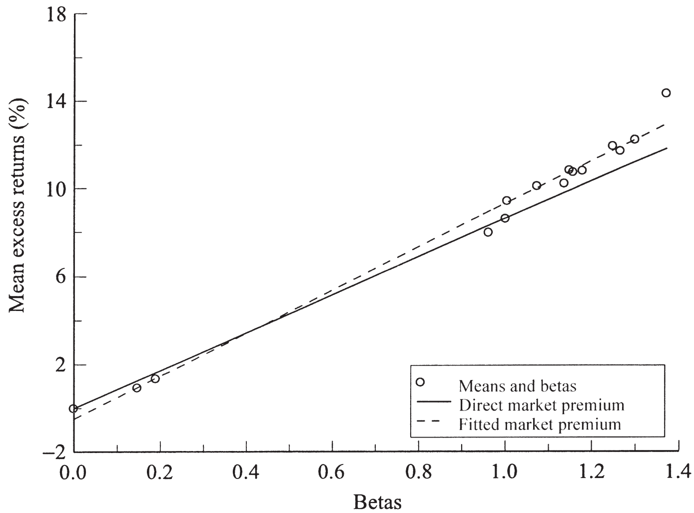

Figure 20.8. *The CAPM. Average returns vs. betas on the NYSE value-weighted portfolio for 10 size-sorted stock portfolios, government bonds, and corporate bonds, 1947–1996. The solid line draws the CAPM prediction by fitting the market proxy and treasury bill rates exactly (a time-series test). The dashed line draws the CAPM prediction by fitting an OLS cross-sectional regression to the displayed data points. The small-firm portfolios are at the top right. The points far down and to the left are the government bond and treasury bill returns.*

long-term and corporate bonds have mean returns in line with their low betas, despite their standard deviations nearly as high as those of stocks. Comparing this graph with the similar Figure 2.4 of the consumption-based model, the CAPM fits very well.

In fact, Figure 20.8 captures one of the first significant *failures* of the CAPM. The smallest firms (the far right portfolio) seem to earn an average return a few percent too high given their betas. This is the celebrated "small-firm effect" (Banz [1981]). Would that all failed economic theories worked so well! It is also atypical in that the estimated market line through the stock portfolios is steeper than predicted, while measurement error in betas usually means that the estimated market line is too flat.

Fama–French 3 Factors

> Book market sorted portfolios show a large variation in average returns that is unrelated to market betas. The Fama and French three-factor model successfully explains the average returns of the 25 size and book market sorted portfolios with a three-factor model, consisting of the market, a small minus big (SMB) portfolio, and a high minus low (HML) portfolio.

In retrospect, it is surprising that the CAPM worked so well for so long. The assumptions on which it is built are very stylized and simplified. Asset pricing theory recognized at least since Merton (1971, 1973a) the theoretical possibility, indeed probability, that we should need factors, state variables, or sources of priced risk beyond movements in the market portfolio in order to explain why some average returns are higher than others.

The Fama–French model is one of the most popular multifactor models that now dominate empirical research. Fama and French (1993) presents the model; Fama and French (1996) gives an excellent summary, and also shows how the three-factor model performs in evaluating expected return puzzles beyond the size and value effects that motivated it.

"Value" stocks have market values that are small relative to the accountant's book value. (Book values essentially track past investment expenditures. Book value is a better divisor for individual-firm price than are dividends or earnings, which can be negative.) This category of stocks has given large average returns. "Growth" stocks are the opposite of value and have had low average returns. Since low prices relative to dividends, earnings, or book value forecast *times* when the market return will be high, it is natural to suppose that these same signals forecast categories of stocks that will do well; the "value effect" is the cross-sectional analogy to price-ratio predictability in the time series.

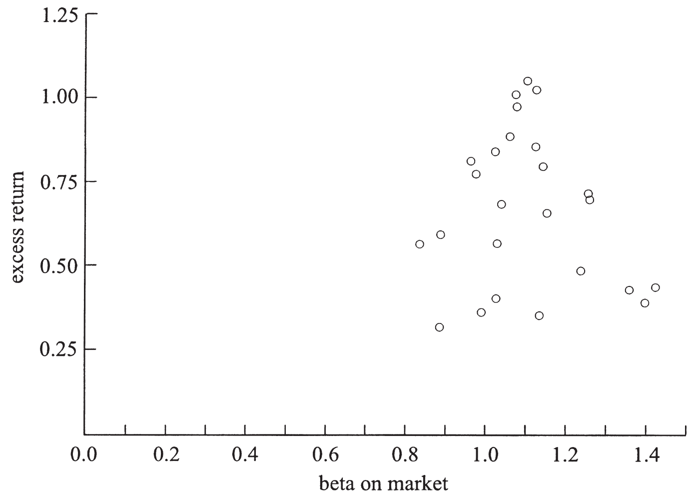

Figure 20.9. *Average returns vs. market beta for 25 stock portfolios sorted on the basis of size and book/market ratio.*

