Asset pricing with observable stochastic discount factors

P. N. Smith and M. R. Wickens

University of York

March 2002

Abstract

The stochastic discount factor model provides a general framework for pricing assets. By specifying the discount factor suitably it encompasses most of the theories currently in use, including CAPM and consumption CAPM. The SDF model has been based on the use of single and multiple factors, and on latent and observed factors. In most situations, and especially for the term structure, single factor models are inappropriate, whilst latent variables require the somewhat arbitrary specification of generating processes and are difficult to interpret. In this paper we survey the principal different implementations of the SDF model for FOREX, equity and bonds and we propose a new approach. This is based on the use of multiple factors that are observable and modelling the joint distribution of excess returns and the factors using a multi-variate GARCH-in-mean process. We argue that in general single equation and VAR models, although widely used in empirical finance, are inappropriate as they do not satisfy the no-arbitrage condition. Since risk premia arise from conditional covariation between returns and the factors, both a multi-variate context and having conditional covariances in the conditional mean process, is essential. We explain how apparent exceptions, such as the CIR and Vasicek models, in fact meet this requirement - but at a price. We explain our new approach, discuss how it might be implemented and present some empirical evidence, mainly from our own researches. Partly, to enable comparisons to be made, the survey also includes evidence from recent empirical work using more traditional approaches.

JEL Classification: G12, C51.

Keywords: Asset Pricing, Stochastic Discount Factors, Forex, Equity, Term Structure, Affine Factor Models, Consumption CAPM, Financial Econometrics, GARCH.

Note: This paper is forthcoming in a special issue of the Journal of Economic Surveys 2002, entitled The Econometrics of Financial Time Series edited by Michael McAleer and Les Oxley.

Correspondence. Address: Mike Wickens, Department of Economics, University of York, York, Y01 5DD, UK; email: mike.wickens@york.ac.uk.

1 Introduction

The stochastic discount factor (SDF) model is rapidly emerging as the most general and convenient way to price assets. Most existing asset pricing methods can be shown to be particular versions of the SDF model. This includes the capital asset pricing model (CAPM) of Sharpe (1964), Lintner (1965) and Black (1972), the general equilibrium consumption-based inter-temporal capital asset pricing model (CCAPM) of Rubinstein (1976) and Lucas (1978) and even the Black-Scholes theorem for pricing options. A detailed analysis of the SDF model may be found in the book by Cochrane (2000) and in the surveys by Ferson (1995) and Campbell (1999).

Cochrane argues that there are two polar approaches to asset pricing: absolute and relative asset pricing. Absolute pricing involves pricing each asset with reference to its exposure to fundamental sources of macroeoconomic risk. This provides a positive theory of asset prices and is the typical academic approach. Relative asset pricing is less ambitious. It aims to price an asset with reference to the prices of other assets. CAPM (i.e. the "beta" model) and the Black-Scholes theorem are examples of this. And bonds are commonly priced off the short rate. The SDF model can be used for both approaches. This survey focuses mainly on the absolute approach to asset pricing.

Another way to classify asset pricing models is whether they involve the use of observable or latent factors. Most of the research that has made explicit use of the SDF model have used latent factors. Sometimes an attempt is then made to give an economic interpretation to the estimated latent factors. Many studies that have implicitly used the SDF model such as CAPM and CCAPM have used observable factors. In this survey we seek to promote the explicit use of the SDF model with observable macroeconomic factors. The attraction of this approach is that it identifies the sources of risk, enables an assessment of their significance and offers the possibility of tilting portfolios to provide a hedge against these risks.

This approach is not without its problems. In the SDF model risk is measured by the conditional covariance of returns with the factors. For risk premia to be time-varying the factors must exhibit conditional heteroskedasticity. The drawback with using observable macroeconomic factors is that they do not display much heteroskedasticity, especially when the data are quarterly, or of lower frequency. There is more evidence of heteroskedasticity at higher frequencies, but there is not much data on macroeconomic variables at monthly or higher frequencies. Related problems are that the statistical models tend to be highly parameterised, and there is a lack of heteroskedasticity in macroeconomic variables. Together they make it difficult to obtain statistically well-determined models.

The SDF model is not the only factor model of asset pricing. The well-known arbitrage pricing theory (APT) of Ross (1976, 1977) is an example of a factor pricing model. A key feature, not possessed by other factor pricing models, is that in the SDF model the factors are linear functions of the conditional covariances between the factors and the excess return on the risky asset. This is not true, for example, of the APT.

In the past much econometric work on asset pricing and tests of market efficiency made use of a VAR model, see for example, Campbell and Shiller (1987, 1988) and Campbell, Lo and MacKinlay (1997). If the excess return is one of the variables in the VAR, then implicitly the risk premium is the right-hand side, the lag structure. Such a model can also be interpreted as a linear factor model in which the lagged variables are the factors. Usually no attempt is made to restrict the VAR to satisfy the no-arbitrage condition. It is not clear therefore what value this approach has except to test the joint hypothesis of market efficiency *and risk neutrality,* i.e. the expectations hypothesis. The null hypothesis for is that the right-hand side variables are jointly insignificant. An exception to this, discussed later, occurs when certain affine factor models are used.

The paper is set out as follows. In section 2 we discuss the SDF asset pricing model in general terms, setting out the theory and then relating this to specific implementations such as CCAPM, CAPM, affine factor models and APT. We then explain the implications of this for empirical finance. In section 3 we discuss the econometric methodology commonly used in estimating SDF models. We include a discussion of VAR models, multivariate conditional heteroskedasticity models, simulation methods using the efficient method of moments and the generalised method of moments. We then consider applications of the SDF approach to the term structure (section 4), the foreign exchange market (section 5) and equity (section 6). We review the use of unobservable and observable SDF models, and we suggest new ways to implement the observable SDF model using the multivariate GARCH-in-mean specification. We report our conclusions in section 7.

2 The SDF asset pricing model

The SDF model is based on a very simple proposition. The price of an asset in period t, is the expected discounted value of the asset's pay-off in period $t+s$ based on information available in period t:

$$P_t = E_t[M_{t+s}X_{t+s}] \tag{1}$$

where P_t = the price of the asset in period t, X_{t+s} = the pay-off of the asset in period $t+s$, M_{t+s} = the discount factor for period $t+s$ ($0 \leq M_{t+s} \leq 1$), and E_t = the expectation taken with respect to information available in period t. Thus P_t is the current value of the period $t+s$ income X_{t+s}. In general this income will not be known in period t and will be a random variable.

The discount factor is sometimes called the pricing kernel and will be a stochastic variable. For convenience, unless explicitly changed, hereafter we will assume that $s = 1$.

Equation (1) can also be written in terms of the asset's gross return $R_{t+1} = \frac{X_{t+1}}{P_t} = 1 + r_{t+1}$:

$$1 = E_t[M_{t+1}\frac{X_{t+1}}{P_t}] = E_t[M_{t+1}R_{t+1}] \tag{2}$$

Gross returns can be defined either in nominal or real terms; correspondingly, the discount factor must then also be expressed in nominal or real terms.

Examples of assets that can be priced in this way together with their pay-offs are:

1. A stock that pays a dividend of D_{t+1} and has a re-sale value of P_{t+1} at $t+1$ has the pay-off $X_{t+1} = P_{t+1} + D_{t+1}$.

2. A T-bill that pays 1 unit one unit of the consumption good regardless of the state of nature next period has $X_{t+1} = 1$.

3. A bond that has a constant coupon payment of C and can be sold for P_{t+1} next period has $X_{t+1} = P_{t+1} + C$.

4. A bank deposit that pays the risk-free rate of return r_t^f between t and $t+1$ has $X_{t+1} = 1+r_t^f$.

5. A call option costing P_t that gives the holder the right to purchase a stock at the exercise price K at date T has $X_{t+1} = \max[S_T - K, 0]$.