High average returns are consistent with the CAPM, if these categories of stocks have high sensitivities to the market, high betas. However, small and especially value stocks seem to have abnormally high returns even after accounting for market beta. Conversely, "growth" stocks seem to do systematically worse than their CAPM betas suggest. Figure 20.9 shows this value-size puzzle. It is just like Figure 20.8, except that the stocks are sorted into portfolios based on size and book/market ratio[1] rather than size alone. As you can see, the highest portfolios have *three* times the average excess return of the lowest portfolios, and this variation has nothing at all to do with market betas.

Figures 20.10 and 20.11 dig a little deeper to diagnose the problem, by connecting portfolios that have different size within the same book/market category, and different book/market within size category. As you can see, variation in size produces a variation in average returns that is positively related to variation in market betas, as we had in Figure 20.9. Variation in book/market ratio produces a variation in average return that is *negatively* related to market beta. Because of this value effect, the CAPM is a disaster when confronted with these portfolios. (Since the size effect disappeared in 1980, it is likely that almost the whole story can be told with book/market effects alone.)

To explain these patterns in average returns, Fama and French advocate a multifactor model with the market return, the return of small less big stocks

[1] I thank Gene Fama for providing me with these data.

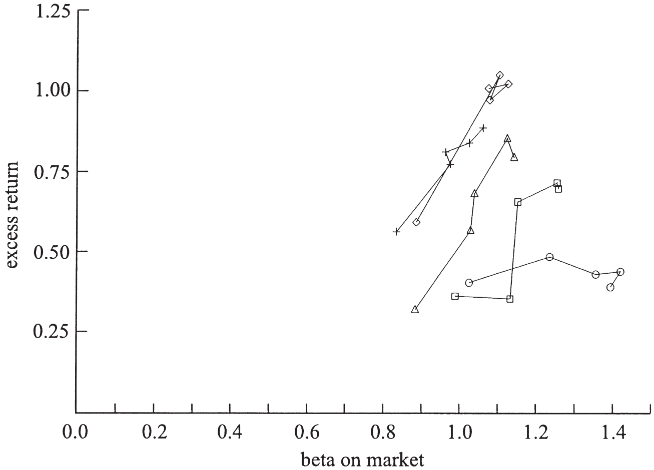

Figure 20.10. *Average excess returns vs. market beta. Lines connect portfolios with different size category within book market categories.*

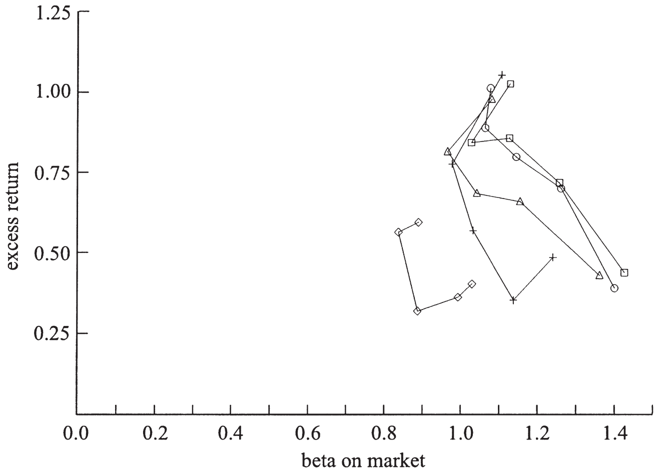

Figure 20.11. *Average excess returns vs. market beta. Lines connect portfolios with different book market categories within size categories.*

(SMB) and the return of high book/market minus low book/market stocks (HML) as three factors. They show that variation in average returns of the 25 size and book/market portfolios can be explained by varying loadings (betas) on the latter two factors. (All their portfolios have betas close to one on the market portfolio. Thus, market beta explains the average return difference between stocks and bonds, but not across categories of stocks.)

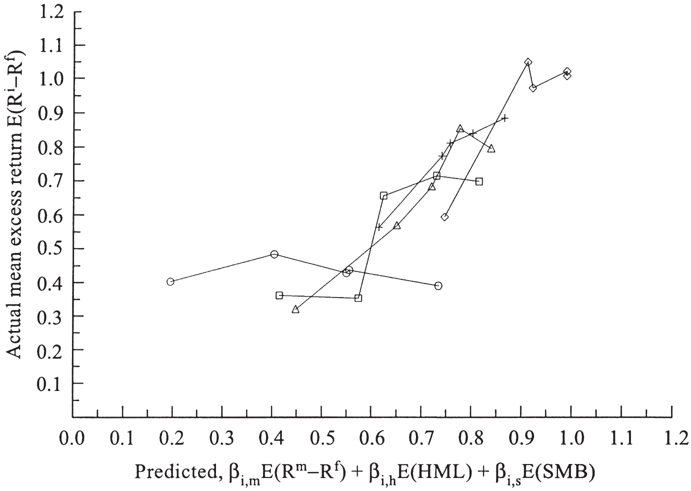

Figure 20.12. *Average excess return vs. prediction of the Fama–French three-factor model. Lines connect portfolios of different size categories within book/market category.*

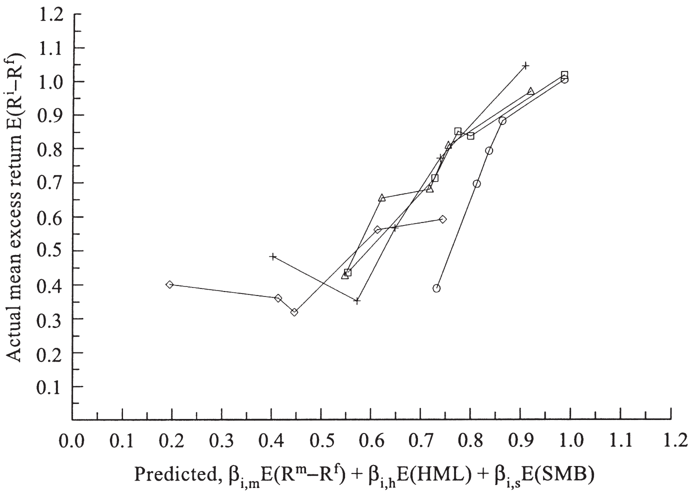

Figure 20.13. *Average excess return vs. prediction of the Fama–French three-factor model. Lines connect portfolios of different book market category within the same size category.*

Figures 20.12 and 20.13 illustrate Fama and French's results. The vertical axis is still the average return of the 25 size and book/market portfolios. Now, the horizontal axis is the predicted values from the Fama–French three-factor model. The points should all lie on a 45° line if the model is correct. The points lie much closer to this prediction than they do in Figures 20.10

and 20.11. The worst fit is for the growth stocks (lowest line, Figure 20.12), for which there is little variation in average return despite large variation in size beta as one moves from small to large firms.

What Are the Size and Value Factors?

> What are the macroeconomic risks for which the Fama–French factors are proxies or mimicking portfolios? There are hints of some sort of "distress" or "recession" factor at work.
>
> A central part of the Fama–French model is the fact that these three pricing factors also explain a large part of the ex post variation in the 25 portfolios—the R^2 in time-series regressions are very high. In this sense, one can regard it as an APT rather than a macroeconomic factor model.
>
> The Fama–French model is not a tautology, despite the fact that factors and test portfolios are based on the same set of characteristics.

We would like to understand the real, macroeconomic, aggregate, nondiversifiable risk that is proxied by the returns of the HML and SMB portfolios. Why are investors so concerned about holding stocks that do badly at the times that the HML (value less growth) and SMB (small-cap less large-cap) portfolios do badly, even though the market does not fall?

Fama and French (1996) note that the typical "value" firm has a price that has been driven down from a long string of bad news, and is now in or near financial distress. Stocks bought on the verge of bankruptcy have come back more often than not, which generates the high average returns of this strategy. This observation suggests a natural interpretation of the value premium: If a credit crunch, liquidity crunch, flight to quality, or similar financial event comes along, stocks in financial distress will do *very* badly, and this is just the sort of time at which one particularly does not want to hear that one's stocks have become worthless! (One cannot count the "distress" of the individual firm as a "risk factor." Such distress is idiosyncratic and can be diversified away. Only aggregate events that average investors care about can result in a risk premium.) Unfortunately, empirical support for this theory is weak, since the HML portfolio does not covary strongly with other measures of aggregate financial distress. Still, it is a possible and not totally tested interpretation, since we have so few events of actual systematic financial stress in recent history.

Heaton and Lucas' (1997b) results add to this story for the value effect. They note that the typical stockholder is the proprietor of a small, privately held business. Such an investor's income is of course particularly sensitive to the kinds of financial events that cause distress among small firms and distressed value firms. Such an investor would therefore demand a substantial premium to hold value stocks, and might hold growth stocks despite a low premium.

Lettau and Ludvigson (2001a) (also discussed in the next section) document that HML has a *time-varying* beta on both the market return and on consumption. Thus, though there is very little unconditional correlation between HML and recession measures, Lettau and Ludvigson document that HML is sensitive to bad news *in bad times*.