As

$$E_t(M_{t+1}R_{t+1}) = E_t(M_{t+1})E_t(R_{t+1}) + Cov_t(M_{t+1}, R_{t+1}), \tag{3}$$

the expected return on the asset is given by

$$E_t(R_{t+1}) = \frac{1 - Cov_t(M_{t+1}, R_{t+1})}{E_t(M_{t+1})}. \tag{4}$$

The conditional covariance between M_{t+1} and R_{t+1} can be estimated as the covariance of the error terms in the joint model of M_{t+1} and R_{t+1}, and possibly other variables.

Equation (4) holds whether the asset is risky or risk free. If it is risk free, then its pay-off in period $t+1$ is known with certainty. Without loss of generality, this can be assumed to be 1. As a result, R_{t+1} will be known in period t and can be written as $R_{t+1} = 1 + r_t^f$, where r_t^f is the risk-free rate of return. Equation (2) can then be written as

$$1 = E_t[M_{t+1}(1 + r_t^f)] = (1 + r_t^f)E_t(M_{t+1}) \tag{5}$$

Thus $E_t(M_{t+1}) = \frac{1}{1+r_t^f}$. This carries the implication that the discount factor is the random variable $M_{t+1} = \frac{1}{1+r_t^f} + \xi_{t+1}$ where the random variable ξ_{t+1} has zero conditional mean, i.e. $E_t\xi_{t+1} = 0$.

The excess return over the risk-free rate is obtained by substituting equation (5) into (4) and using $R_{t+1} = 1 + r_{t+1}$ to give

$$\begin{aligned} E_t r_{t+1} - r_t^f &= -(1+r_t^f)Cov_t(M_{t+1}, R_{t+1}) \\ &= -(1+r_t^f)Cov_t(M_{t+1}, r_{t+1} - r_t^f) \end{aligned} \tag{6}$$

This is the no-arbitrage condition that all correctly priced assets must satisfy. The term on the right-hand side is the risk premium. It is the extra return over the risk-free rate that is required to compensate investors for holding the risky asset.

2.1 Risk

Since risk-averse investors require compensation for taking on risk, the risk premium must be non-negative. This implies that $Cov_t(M_{t+1}, R_{t+1}) \leq 0$. The source of the risk is that when the discount factor is in a high state, the return on the asset is in a low state. A risk-neutral investor is indifferent to such considerations and does not require a risk premium

Another way to write (6), the no-arbitrage equation, is

$$\{E_t[r_{t+1} + (1+r_t^f)Cov_t(M_{t+1}, R_{t+1})]\} - r_t^f = 0$$

The first term (in braces) is the expected return on a risky asset after adjusting for risk. It is also known as the expected return on the risky asset where expectations are taken with respect to the risk-neutral distribution, i.e. using E_t^N instead of E_t. In effect, to obtain the risk-neutral distribution, the true distribution is shifted to the left by subtracting the risk premium from r_{t+1}. The risk-neutral distribution is used in pricing derivative assets such as options. Thus

$$E_t^N(r_{t+1}) = E_t[r_{t+1} + (1+r_t^f)Cov_t(M_{t+1}, R_{t+1})] = r_t^f$$

Consequently, a self-financing portfolio consisting of borrowing at the risk-free rate and investing the proceeds in the risky asset will have zero expected return when the expected return on the risky asset is evaluated using its risk-neutral distribution.

The price and quantity of risk are concepts sometimes used in relation to the risk premium. They are analagous to the notion of total expenditure being equal to price multiplied by quantity. Thus

$$\begin{aligned}\text{risk premium} &= \text{price of risk} \times \text{quantity of risk} \\ &= -(1+r_t^f)cov_t(M_{t+1}, r_{t+1}-r_t^f) \\ &= \beta_t\lambda_t\end{aligned}$$

$$\beta_t = \text{price of risk} = -(1+r_t^f)\frac{cov_t(M_{t+1}, r_{t+1}-r_t^f)}{SD_t(r_{t+1}-r_t^f)}$$

$$\lambda_t = \text{quantity of risk} = SD_t(r_{t+1}-r_t^f)$$

where SD_t means the conditional standard deviation. The price of risk can also be written as

$$\beta_t = cov_t[\frac{M_{t+1}}{E_t(M_{t+1})}, \frac{r_{t+1}-r_t^f}{SD_t(r_{t+1}-r_t^f)}].$$

This is also the conditional correlation coefficient, and the regression coefficient of $\frac{M_{t+1}}{E_t(M_{t+1})}$ on $\frac{r_{t+1}-r_t^f}{SD_t(r_{t+1}-r_t^f)}$ (or vice-versa). It measures the effect on the stochastic discount factor (scaled to be unit-free) of a unit change in the standardised excess return (also scale free), etc. The quantity of risk is just due to variability in the excess return.

We have not yet discussed how to specify the discount factor or, when it is unkown, what properties we expect it to possess. We can, however, derive bounds for both the risk premium and the discount factor. As they are free from any additional behavioural assumptions, these bounds might be useful when looking for suitable proxy variables for M_{t+1}.

If $\rho_t(M_{t+1}, r_{t+1})$ is denoted as the correlation coefficient of the joint conditional distribution of M_{t+1} and r_{t+1}, and $V_t(M_{t+1})$ and $V_t(R_{t+1})$ are their conditional variances, then

$$\begin{aligned}Cov_t(M_{t+1}, r_{t+1})^2 &= \rho_t^2(M_{t+1}, r_{t+1})V_t(M_{t+1})V_t(r_{t+1}) \\ &\leq V_t(M_{t+1})V_t(r_{t+1}),\end{aligned}$$

as $-1 \leq \rho_t(M_{t+1}, r_{t+1} \leq 1$, this provides upper and lower bounds for the risk premium. As the risk premium must be non-negative, these are

$$0 \leq E_t(r_{t+1}-r_t^f) \leq (1+r_t^f)SD_t(M_{t+1})SD_t(r_{t+1}).$$

This provides a lower bound for $V_t(M_{t+1})$ that can be calculated from just knowledge of the returns and the conditioning information:

$$V_t(M_{t+1}) \geq [\frac{E_t(r_{t+1}-r_t^f)}{1+r_t^f}]^2\frac{1}{V_t(r_{t+1})}. \tag{7}$$

A widely used assumption, due partly to its convenience and because it is a reasonably good approximation, is that the stochastic discount factor and the gross return are jointly distributed as lognormal. If $\ln x$ is $N(\mu, \sigma^2)$ then $E(x) = \exp(\mu + \frac{\sigma^2}{2})$ and hence $\ln E(x) = \mu + \frac{\sigma^2}{2}$. If $m_{t+1} = \ln M_{t+1}$ and $r_{t+1} = \ln R_{t+1}$ are jointly normally distributed then

$$1 = E_t[M_{t+1}R_{t+1}] = \exp\{E_t[\ln(M_{t+1}R_{t+1})] + V_t[\ln(M_{t+1}R_{t+1}]/2\}.$$

Taking the logarithms yields

$$\begin{aligned}\ln E_t[M_{t+1}R_{t+1}] &= E_t[\ln(M_{t+1}R_{t+1})] + V_t[\ln(M_{t+1}R_{t+1})]/2 \\ &= E_t(m_{t+1}) + E_t(r_{t+1}) + V_t(m_{t+1})/2 + V_t(r_{t+1})/2 \\ &\quad + cov_t(m_{t+1}, r_{t+1}) = 0. \end{aligned} \tag{8}$$

When $r_{t+1} = r_t^f$, the risk-free rate, equation (8), becomes

$$E_t(m_{t+1}) + r_t^f + \frac{1}{2}V_t(m_{t+1}) = 0 \tag{9}$$

as $E_t(r_t^f) = r_t^f$ and $V_t(r_t^f) = 0$. Subtracting (9) from (8) produces the key no-arbitrage condition under log-normality:

$$E_t(r_{t+1} - r_t^f) + \frac{1}{2}V_t(r_{t+1}) = -Cov_t(m_{t+1}, r_{t+1}). \tag{10}$$

Comparing equations (10) and (6), apart from the switch to logarithms, (10) involves an additional term on the left-hand side, namely, half the conditional variance of returns. This is called the Jensen effect. It arises because expectations are being taken of a non-linear function, and $E[f(x)] \neq f[E(x)]$ unless $f(x)$ is linear. The right-hand side of (10) is the risk premium.