Liew and Vassalou (1999) is an example of current attempts to link value and small-firm returns to macroeconomic events. They find that in many countries counterparts to HML and SMB contain information above and beyond that in the market return for forecasting GDP growth. For example, they report a regression

$$GDP_{t\to t+1} = a + 0.065\ MKT_{t-1\to t} + 0.058\ HML_{t-1\to t} + \varepsilon_{t+1}.$$

$GDP_{t\to t+1}$ denotes the next year's GDP growth and MKT, HML denote the previous year's return on the market index and HML portfolio. Thus, a 10% HML return reflects a 1/2 percentage point rise in the GDP forecast.

On the other hand, one can ignore Fama and French's motivation and regard the model as an arbitrage pricing theory. If the returns of the 25 size and book/market portfolios could be perfectly replicated by the returns of the three-factor portfolios—if the R^2 in the time-series regressions were 100%—then the multifactor model would have to hold exactly, in order to preclude arbitrage opportunities. In fact the R^2 of Fama and French's time-series regressions are all in the 90–95% range, so extremely high Sharpe ratios for the residuals would have to be invoked for the model *not* to fit well. Equivalently, given the average returns and the failure of the CAPM to explain those returns, there would be near-arbitrage opportunities if value and small stocks did not move together in the way described by the Fama–French model.

One way to assess whether the three factors proxy for real macroeconomic risks is by checking whether the multifactor model prices additional portfolios, and especially portfolios that do *not* have high R^2 values. Fama and French (1996) extend their analysis in this direction: They find that the SMB and HML portfolios comfortably explain strategies based on alternative price multiples (P/E, B/M), strategies based on five-year sales growth (this is especially interesting since it is the only strategy that does not form portfolios based on price variables), and the tendency of five-year returns to reverse. All of these strategies are not explained by CAPM betas. However, they all also produce portfolios with high R^2 values in a time-series regression on the HML and SMB portfolios! This is good and bad news. It might mean that the model is a good APT: that the size and book/market characteristics describe the major sources of priced variation in all stocks. On the other hand, it might mean that these extra sorts just have not identified other sources of priced variation in stock returns. (Fama and French also find that HML and SMB do not explain "momentum," despite large R^2 values. More on momentum later.)

One's first reaction may be that explaining portfolios sorted on the basis of size and book/market by factors sorted on the same basis is a tautology. This is not the case. For example, suppose that average returns were higher for stocks whose ticker symbols start later in the alphabet. (Maybe investors search for stocks alphabetically, so the later stocks are "overlooked.") This need not trouble us if Z stocks happened to have higher betas. If not—if letter of the alphabet were a CAPM anomaly like book/market—however, it would not necessarily follow that letter-based stock portfolios *move together*. Adding A–L and M–Z portfolios to the right-hand side of a regression of the 26 A,B,C, etc. portfolios on the market portfolio need not (and probably does not) increase the R^2 at all. The size and book/market premia are hard to measure, and seem to have declined substantially in recent years. But even if they decline back to CAPM values, Fama and French will still have found a surprisingly large source of common movement in stock returns.

More to the point, in testing a model, it is exactly the right thing to do to sort stocks into portfolios based on characteristics related to expected returns. When Black, Jensen, and Scholes and Fama and MacBeth first tested the CAPM, they sorted stocks into portfolios based on betas, because betas are a good characteristic for sorting stocks into portfolios that have a spread in average returns. If your portfolios have no spread in average returns—if you just choose 25 random portfolios, then there will be nothing for the asset pricing model to test.

In fact, despite the popularity of the Fama–French 25, there is really no fundamental reason to sort portfolios based on two-way or larger sorts of individual characteristics. You should use all the characteristics at hand that (believably!) indicate high or low average returns and simply sort stocks according to a one-dimensional measure of expected returns.

The argument over the status of size and book/market factors continues, but the important point is that it does so. Faced with the spectacular failure of the CAPM documented in Figures 20.9 and 20.11 one might have thought that any hope for a rational asset pricing theory was over. Now we are back where we were, examining small anomalies and arguing over refinements and interpretations of the theory. That is quite an accomplishment!

Macroeconomic Factors

> Labor income, industrial production, news variables, and conditional asset pricing models have also all had some successes as multifactor models.

I have focused on the size and value factors since they provide the most empirically successful multifactor model to date, and have therefore attracted much attention.

Several authors have used macroeconomic variables as factors in order to examine directly the story that stock performance during bad macroeconomic times determines average returns. Jagannathan and Wang (1996) and Reyfman (1997) use labor income; Chen, Roll, and Ross (1986) use industrial production and inflation among other variables. Cochrane (1996) uses investment growth. All these authors find that average returns line up against betas calculated using these macroeconomic indicators. The factors are theoretically easier to motivate, but none explains the value and size portfolios as well as the (theoretically less solid, so far) size and value factors.

Lettau and Ludvigson (2001a) specify a macroeconomic model that does just as well as the Fama–French factors in explaining the 25 Fama–French portfolios. Their plots of actual average returns versus model predictions show a relation as strong as those of Figures 20.12 and 20.13. Their model is

$$m_{t+1} = a + b(\mathrm{cay}_t)\Delta c_{t+1},$$

where cay is a measure of the consumption-wealth ratio. This is a "scaled factor model" of the sort advocated in Chapter 8. You can think of it as capturing a time-varying risk aversion.

Though Merton's (1971, 1973a) theory says that variables which predict market returns should show up as factors which explain cross-sectional variation in average returns, surprisingly few papers have actually tried to see whether this is true, now that we do have variables that we think forecast the market return. Campbell (1996) and Ferson and Harvey (1999) are among the few exceptions.

Momentum and Reversal

> Sorting stocks based on past performance, you find that a portfolio that buys long-term losers and sells long-term winners does better than the opposite—individual stock long-term returns mean-revert. This "reversal" effect makes sense given return predictability and mean-reversion, and is explained by the Fama–French three-factor model. However, a portfolio that buys short-term winners and sells short-term losers also does well—"momentum." This effect is a puzzle.

Since a string of good returns gives a high price, it is not surprising that stocks that do well for a long time (and hence build up a high price) subsequently do poorly, and stocks that do poorly for a long time (and

Table 20.12. *Average monthly returns from reversal and momentum strategies*

Strategy	Period	Portfolio Formation Months	Average Return, 10-1 (Monthly %)
Reversal	6307-9312	60-13	−0.74
Momentum	6307-9312	12-2	+1.31
Reversal	3101-6302	60-13	−1.61
Momentum	3101-6302	12-2	+0.38

Each month, allocate all NYSE firms on CRSP to 10 portfolios based on their performance during the "portfolio formation months" interval. For example, 60–13 forms portfolios based on returns from 5 years ago to 1 year, 1 month ago. Then buy the best-performing decile portfolio and short the worst-performing decile portfolio.
Source: Fama and French (1996, Table VI).

hence dwindle down to a low price, market value, or market/book ratio) subsequently do well. Table 20.12, taken from Fama and French (1996), reveals that this is in fact the case. (As usual, this table is the tip of an iceberg of research on these effects, starting with DeBont and Thaler [1985] and Jegadeesh and Titman [1993].)

Reversal

Here is the "reversal" strategy. Each month, allocate all stocks to 10 portfolios based on performance in year −5 to year −1. Then, buy the best-performing portfolio and short the worst-performing portfolio. The first row of Table 20.12 shows that this strategy earns a hefty −0.74% monthly return.[2] Past long-term losers come back and past winners do badly. This is a cross-sectional counterpart to the mean-reversion that we studied in Section 1.4. Fama and French (1988a) already found substantial mean-reversion—negative long-horizon return autocorrelations—in disaggregated stock portfolios, so one would expect this phenomenon.

Spreads in average returns should correspond to spreads in betas. Fama and French verify that these portfolio returns are explained by their three-factor model. Past losers have a high *HML* beta; they move together with value stocks, and so inherit the value stock premium.

[2] Fama and French do not provide direct measures of standard deviations for these portfolios. One can infer, however, from the betas, R^2 values, and standard deviation of market and factor portfolios that the standard deviations are roughly 1–2 times that of the market return, so that Sharpe ratios of these strategies are comparable to that of the market return.

Momentum

The second row of Table 20.12 tracks the average monthly return from a "momentum" strategy. Each month, allocate all stocks to 10 portfolios based on performance in the last *year*. Now, quite surprisingly, the winners continue to win, and the losers continue to lose, so that buying the winners and shorting the losers generates a positive 1.31% monthly return.

At every moment there is a most-studied anomaly, and momentum is that anomaly as I write. It is not explained by the Fama–French three-factor model. The past losers have low prices and tend to move with value stocks. Hence the model predicts they should have *high* average returns, not *low* average returns. Momentum stocks move together, as do value and small stocks, so a "momentum factor" works to "explain" momentum portfolio returns. This is so obviously ad hoc (i.e., an APT factor that will only explain returns of portfolios organized on the same characteristic as the factor) that nobody wants to add it as a risk factor.

A momentum factor is more palatable as a performance attribution factor. If we run fund returns on factors including momentum, we may be able to say that a fund did well by following a mechanical momentum strategy rather than by stock-picking ability, leaving aside why a momentum strategy should work. Carhart (1997) uses it in this way.