2.2 How to choose $\mathbf{M}_{t+1}$?

The various asset pricing models differ primarily due to their choice of discount factor. This can be either an implicit or an explicit choice. Further, we can either use observable or unobservable (latent variable) factors. First we consider the two leading examples of implicit observable factor models, the consumption-based inter-temporal capital assset pricing model (CCAPM) and the capital asset pricing model (CAPM).

2.2.1 CCAPM

This is a general equilibrium model. All investments are evaluated in terms of the investor's expected present value of current and future utility $U(C_t)$, where C_t is *real* consumption of goods and services. The investor's problem in period t is to choose consumption C_{t+s}, and the real stock of financial assets W_{t+s+1}, for all $s \geq 0$, subject to the investor's budget constraint. Assuming time-separable utility, the problem can be solved using stochastic dynamic programming. It can be written:

$$\max_{C_t}\{\mathcal{U}_t = U(C_t) + \beta E_t(\mathcal{U}_{t+1})\} \tag{11}$$

subject to the budget constraint

$$C_t + W_{t+1} = Y_t + W_t R_t \tag{12}$$

where Y_t is real income, R_t is the real gross return on equity and $0 \leq \beta = \frac{1}{1+\theta} \leq 1$ is the rate of time-discount of preferences. The solution is the Euler equation

$$E_t[\frac{\beta U'(C_{t+1})}{U'(C_t)} R_{t+1}] = 1. \tag{13}$$

Comparing this with equation (2), we can see that CCAPM has implicitly defined the SDF (or pricing kernel) as

$$M_{t+1} = \frac{\beta U'(C_{t+1})}{U'(c_t)} \simeq \beta\,(1 - \sigma_t \Delta \ln C_{t+1}) \tag{14}$$

where $\sigma_t = -\frac{C_t U_t''}{U_t'} > 0$ is the coefficient of relative risk aversion. Hence, the no-arbitrage condition for the real rate of return on a risky asset is

$$E_t(r_{t+1} - r_t^f) = \beta(1 + r_t^f)\sigma_t Cov_t\,(\Delta \ln C_{t+1}, r_{t+1}) \tag{15}$$

If we assume that $\Delta \ln C_{t+1}$ and r_{t+1} are jointly lognormal then, using $m_{t+1} \simeq -(\theta + \sigma_t \Delta \ln C_{t+1})$, we obtain the no-arbitrage condition

$$E_t(r_{t+1} - r_t^f) + \frac{1}{2} V_t(r_{t+1}) = \sigma_t Cov_t(\Delta \ln C_{t+1}, r_{t+1}). \tag{16}$$

Equations (15) and (16) both imply that the greater the correlation between consumption growth and the risky return, the higher the risk premium, i.e. lower future returns are associated with lower future consumption growth, a common feature of the business cycle.

For the special case of the widely-used power utility function

$$U = \frac{C_t^{1-\sigma} - 1}{1 - \sigma} \tag{17}$$

which has constant coefficient of relative risk aversion σ,

$$M_{t+1} = \beta \left(\frac{C_{t+1}}{C_t}\right)^{-\sigma} \simeq \beta\,(1 - \sigma \Delta \ln C_{t+1}) \tag{18}$$

In applications to equity pricing it was found that CCAPM produced a risk premium that is too low. This result is known as the equity premium puzzle, see Mehra and Prescott (1985). The failure of CCAPM has been attributed to a lack of sufficient volatility in the stochastic discount factor to match that of equity returns. Inspection of equation (16) shows that this failure could be due either to too low a consumption growth rate, or to a too low a coefficient of relative risk aversion, σ_t, or to both. As shown by Mehra and Prescott, implausibly large values of σ_t are required to eliminate the equity premium puzzle. This led to a search for utility functions that produce a larger risk premium. Attention focused on the assumption of the time separability of utility.

An example of the alternative of time non-separable utility is the habit persistence model where $U_t = U(C_t, X_t)$ and X_t is the habitual level of consumption. Constantinides (1990) proposed the form

$$U_t = \frac{(C_t - \lambda X_t)^{1-\sigma} - 1}{1-\sigma}$$

The stochastic discount factor implied by this is $M_{t+1} = \beta \left(\frac{C_{t+1} - \lambda X_{t+1}}{C_t - \lambda X_t}\right)^{-\sigma}$. By a suitable choice of λ and X_t it is possible to produce a discount factor that displays greater volatility and hence a larger risk premium. Later, we shall consider habit persistence models in more detail together with other versions of the habit persistence model, and also their implementation.

A second example of time non-separable utility has been proposed by Kreps-Porteus (1978). Their general formulation is

$$\mathcal{U}_t = \mathcal{U}[C_t,\ E_t(\mathcal{U}_{t+1})]$$

Epstein-Zin (1989, 1990, 1991) have implemented a special case of this based on the constant elasticity of substitution (CES) function. This is

$$\mathcal{U}_t = \left[(1-\beta)C_t^{1-\frac{1}{\gamma}} + \beta\left(E_t(\mathcal{U}_{t+1}^{1-\sigma})\right)^{\frac{1-\frac{1}{\gamma}}{1-\sigma}}\right]^{\frac{1}{1-\frac{1}{\gamma}}}$$

where $\beta =$ the discount factor, $\sigma =$ the coefficient of relative risk aversion and $\gamma =$ the elasticity of inter-temporal substitution. Thus, whereas the additively separable inter-temporal utility function restricts the coefficient of relative risk aversion and the elasticity of inter-temporal substitution to be identical, the time non-separable model allows them to be different. Epstein and Zin show that maximising $\mathcal{U}_t$ subject to the slightly different budget constraint $W_{t+1} = R^m_{t+1}(W_t - C_t)$, where $r^m_{t+1} = R^m_{t+1} - 1$ is the return on the portfolio, gives the following asset pricing equation for any individual asset:

$$E_t\left\{\left[\beta\left(\frac{C_{t+1}}{C_t}\right)^{-\frac{1}{\gamma}}\right]^{\frac{1-\sigma}{1-\frac{1}{\gamma}}}\left(R^m_{t+1}\right)^{1-\frac{1-\sigma}{1-\frac{1}{\gamma}}} R_{t+1}\right\} = 1. \tag{19}$$

Thus the stochastic discount factor is

$$M_{t+1} = \left[\beta\left(\frac{C_{t+1}}{C_t}\right)^{-\frac{1}{\gamma}}\right]^{\frac{1-\sigma}{1-\frac{1}{\gamma}}}\left(R^m_{t+1}\right)^{1-\frac{1-\sigma}{1-\frac{1}{\gamma}}} \tag{20}$$

Compared with time separable utility, this has two additional degrees of freedom which can boost the size of the risk premium. First, the power index is no longer the coefficient of relative risk aversion, and is therefore free to take on large values. Second, M_{t+1} now varies with the return on the portfolio. Assuming log-normality, Campbell, Lo and MacKinlay (1997) have rewritten the no-arbitrage condition equation (19) as

$$E_t(r_{t+1} - r_t^f) + \frac{1}{2}V_t(r_{t+1}) = \frac{1-\sigma}{1-\gamma}Cov_t(\Delta \ln C_{t+1}, r_{t+1}) + (1 - \frac{1-\sigma}{1-\frac{1}{\gamma}})Cov_t(r_{t+1}^m, r_{t+1}) \quad (21)$$