Momentum is really a new way of looking at an old phenomenon, the small apparent predictability of monthly individual stock returns. A tiny regression R^2 for forecasting monthly returns of 0.0025 (1/4%) is more than adequate to generate the momentum results of Table 20.12. The key is the large standard deviation of individual stock returns, typically 40% or more at an annual basis. The average return of the best performing decile of a normal distribution is 1.76 standard deviations above the mean,[3] so the winning momentum portfolio typically went up about 80% in the previous year, and the typical losing portfolio went down about 60% per year. Only a small amount of continuation will give a 1% monthly return when multiplied by such large past returns. To be precise, the monthly individual stock standard deviation is about $40\%/\sqrt{12} \approx 12\%$. If the R^2 is 0.0025, the standard deviation of the predictable part of returns is $\sqrt{0.0025} \times 12\% = 0.6\%$. Hence, the decile predicted to perform best will earn $1.76 \times 0.6\% \approx 1\%$

[3] We are looking for

$$E(r|r \geq x) = \frac{\int_x^\infty rf(r)\,dr}{\int_x^\infty f(r)\,dr},$$

where x is defined as the top 10th cutoff,

$$\int_x^\infty f(r)\,dr = \frac{1}{10}.$$

With a normal distribution, $x = 1.2816\sigma$ and $E(r|r \geq x) = 1.755\sigma$.

above the mean. Since the strategy buys the winners and shorts the losers, an R^2 of 0.0025 implies that one should earn a 2% monthly return by the momentum strategy—more even than the 1.3% shown in Table 20.12. Lewellen (2000) offers a related explanation for momentum coming from small *cross*-correlations of returns.

We have known at least since Fama (1965) that monthly and higher-frequency stock returns have slight, statistically significant predictability with R^2 in the 0.01 range. However, such small though statistically significant high-frequency predictability, especially in small stock returns, has also since the 1960s always failed to yield exploitable profits after one accounts for transactions costs, thin trading, high short-sale costs, and other microstructure issues. Hence, one naturally worries whether momentum is really exploitable after transactions costs.

Momentum does require frequent trading. The portfolios in Table 20.12 are reformed every month. Annual winners and losers will not change that often, but the winning and losing portfolios must still be turned over at least once per year. Carhart (1997) calculates transactions costs and concludes that momentum is not exploitable after those costs are taken into account. Moskowitz and Grinblatt (1999) note that most of the apparent gains come from short positions in small, illiquid stocks, positions that also have high transactions costs. They also find that a large part of momentum profits come from short positions taken November, anticipating tax-loss selling in December. This sounds a lot more like a small microstructure glitch rather than a central parable for risk and return in asset markets.

Table 20.12 already shows that the momentum effect essentially disappears in the earlier data sample, while reversal is even stronger in that sample. Ahn, Boudoukh, Richardson, and Whitelaw (2002) show that apparent momentum in international index returns is missing from the futures markets, also suggesting a microstructure explanation.

Of course, it is possible that a small positive autocorrelation is there and related to some risk. However, it is hard to generate real positive autocorrelation in realized returns. As we saw in Section 20.2, a slow and persistent variation in *expected* returns most naturally generates negative autocorrelation in *realized* returns. News that expected returns are higher means future dividends are discounted at a higher rate, so today's price and return declines. The only way to overturn this prediction is to suppose that expected return shocks are positively correlated with shocks to current or expected future dividend growth. A convincing story for such correlation has not yet been constructed. On the other hand, the required positive correlation is very small and not very persistent.

20.3 Summary and Interpretation

While the list of new facts appears long, similar patterns show up in every case. Prices reveal slow-moving market expectations of subsequent excess returns, because potential offsetting events seem sluggish or absent. The patterns suggest that there are substantial expected return premia for taking on risks of recession and financial stress unrelated to the market return.

Magnifying Glasses

The effects are not completely new. We knew since the 1960s that high-frequency returns are slightly predictable, with R^2 of 0.01 to 0.1 in daily to monthly returns. These effects were dismissed because there did not seem to be much that one could do about them. A 51/49 bet is not very attractive, especially if there is any transactions cost. Also, the increased Sharpe ratio one can obtain by exploiting predictability is directly related to the forecast R^2, so tiny R^2, even if exploitable, did not seem like an important phenomenon.

Many of the new facts amount to clever magnifying glasses, ways of making small facts economically interesting. For forecasting market returns, we now realize that R^2 rise with horizon when the forecasting variables are slow-moving. Hence small R^2 at high frequency can mean really substantial R^2, in the 30–50% range, at longer horizons. Equivalently, we realize that small expected return variation can add up to striking price variation if the expected return variation is persistent. For momentum and reversal effects, the ability to sort stocks and funds into momentum-based portfolios means that small predictability times portfolios with huge past returns gives important subsequent returns.