Thus, this more general class of inter-temporal utility functions is also an SDF model with observable factors. Equation (21) has two factors, consumption growth and the portfolio (or market) return. The last term in equation (21) is a potential additional source of risk. We examine evidence on the Epstein-Zin model below.[1]

2.2.2 CAPM

Static CAPM relates the expected excess return on the risky asset over the risk-free rate to the excess return of the market portfolio, r_{t+1}^m which is given by $(1+r_{t+1}^m) = \frac{W_{t+1}}{W_t}$, where W_t is wealth. The key results are:

$$\begin{aligned} E_t(r_{t+1} - r_t^f) &= \beta_t E_t(r_{t+1}^m - r_t^f), \\ \beta_t &= \frac{Cov_t(r_{t+1}^m, r_{t+1})}{V_t(r_{t+1}^m)} \\ E_t(r_{t+1}^m - r_t^f) &= \sigma_t V_t(r_{t+1}^m). \end{aligned} \quad (22)$$

Thus the risk premium depends in part on the market return. But we can also write equation (22) as

$$E_t(r_{t+1} - r_t^f) = \sigma_t Cov_t(r_{t+1}^m, r_{t+1}) = \sigma_t Cov_t(\frac{\Delta W_{t+1}}{W_t}, r_{t+1})..$$

This shows that the risk premium arises from the conditional covariance of the asset's return with the market return or, equivalently, the rate of growth of wealth. We conclude therefore that, in effect, CAPM is an SDF model in which the discount factor is $\sigma_t(1+r_{t+1}^m)$ or $\sigma_t\frac{\Delta W_{t+1}}{W_t}$. In implementing CAPM there is some disagreement about which assets should be included in wealth. In principle all assets should be included, but in practice it is common to use only financial wealth.

[1] The complexity of the derivation of the Euler equation has led subsequent authors to simply quote the original result. A simpler derivation is possible for the two period case where $\mathcal{U}_t = \left[(1-\beta)U(C_t)^\alpha + \beta\left(E_t(U(C_{t+1})^\alpha)^\theta\right)\right]^{\frac{1}{\alpha}}$ and the budget constraint is $a_{t+1} + c_t = (1+r_t)a_t$. It can be shown that $\frac{\partial \mathcal{U}_t}{\partial C_t} = 0$ implies the Euler equation $\beta\theta[E_t U(C_{t+1})^\alpha]^{\theta-1}E_t[\left(\frac{U(C_{t+1})}{U(C_t)}\right)^{\alpha-1}\frac{U'(C_{t+1})}{U'(C_t)}(1+r_{t+1})] = 1$ Subtracting the corresponding equation for the risk-free rate away from this gives the no-arbitrage condition $E_t[\left(\frac{U(C_{t+1})}{U(C_t)}\right)^{\alpha-1}\frac{U'(C_{t+1})}{U'(C_t)}(r_{t+1} - r_t^f)] = 0$. For power utility $U(C_t) = \frac{C_t^{1-\sigma}}{1-\sigma}$ this becomes $E_t[\left(\frac{C_{t+1}}{C_t}\right)^{\alpha(1-\sigma)-1}(r_{t+1} - r_t^f)] = 0$.

Comparing CCAPM - and in particular equation (15) - with CAPM, we observe that CCAPM uses the covariance of consumption growth with the risky asset to measure risk while CAPM uses the covariance of the growth rate of wealth with the risky asset. If consumption is proportional to wealth, then the determination of the risk premia in CAPM and CCAPM are identical. Apart from CAPM and CCAPM the problem of how to implement the stochastic discount factor model remains. How might we model the discount factor (or pricing kernel) M_{t+1} without making either of these assumptions?

2.3 Multi-factor models

CAPM and CCAPM under time separability are examples of single-factor models. CAPM assumes that the stochastic discount factor is $M_{t+1} = \sigma_t(1 + r^m_{t+1})$, time-separable CCAPM assumes that it is $M_{t+1} = \beta \frac{U'_{t+1}}{U'_t} \simeq \beta [1 - \sigma_t \Delta \ln C_{t+1}]$. In each case, for $\sigma_t = \sigma$, a constant, and we can write

$$M_{t+1} = a + bz_{t+1}$$

or, assuming log-normality,

$$\ln M_{t+1} = m_{t+1} = a + bz_{t+1}$$

where z_{t+1} is r^m_{t+1} or $\Delta \ln C_{t+1}$. Multi-factor generalisations are:

$$M_{t+1} = a + \sum_i b_i z_{i,t+1},$$

$$\ln M_{t+1} = m_{t+1} = a + \sum_i b_i z_{i,t+1}.$$

Because these are linear models, they are also called *affine* factor models (meaning linear).

The implication for asset pricing is that

$$\begin{aligned} E_t(r_{t+1} - r^f_t) &= -(1 + r^f_t) Cov_t(M_{t+1}, r_{t+1}) \\ &= -(1 + r^f_t) \sum_i b_i Cov_t(z_{i,t+1}, r_{t+1}) = \sum_i \beta_i \ f_{it}, \end{aligned} \tag{23}$$

or, assuming log-normality,

$$\begin{aligned} E_t(r_{t+1} - r^f_t) + \frac{1}{2} V_t(r_{t+1}) &= -Cov_t(m_{t+1}, r_{t+1}) \\ &= -\sum_i b_i Cov_t(z_{i,t+1}, r_{t+1}) = \sum_i \beta_i \ f_{it}, \end{aligned} \tag{24}$$

where the f_{it} are known as *common factors*. Time non-separable CCAPM is an example of a two-factor SDF model where the factors are $\Delta \ln C_{t+1}$ and r^m_{t+1}. If returns are nominal and not real then CCAPM has two factors: $\Delta \ln C_{t+1}$ and π_{t+1}, the inflation rate.

2.4 Latent variable affine factor models

We now consider explicit but unobservable factor models. The most widely-used are the latent variable affine factor models of Vasicek (1977) and Cox, Ingersoll and Ross (1985). Both assume that the logarithm of the stochastic discount factor m_{t+1} can be expressed as a linear function of one or more random variables each of which follows a mean-reverting first-order autoregressive process. Moreover, instead of m_{t+1} being an exact linear function of the factors as above, allowance is made for an additional random error. We illustrate these models using the single factor z_{t+1}.

2.4.1 Vasicek model

In the Vasicek model it is assumed that

$$\begin{aligned} m_{t+1} &= \alpha + \beta z_{t+1} + \lambda\sigma\varepsilon_{t+1} \\ z_{t+1} - \mu &= \theta(z_t - \mu) + \sigma\varepsilon_{t+1}, \qquad 0 \leq |\theta| < 1 \end{aligned} \tag{25}$$

where $\varepsilon_{t+1} \sim iid(0,1)$. These assumptions imply that m_{t+1} is the ARMA(1,1):

$$m_{t+1} = (\alpha + \beta\mu)(1-\theta) + \theta m_t + (\lambda + \beta)\sigma\varepsilon_{t+1} - \theta\lambda\sigma\varepsilon_t.$$

Using the log-normal version of the asset pricing model, the no-arbitrage equation becomes

$$\begin{aligned} E_t(r_{t+1} - r_t^f) + \frac{1}{2}V_t(r_{t+1}) &= -Cov_t(m_{t+1}, r_{t+1}) \\ &= -(\lambda+\beta)\sigma Cov_t(\varepsilon_{t+1}, r_{t+1}). \end{aligned}$$

2.4.2 Cox, Ingersoll and Ross (CIR) model

In this model it is assumed that

$$\begin{aligned} m_{t+1} &= \alpha + \beta z_{t+1} + \lambda\sigma\sqrt{z_t}\varepsilon_{t+1} \\ z_{t+1} - \mu &= \theta(z_t - \mu) + \sigma\sqrt{z_t}\varepsilon_{t+1}. \end{aligned}$$

The attraction of the CIR model is that the return on the asset is bounded at, or above, zero. In the Vasicek model it can become negative. In the CIR model $V_t(z_{t+1}) = \sigma^2 z_t$, hence $\lim_{z_t \to 0} V_t(z_{t+1}) = 0$, i.e. as the factor approaches zero, its variance also converges to zero. The no-arbitrage equation is

$$E_t(r_{t+1} - r_t^f) + \frac{1}{2}V_t(r_{t+1}) = -(\lambda+\beta)\sigma\sqrt{z_t}Cov(\varepsilon_{t+1}, r_{t+1}).$$

The CIR model can be generalised to,

$$\begin{aligned} m_{t+1} &= \alpha + \beta z_{t+1} + \lambda(z_t)\varepsilon_{t+1} \\ \Delta z_{t+1} &= \mu(z_t) + \sigma(z_t)\varepsilon_{t+1} \end{aligned}$$

A common version of this is to choose $\sigma(z_t) = \sqrt{\phi + \omega z_t}$ and $\lambda(z_t) = \lambda\sigma(z_t)$, see Dai and Singleton (2000). Later we show how this can be implemented for the term structure.

2.5 Multi-asset multi-factor models

The pricing equations derived so far refer to a single asset. The form of the pricing equation for every asset will be the same. If the return on the jth asset is $r_{j,t+1}$ then its no-arbitrage equation is

$$E_t(r_{j,t+1} - r_t^f) = -(1 + r_t^f)\sum_i b_{ij}Cov_t(z_{i,t+1}, r_{j,t+1}) \tag{26}$$

The equivalent equation under log-normality is

$$E_t(r_{j,t+1} - r_t^f) + \frac{1}{2}V_t(r_{j,t+1}) = -\sum_i b_{ij}Cov_t(z_{i,t+1}, r_{j,t+1}) \tag{27}$$

In matrix terms, equations (26) and (27) can be written

$$E_t(\mathbf{r}_{t+1} - r_t^f\ell) = -(1 + r_t^f)\mathcal{C}_t$$

$$E_t(\mathbf{r}_{t+1} - r_t^f\ell) + \frac{1}{2}\mathcal{V}_t = -(1 + r_t^f)\mathcal{C}_t$$

where $\mathbf{r}_{t+1}$ is vector of returns, $\mathcal{C}_t$ is a column vector formed from the diagonal elements of $\mathbf{BV}_t$, where $\mathbf{B} = \{b_{ij}\}$ and $\mathbf{V}_t = \{Cov_t(z_{i,t+1}, r_{j,t+1})\}$, $\mathcal{V}_t$ is column vector formed from the diagonal elements of $\mathbf{V}_t$ and ℓ is a vector of ones

2.6 Other types of factor models

Some asset pricing models involve the use of factors but are not examples of SDF models - even implicitly. The best known example is APT. This satisfies the condition of no-arbitrage. Another widely used model in empirical finance, especially for testing market efficiency, is a VAR model in which the excess return on the risky assets is one of the variables. In general a VAR is neither an SDF model, nor does it satisfy the no-arbitrage condition. This calls into question such uses of a VAR in finance.

2.6.1 Arbitrage Pricing Theorem (APT)

This is one of the best-known multi-factor models of asset pricing. For a survey of APT see Connors and Korajczyk (1995). The basic idea underlying the APT is to avoid the formality of CAPM or CCAPM and simply model the return on any asset j as the sum of k common factors plus an idiosyncratic component that is uncorrelated with the common factors. These factors are sometimes assumed to be obtainable from the principal components of the returns on a vector of assets.

The return on the j^{th} asset is

$$r_j = a_j + \sum_i^k b_{ij} f_i + \varepsilon_j, \qquad j = 1, ..., n, \tag{28}$$

where $E(\varepsilon_j) = 0$,and $E(\varepsilon_j, f_i) = 0$. In this expression, f_i are some variables useful for forecasting r_j, e.g., f_i are returns on other assets. Equation (28) therefore has the same form as (23). The difference lies primarily in the choice of factors, f_i and whether a risk-free asset exists.

The aim is to "diversify away" the risk in asset j arising from the idiosyncratic component ε_j by forming no-arbitrage portfolios. A portfolio $r^p = \sum_{j=1}^{n} w_j r_j$ is a no-arbitrage portfolio if $r^p = 0$, with some $w_j < 0$. Hence, $0 = r^p = \sum_j w_j a_j + \sum_i^k (\sum_j^n w_j b_{ij}) f_i + \sum_j w_j \varepsilon_j$.where $w_n = 1 - \sum_{j=1}^{n-1} w_j$. In general therefore, each of the terms $\sum_j w_j a_j$, $\sum_j^n w_j b_{ij}$ and $\sum_j w_j \varepsilon_j$ will be zero. If one of the assets the n^{th} (say) is risk-free, then $f_k = r_n = r^f$. The return on the no-arbitrage portfolio can then be written as

$$r^p = \sum^{n-1} w_j r_j + (1 - \sum_j^{n-1} w_j) r^f = \sum^{n-1} w_j (r_j - r^f) + r^f = 0$$

Hence, for any single asset, the excess return is

$$\begin{aligned} r_i - r^f &= \sum_{i \neq j}^{n-1} \frac{w_j}{w_i} (r_j - r^f) = -\sum_{i \neq j}^{n-1} \frac{w_j}{w_i} \{[a_j + \sum_i^{k-1} b_{ij} f_i + b_{kj} r^f + \varepsilon_j] - r^f\} \\ &= \sum \beta_i f_i + \beta_f r^f \end{aligned}$$

Thus the APT has factors f_i and r^f. Since f_i need not be a conditional covariance, the APT is not in general an SDF model.

2.7 The implications of the SDF model for empirical finance

So far the main broad conclusions to emerge for empirical finance are that, unless it is explictly assumed that investors are risk neutral, a risk premium should be included in the model. The SDF approach shows that this is a (possibly linear) function of the conditional covariance between the factors and the excess return. This implies that in general in empirical finance the model must be multivariate, not univariate, as the joint distribution of the excess return and the factors is required to model the risk premium. Further, it is not sufficient simply to specify the model with a time-varying conditional covariance matrix. The model must also have the conditional covariances in the conditional mean of the excess return equation in order for this equation to satisfy the no-arbitrage condition. Very few of the models that have been used in empirical finance (to study, for example, equity, bonds or foreign exchange) satisfy these requirements.

Two qualifications may be made to these strictures, both of which are examined in more detail later. First, there is a special case where it is not necessary to explictly include conditional covariances in the mean of an SDF model in order to satisfy the condition of no arbitrage, as the conditional covariances are included implictly. In this case not all of the strictures above apply. This happens where, due to the assumed structure for the stochastic discount factor, the conditional covariances are functions of the variables themselves. An example is the CIR latent variable model of the term structure. It can be shown that, in the case of a single-factor model, the conditional covariance between the factor and the excess return on a long bond (and hence its risk premium) is a linear function of the short rate and hence, in effect, is an observable factor. If there is more than one factor, then the risk premium is a linear function of short and long rates. In these cases it would be possible to model the term structure using a VAR. We examine this idea further below. The no-arbitrage condition would imply restrictions on the coefficients of the VAR.