Dogs that Did Not Bark

In each case, an apparent difference in yield should give rise to an offsetting movement, but seems not to do so. Something *should* be predictable so that returns are *not* predictable, and it is not.

The d/p forecasts of the market return were driven by the fact that dividends *should* be predictable, so that returns are not. Instead, dividend growth seems nearly unpredictable. As we saw, this fact and the speed of the d/p mean-reversion imply the observed magnitude of return predictability.

The term structure forecasts of bond returns were driven by the fact that bond yields *should* be predictable, so that returns are not. Instead, yields seem nearly unpredictable at the one-year horizon. This fact means that the forward rate moves one for one with expected returns, and that a one percentage point increase in yield spread signals as much as a 5 percentage point increase in expected return.

Exchange rates should be forecastable so that foreign exchange returns are not. Instead, a one percentage point increase in interest rate abroad seems to signal a greater than one percentage point increase in expected return.

Prices Reveal Expected Returns

If expected returns rise, prices are driven down, since future dividends or other cash flows are discounted at a higher rate. A "low" price, then, can *reveal* a market expectation of a high expected or required return.

Most of our results come from this effect. Low price/dividend, price/earnings, price/book values signal times when the market as a whole will have high average returns. Low market value (price times shares) relative to book value signals securities or portfolios that earn high average returns. The "small-firm" effect derives from low prices—other measures of size such as number of employees or book value alone have no predictive power for returns (Berk [1997]). The "5 year reversal" effect derives from the fact that five years of poor returns lead to a low price. A high long-term bond yield means that the price of long-term bonds is "low," and this seems to signal a time of good long-term bonds returns. A high foreign interest rate means a low price on foreign bonds, and this seems to indicate good returns on the foreign bonds.

The most natural interpretation of all these effects is that the expected or required return—the risk premium—on individual securities as well as the market as a whole varies slowly over time. Thus we can track market expectations of returns by watching price/dividend, price/earnings, or book/market ratios.

Macroeconomic Risks

The price-based patterns in time-series and cross-sectional expected returns suggest a premium for holding risks related to recession and economy-wide financial distress. All of the forecasting variables are connected to macroeconomic activity (Fama and French [1989]). The dividend/price ratio is highly correlated with the default spread and rises in bad times. The term spread forecasts bond and stock returns, and is also one of the best recession forecasters. It rises steeply at the bottoms of recessions, and is inverted at the top of a boom. Thus, return forecasts are high at the bottom of business cycles and low at the top of booms. "Value" and "small-cap" stocks are typically distressed. Formal quantitative and empirically successful economic models of the recession and distress premia are still in their infancy (I think Campbell and Cochrane [1999] is a good start), but the story is at least plausible, and the effects have been expected by theorists for a generation.

To make this point come to life, think concretely about what you have to do to take advantage of the value or predictability strategies. You have to buy stocks or long-term bonds at the bottom, when stock prices are low after a long and depressing bear market; in the bottom of a recession or financial panic; a time when long-term bond prices and corporate bond prices are unusually low. This is a time when few people have the guts (the risk-tolerance) or the wallet to buy risky stocks or risky long-term bonds. Looking across stocks rather than over time, you have to invest in "value" companies, dogs by any standards. These are companies with years of poor past returns, years of poor sales, companies on the edge of bankruptcy, far off of any list of popular stocks to buy. Then, you have to sell stocks and long-term bonds in good times, when stock prices are high relative to dividends, earnings, and other multiples, when the yield curve is flat or inverted so that long-term bond prices are high. You have to sell the popular "growth" stocks with good past returns, good sales, and earnings growth.

I am going on a bit here to counter the widespread impression, best crystallized by Shiller (2000) that high price/earnings ratios must signal "irrational exuberance." Perhaps, but is it just a coincidence that this exuberance comes at the top of an unprecedented economic expansion, a time when the average investor is surely feeling less risk averse than ever, and willing to hold stocks despite historically low risk premia? I do not know the answer, but the rational explanation is surely not totally impossible! Is it just a coincidence that we are finding premia just where a generation of theorists said we ought to—in recessions, credit crunches, bad labor markets, investment opportunity set variables, and so forth?

This line of explanation for the foreign exchange puzzle is still a bit farther off, though there are recent attempts to make economic sense of the puzzle. (See Engel's [1996] survey; Atkeson, Alvarez, and Kehoe [1999] is a recent example.) At a verbal level, the strategy leads you to invest in countries with high interest rates. High interest rates are often a sign of monetary instability or other economic trouble, and thus may mean that the investments may be more exposed to the risks of global financial stress or a global recession than are investments in the bonds of countries with low interest rates, who are typically enjoying better times.