Second, it is often possible to estimate the parameters and to carry out statistical tests using the Euler equation itself. Solving the Euler equation to obtain an expression for the risky return would not then be necessary. An example of this approach is direct GMM estimation of the Euler equation. The GMM estimator exploits the lack of correlation between the discounted pay-off and the information set used in conditioning. The null hypothesis is that

$$E[(M_{t+1}(\theta)(R_{t+1}-1)I_t]=0 \tag{29}$$

or as

$$E[(M_{t+1}(\theta)(r_{t+1}-r_t^f)I_t]=0$$

where I_t is the information set and θ is the set of parameters to be estimated. The implication is that I_t contains no information about $M_{t+1}(\theta)(r_{t+1}-r_t^f)$, the discounted excess return in period $t+1$. Nevertheless, the results above suggest that unless the information set has time-varying volatility, it will be unlikely to prove suitable. Surprisingly perhaps, in practice, this has not usually been a consideration. We now turn to a more detailed consideration of econometric methods. If an estimate of the risk premium is required, GMM estimation of the Euler equation would not be appropriate.

3 Econometric Methods

To a large extent the choice of econometric methodology will be dictated by the SDF model being used. We discuss three different methods: the use of a VAR, multivariate conditional heteroskedasticity models, the efficient method of moments (a simulation estimator), and GMM.

This is a large subject in its own right. To make it manageable, we focus on their application to observed factor models. Another estimation method commonly used for latent variable affine factor pricing models is the Kalman filter as this does not require data on the factors.

3.1 VAR models

There is a vast literature on the use of VAR models in empirical finance, see for example the surveys of Campbell, Lo and MacKinlay (1997) and Cuthbertson (1996). The VAR is typically specified in terms of the excess return and other variables appropriate to the underlying theory, or in terms of the returns themselves together with any other variables. When a vector of returns is used the coefficient matrix of the VAR is restricted to satisy the no-arbitrage condition.

Despite its popularity it is our contention that, in general, using a VAR is not a valid way to estimate asset pricing models, or to test market efficiency. This is because for risk averse investors the no-arbitrage condition of asset pricing will require a risk premium and this involves conditional covariance terms. As a VAR does not include conditional covariances, in general, it is not an appropriate model to use.

There are two main exceptions to this for which a VAR may be appropriate. One is where risk neutrality is assumed, and so on the null hypothesis there is no risk premium. This is the assumption underlying the expectations or unbiasedness hypothesis. For a survey, see Bekaert and Hodrick (2001). Under the null hypothesis of unbiasedness none of the lags of the VAR should be non-zero in the equation for the excess return. When a vector of returns is used the unbiasedness hypothesis implies that the coefficient matrix of the VAR is restricted. The aim is then to test these restrictions.

It is not clear what can be inferred from a failure to reject the null hypothesis. It does not necessarily imply a lack of market efficiency, since if investors are risk averse, the model should have included a well-specified risk premium, or a suitable proxy for the risk premium. For an example of this, see Tzavalis and Wickens (1997) who show that rejection of the rational expectations hypothesis of the term structure (REHTS) is probably due to omitting the term premium. When the term premium is appropriately proxied in the VAR the REHTS is no longer rejected. The power of a test is likely to depend on whether a model is well specified on the alternative hypothesis. The absence of conditional covariance terms in these tests implies that the model is well not specified against the alternative hypothesis of risk aversion, and so may have low power.

Implictly, the lagged variables in the VAR are acting as proxies for the risk premium. The second exception we mention, therefore, is where the right-hand side of the VAR is a suitable proxy for the risk premium. Without a careful consideration of how good a proxy the VAR provides, it should be assumed that it is not appropriate. Tzavalis and Wickens (1997) illustrate how this might be achieved in tests of the REHTS. Another exception is where the risk premium can be shown to be a linear function of the variables in the VAR. Although hardly ever exploited in the literature, it is shown below that the Vasicek and CIR affine factor models of the term structure have this implication. In the Vasicek model the risk premium is constant over time. In the CIR model it is a linear function of the yields.

3.2 Multivariate conditional heteroskedasticity models

Two important implications follow from equations (6) and (10). First, in general, econometric models that satisfy the no-arbitrage condition must have (time-varying) conditional covariance terms in the mean. Second, since covariance is a multi-variate concept, the econometric model will need to be multivariate and not univariate, and the conditional covariance will need to be modelled at the same time as everything else. These requirements eliminate much of the literature on financial econometrics from further consideration.

An obvious econometric model that does satisfy these requirements is the multivariate GARCH-in-mean model, but not of course the multivariate GARCH model.[2] GARCH is preferred to ARCH on the grounds of its greater flexibility, and on the basis of previous empirical findings. We illustrate the use of the multivariate GARCH-in-mean (MGM) model under the assumption of log-normality. It is also possible to use other distributions such as the t-distribution, see Wickens and Smith (2001).

Many practical problems arise in using the MGM model. One is the computational problem of achieving numerical convergence due to the large number of parameters that need to be estimated. As a result, a trade-off arises between choosing a model that has sufficient flexibility, and one that is sufficiently parsimonious to be estimable. A second problem, is the availability of suitable observable factors. We would like to be able to identify the fundamental sources of risk, and for the most part these will be macroeconomic. The problem is that a time-varying risk premium requires conditional heteroskedasticity both in the excess return and the macroeconomic factors. Even for returns, conditional heteroskedasticity tends to be observable only at frequencies of a month or higher (eg stock returns). There is very little macroeconomic data at frequencies higher than quarterly. The main macroeconomic series likely to prove useful for our purposes that are available monthly are industrial production, retail sales, consumer price inflation and the money

[2] A further illustration of the general lack of appreciation of the requirements is that FANPAC, a recent computer package especially designed for financial analysis and based on GAUSS. does not contain a single program that satisfies these requirements.

supply.

In principle, a multivariate stochastic volatility model could be used instead of the MGM model. The advantage of this would be the greater flexibility it offers in modelling volatility. When volatility is fluctuating violently - as it can over short periods - the coefficients of the GARCH process describing volatility struggle to capture these fluctuations. As a result, the GARCH process often displays a unit root in the variance. The advantage of a stochastic volatility model is that it includes a disturbance in the volatility process. This can be used to absorb extreme fluctuations in volatility, which should be of help computationally. The conceptual value of this is perhaps less clear. A practical problem is that, to our knowledge, to date no multivariate stochastic volatility model which includes the conditional second moments in the mean has been proposed in the literature.[3]

For a single asset, the MGM model must be capable of satisfying the no-arbitrage condition, equations (23) or (24) where the factors z_{it} are assumed to be observable. It follows that we require an MGM in the vector $\mathbf{x}_{t+1} = (r_{t+1} - r_t^f, z_{1,t+1}, z_{2,t+1}, ...)'$. We now require the conditional distribution of $\mathbf{x}_{t+1}$. In practice, initially we would specify the joint distribution of $\mathbf{x}_{t+1}$ and any other variables included in the information set that are used to form conditional expectations. Where these are not simply past values of $\mathbf{x}_{t+1}$, it will be necessary to augment $\mathbf{x}_{t+1}$ with these additional variables. This will ensure that the conditional distribution of $\mathbf{x}_{t+1}$ (including the conditional covariance terms) is well specified. The MGM model can be written

$$\mathbf{x}_{t+1} = \boldsymbol{\alpha} + \boldsymbol{\Gamma}\mathbf{x}_t + \boldsymbol{\Phi}\mathbf{g}_t + \varepsilon_{t+1} \tag{30}$$

where the distribution of ε_{t+1} conditional on information available at time t, I_t, is

$$\boldsymbol{\varepsilon}_{t+1} \mid I_t \sim N[0, \mathbf{H}_{t+1}] \tag{31}$$

$$\mathbf{g}_t = vech\{\mathbf{H}_{t+1}\} \tag{32}$$

The *vech* operator converts the the lower triangle of a symmetric matrix into a vector. Equation (24) imposes no-arbitrage restrictions on the first equation of (30). Thus the first row of $\boldsymbol{\Gamma}$ is zero and the first row of $\boldsymbol{\Phi}$ is $(-\frac{1}{2}, -b_{11}, -b_{12}, -b_{13}, ...)$. Contrasting the MGM model with a VAR, we note that in a VAR $\boldsymbol{\Phi}$ is implictly zero and $\mathbf{H}_{t+1}$ is assumed to be homoskedastic.[4]

[3] Stochastic volatility models have been used for univariate models of returns. Interestingly, it can be shown that a model with an AR(1) volatility process can be transformed into an ARMA(1,1) model in the return with homoskedastic errors. If this were a valid no-arbitrage condition it would imply that the risk premium is a linear function of the past return and the past forecast error of returns.

[4] It may be noted that the factors may themselves be jointly determined with other variables. It could, therefore, be argued that the joint distribution should also include these other variables, but the variables themselves should be constrained from entering the equation for the excess return.

The remaining specification issue is how to choose $\mathbf{H}_{t+1}$. There have been several good surveys of these issues, see for example Bollerslev (2001). A model with considerable generality is the BEKK model described and generalized in Engle and Kroner (1995). This can be formulated as

$$vech(\mathbf{H}_{t+1}) = \mathbf{\Lambda} + \Sigma_{i=0}^{p-1}\mathbf{\Phi}_i vech(\mathbf{H}_{t-i}) + \Sigma_{j=0}^{q-1}\mathbf{\Theta}_j vech(\varepsilon_{t-j}\varepsilon_{t-j}') \tag{33}$$

where the matrices $\mathbf{\Lambda}, \mathbf{\Phi}$ and $\mathbf{\Theta}$ may be unrestricted. If there are $n-1$ factors z_{it} then $\mathbf{\Phi}$ and $\mathbf{\Theta}$ are both square matrices of size $n(n+1)/2$ and $\mathbf{\Lambda}$ is a size $n(n+1)/2$ vector.

A variant of the BEKK model that ensures the time-varying covariance matrices are symmetric and positive definite, and involves far fewer coefficients, is to specify the conditional covariance matrix as an error correction model (ECM):

$$\mathbf{H}_{t+1} = \mathbf{V}'\mathbf{V} + \mathbf{A}'(\mathbf{H}_t - \mathbf{V}'\mathbf{V})\mathbf{A} + \mathbf{B}'(\varepsilon_t\varepsilon_t' - \mathbf{V}'\mathbf{V})\mathbf{B}. \tag{34}$$

The first term on the right-hand side of equation (34) is the long-run, or unconditional, covariance matrix. The other two terms capture the short-run deviation from the long run. This formulation enables us to see more easily how volatility in the short run differs from that in the long run. To reduce the number of parameters further, we can specify $\mathbf{V}$ to be lower triangular and $\mathbf{A}$ and $\mathbf{B}$ to be symmetric matrices.

A specification that involves even fewer parameters is the constant correlation model. This has been found by Ding and Engle (1994) to give a fairly good performance in comparison with the more general BEKK model. It can be written

$$\mathbf{h}_{ij,t+1} = \rho_{ij}[\mathbf{h}_{ii,t+1} \times \mathbf{h}_{jj,t+1}]^{\frac{1}{2}} \tag{35}$$

$$\mathbf{h}_{ii,t+1} = v_i + a_i\mathbf{h}_{ii,t} + b_i\boldsymbol{\varepsilon}_{it}{}^2 \tag{36}$$

where ρ_{ij} is the (constant) correlation between $\boldsymbol{\varepsilon}_i(t+1)$ and $\boldsymbol{\varepsilon}_j(t+1)$. The conditional variances $\mathbf{h}_{ii}(t+1)$ each have a GARCH(1,1) structure.

It is instructive to compare the number of parameters involved in each of these formulations. If $n = 3$ and $p = q = 1$, then we find that BEKK= $n(n+1)/2 + (p+q)n^2(n+1)^2/4 = 78$, ECM unrestricted $= 3n^2 = 27$, ECM restricted $= 3n(n+1)/2 = 18$, and constant correlation $= 3n + n(n-1)/2 = 12$. Ideally, we would choose the most general model, but it is clear that this involves estimating a very large number of parameters. The ECM may be a useful compromise. But the constant correlation model may be the best that one can achieve. Further variants of these models can, of course, be considered.

3.3 Efficient Method of Moments (EMM)

The EMM estimator is an alternative way to estimate equations with unobservable variables. It is particularly suitable for more general affine models than the Vasicek model, for non-Gaussian distributions models and where dependencies in the processes for the (unobserved) state variables make evaluation of the likelihood function impossible. Estimation of the CIR affine model is such a case. Simulation-based method of moments estimators have been developed to analyse dynamic highly non-linear models such as these. Variants of these methods were originally proposed by Gallant and Tauchen (1996) (Efficient Method of Moments), Duffie and Singleton (1993) and Ingram and Lee (1991) (Simulated Method of Moments) and Smith A. (1993) and Gourieroux, Monfort and Renault (1993) (Indirect Inference).

The EMM method proceeds in two stages. In the first stage, an auxiliary model to the theoretical (or structural) model is chosen whose variables are observable and whose likelihood can be evaluated. The score of the likelihood for this model can then be evaluated using a GMM criterion where data for the dependent variables of interest (here the returns) is generated by simulation of the dynamic non-linear structural model. If the auxiliary model is a "good" approximation of the structural model, Gallant and Long (1997) show that the EMM estimator will be asymptotically efficient. If the structural model is close to the true model, the GMM criterion will be close to zero.

Using the affine models above as an example, let $f(y_t \mid x_{t-1}, \Omega)$ be the auxiliary model where the returns y_t are determined by past values of returns $x_{t-1} = (y_{t-1}, y_{t-2}, ...y_{t-j})$ with parameter Ω. This could be thought of as a reduced-form equation. The score of the maximised likelihood function for this model can be evaluated for two sets of data for current and past values of the returns. First, for the actual data we know that the average of the scores must be zero ie

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\partial}{\partial\Omega}\ln f(y_t \mid x_{t-1}, \Omega) = 0 \tag{37}$$

as these are the first-order conditions for a maximum. The EMM proceeds by replacing both y_t and x_{t-1} with values $\tilde{y}_t$ and $\tilde{x}_{t-1}$generated by simulation of the structural model $g(y_t, x_{t-1}, \beta)$ for a particular set of parameters β . The resulting scores are then

$$S(\beta, \tilde{\Omega}) = \frac{1}{T}\sum_{t=1}^{T}\frac{\partial}{\partial\Omega}\ln f(\tilde{y}_t \mid \tilde{x}_{t-1}, \tilde{\Omega}) \tag{38}$$

EMM then proceeds by choosing estimates of structural parameters β that minimise the GMM criterion:

$$S'(\beta, \tilde{\Omega}).W.S(\beta, \tilde{\Omega}) \tag{39}$$

for a weighting matrix W.

Recent research reviewed by Gallant and Tauchen (2001) shows that the choice of auxiliary function $f(y_t \mid x_{t-1}, \Omega)$ is crucial to the performance of the EMM estimator. On the one hand it needs to be computationally simple enough to make repeated solution not excessively burdensome. On the other hand it needs to be close to the theoretical model to deliver maximum likelihood estimates - were they to exist. Gallant and Tauchen (1996, 2001) propose the use of a semi-non parametric (SNP) auxiliary function. Others have concentrated on choosing a function which matches characteristics of the data. In evaluating various affine models of the term structure Dai and Singleton (2000) choose an SNP function they describe as a "non-Gaussian, VAR(1), ARCH(2), Homogenous-Innovation" function. This attempts to capture the non-Gaussian dependencies in bond returns.

The evaluation of models estimated by EMM can be more intuitive than for a likelihood-based method in that elements of the score can be examined along with the t-statistics associated with them. The testing of individual components of the score allows particular areas of model weakness to be identified. Following a similar logic, van der Sluis (1998) presents various tests of structural stability for the EMM estimator.

The computational burden of simulating the full structural model for each candidate set of structural parameters has limited the scale of the models which have been evaluated using this technique. A closely associated, more limited information, estimation method is direct estimation by GMM but evaluating a limited set of moments.