Overall, the new view of finance amounts to a profound change. We have to get used to the fact that most returns and price variation come from variation in *risk premia*, not variation in expected cash flows, interest rates, etc. Most interesting variation in priced risk comes from nonmarket factors. These are easy to say, but profoundly change our view of the world.

Doubts
Momentum is, so far, unlike all the other results. The underlying phenomenon is a small predictability of high-frequency returns. However, the price-based phenomena make this predictability important by noting that, with a slow-moving forecasting variable, the R^2 build over horizon. Momentum is based on a fast-moving forecast variable—the last year's return. Therefore the R^2 decline with horizon. Instead, momentum makes the tiny autocorrelation of high-frequency returns significant by forming portfolios of extreme winners and losers, so a small continuation of huge past returns gives a large current return. All the other results are easily digestible as a slow, business-cycle-related time-varying expected return. This specification gives negative autocorrelation (unless we add a distasteful positive correlation of expected return and dividend shocks) and so does not explain momentum. Momentum returns have also not yet been linked to business cycles or financial distress in even the informal way that I suggested for the price-based strategies. Thus, it still lacks much of a plausible economic interpretation. To me, this adds weight to the view that it is not there, it is not exploitable, or it represents a small illiquidity (tax-loss selling of small illiquid stocks) that will be quickly remedied once a few traders understand it. In the entire history of finance there has always been an anomaly-du-jour, and momentum is it right now. We will have to wait to see how it is resolved.

Many of the anomalous risk premia seem to be declining over time. The small-firm effect completely disappeared in 1980; you can date this as the publication of the first small-firm effect papers or the founding of small-firm mutual funds that made diversified portfolios of small stocks available to average investors. To emphasize this point, Figure 20.14 plots size portfolio average returns versus beta in the period since 1979. You can see that not only has the small-firm *premium* disappeared, the size-related variation in beta and expected return has disappeared.

The value premium has been cut roughly in half in the 1990s, and 1990 is roughly the date of widespread popularization of the value effect, though $\sigma/\sqrt{T}$ leaves a lot of room for error here. As you saw in Table 20.4, the last five years of high market returns have cut the estimated return predictability from the dividend/price ratio in *half*.

These facts suggest an uncomfortable implication: that at least some of the premium the new strategies yielded in the past was due to the fact that they were simply overlooked, they were artifacts of data-dredging, or they survived only until funds were created that allow many investors to hold diversified portfolios that exploit them.

Since they are hard to measure, one is tempted to put less emphasis on these average returns. However, they are crucial to our interpretation of the facts. The CAPM is perfectly consistent with the fact that there are additional sources of common variation. For example, it was long understood that

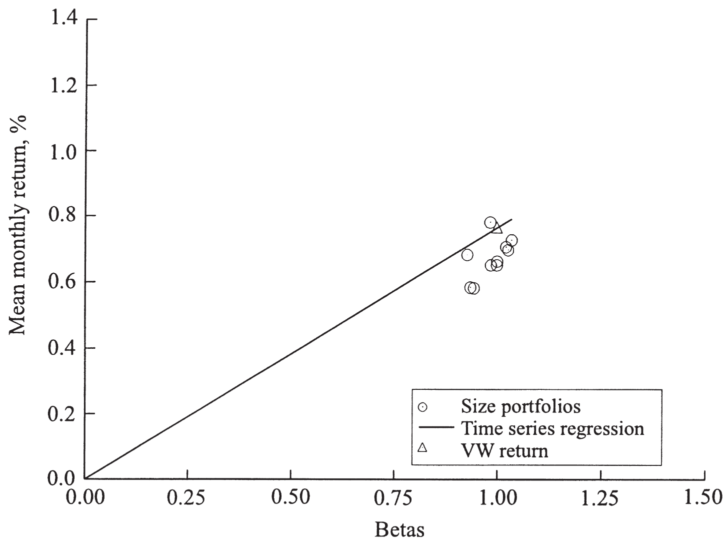

Figure 20.14. *Average returns vs. market betas. CRSP size portfolios less treasury bill rate, monthly data 1979–1998.*

stocks in the same industry move together; the fact that value or small stocks also move together need not cause a ripple. The surprise is that investors seem to earn an average return premium for holding these additional sources of common movement, whereas the CAPM predicts that (given beta) such common movements should have no effect on a portfolio's average returns.

Problems—Chapter 20

1. Does equation (20.11) condition down to information sets coarser than those observed by agents? Or must we assume that whatever VAR is used by the econometrician contains all information seen by agents?

2. Show that the two regressions in Table 20.9 are complementary—that the coefficients add up to one, mechanically, in sample.

3. Derive the return innovation decomposition (20.11), directly. Write the return

$$r_t = \Delta d_t + \rho \left(p_t - d_t\right) - \left(p_{t-1} - d_{t-1}\right).$$