3.4 Generalised Method of Moments (GMM)

GMM is a non-linear instrumental variables estimator originally proposed by Hansen (1982) and widely employed in estimating affine and other non-linear asset pricing models. It is particularly suited to the direct estimation of the Euler equation and avoids the need to solve for the risky rate of return. The robustness of FGMM to heteroskedasticity has encouraged its use with financial data. Comprehensive presentations of GMM are available in Hall (1992) and Hamilton (1994).

In implementing the estimation of equation (40) by GMM, the information set I_t consists of a set of instruments h_t. We define the set of moments that are being set equal to zero as $E[f_t(\theta)]$. Hence,

$$E[f_t(\theta)] = E[M_{t+1}(\theta)(R_{t+1} - 1) \otimes h_t] = 0 \tag{40}$$

where $\otimes$ is the Kronecker product. The model and the errors have the property that there exists a true set of parameters θ_0 for which the error vector is orthogonal to the set of instruments h_t. Defining the sample average of $f_t(\theta)$ as $g_T(\theta) = T^{-1} \sum_{t=1}^{T} f_t(\theta)$, the GMM estimator minimises

$Q_T(\theta) \equiv g_T(\theta)'.W_T.g_T(\theta)$ where $W_T = \frac{1}{T}\sum f_t(y_t,\tilde{\theta}).f_t(y_t,\tilde{\theta})'$ is a weighting matrix and $\tilde{\theta}$ is a consistent estimate of θ. The first-order conditions are $D_T(\hat{\theta}_T)'.W_T.g_T(\hat{\theta}) = 0$ where $D_T(\theta_T) = \partial g_T(\theta)/\partial\theta$. The variance of the estimator of θ is $\Omega = ((D_T(\hat{\theta}_T)'.W_T.D_T(\hat{\theta}_T))^{-1}$.

Given that W_T must itself be estimated, asymptotically efficient estimates can be obtained in two steps. First an arbitrary weighting matrix W can be used to obtain an initial set of consistent estimates $\tilde{\theta}$. Second, these estimates can be used to construct a second weighting matrix W_T and thence efficient estimates. It has been shown that in large-scale systems the finite sample performance of GMM can be improved by further iterations (see Ferson and Foerster (1994)). A further attractive feature of GMM is that the error ϵ_t can be heteroskedastic and autocorrelated. In the case of autocorrelation, the adjustment to the weighting matrix W_T proposed by Newey and West (1987) is commonly employed, although other weighting schemes have been proposed, and are canvassed by Hamilton (1994).

The leading specification test for models estimated by GMM is the J-test of over-identifying restrictions proposed by Hansen (1982). This is easily constructed from any remaining orthogonality conditions remaining beyond the number of estimated parameters. Further hypothesis tests based on the evaluation of individual orthogonality conditions can also be constructed. For example, structural parameter stability can be assessed in a similar way to EMM. Examples of applications of this approach to latent variable models are Gibbons and Ferson (1985) to equity markets, Campbell (1987) to stock returns and the term structure, and Smith P. (1993) to FOREX and the term structure.

In applications such as the latent variable model, the lack of guidance from theory on the number and type of instruments may potentially lead to lack of power of the $J-$test in rejecting the model restrictions when wrong. A price sometimes paid for the robustness is bias in the estimates and the test statistics arising from a version of the "weak instruments" problem, see Staiger and Stock (1997).

4 SDF models of the term structure

The SDF model has been used in the most part to price bonds, and to model the term structure. Of the different versions of the SDF model mentioned earlier, the Vasicek and CIR models are usually used. There is a vast literature on this that has been surveyed by Shiller (1990), Marsh (1995) and Cochrane (2001). The present discussion will focus mainly on the use of latent variable affine factor models as there seems to be little in the literature on the use of observable factors models at present. We suggest how observable factors models might be used.

First, we specialise the previous discussion to the pricing of zero-coupon bonds. A bond with a coupon can be converted to one without. Given the paucity of bonds of different maturities available at any time, data on zero-coupon bond yields are usually a constructed series based on coupon bonds, and the use of interpolation methods to fill in missing maturities.

We define the following variables: $P_{n,t}$ = nominal price of a zero-coupon n-period bond at t (i.e. has n periods to maturity), $P_{0,t} = 1$ (i.e. the pay-off at maturity is 1), $R_{n,t}$ = yield to maturity on an n-period bond, $R_{1,t} = s_t$ (i.e. one period rate - the risk free rate), $h_{n,t+1}$ = return to holding an n-period bond for one period (i.e. from t to $t+1$). $R_{n,t}$ is defined as $P_{n,t} = \frac{1}{[1+R_{n,t}]^n}$ and $h_{n,t+1}$ is defined as $1 + h_{n,t+1} = \frac{P_{n-1,t+1}}{P_{n,t}}$. Using the logarithmic approximation $\ln(1+x) \simeq x$ for small x, $\ln P_{n,t} = -n\ln(1+R_{n,t}) \simeq -nR_{n,t}$ and $h_{n,t} \simeq p_{n-1,t+1} - p_{n,t} = nR_{n,t} - (n-1)R_{n-1,t+1}$, where $p_{n,t} = \ln P_{n,t}$.

$>$From the SDF model, $P_{n,t}$, the price of an n-period zero-coupon bond in period t is the discounted value of $P_{n-1,t+1}$, the price of the bond in $t+1$ when it has $n-1$ periods to maturity $P_{nt} = E_t[M_{t+1}P_{n-1,t+1}]$. Thus

$$E_t[M_{t+1}(1 + h_{n,t+1})] = 1$$

When $n = 1$ the bond is risk free and

$$(1 + s_t)E_t[M_{t+1}] = 1$$

Assuming that $P_{n,t}$ and M_{t+1} are jointly log-normal with $m_{t+1} = \ln M_{t+1}$,

$$\begin{aligned} p_{nt} &= E_t(m_{t+1} + p_{n-1,t+1}) + \frac{1}{2}V_t(m_{t+1} + p_{n-1,t+1}) \qquad (41) \\ &= E_t(m_{t+1}) + E_t(p_{n-1,t+1}) + \frac{1}{2}V_t(m_{t+1}) + \frac{1}{2}V_t(p_{n-1,t+1}) + Cov_t(m_{t+1}, p_{n-1,t+1}) \end{aligned}$$

Similarly, as $p_{o,t} = 0$,

$$p_{1,t} = E_t(m_{t+1}) + \frac{1}{2}V_t(m_{t+1}) \qquad (42)$$

Subtracting (42) from (41) and re-arranging gives the no-arbitrage equation

$$E_t(p_{n-1,t+1}) - p_{n.t} + p_{1,t} + \frac{1}{2}V_t(p_{n-1,t+1}) = -Cov_t(m_{t+1}, p_{n-1,t+1}) \qquad (43)$$

This can be re-written in terms of yields as

$$-(n-1)E_t(R_{n-1,t+1}) + nR_{n,t} - s_t + \frac{(n-1)^2}{2}V_t(R_{n-1,t+1}) = (n-1)Cov_t(m_{t+1}, R_{n-1,t+1})$$

and, since $h_{n,t} \simeq -(n-1)R_{n-1,t+1} + nR_{n,t}$, as

$$E_t(h_{n,t+1} - s_t) + \frac{1}{2}V_t(h_{n,t+1}) = -Cov_t(m_{t+1}, h_{n,t+1}) \qquad (44)$$

This is the fundamental no-arbitrage condition for an n-period bond. The term on the right-hand side is the term premium

Each point on the yield curve must satisfy this no-arbitrage condition. As $V_t(h_{n,t+1}) = V_t(h_{n,t+1} - s_t) = (n-1)^2 V_t(R_{n-1,t+1})$ and $Cov_t(m_{t+1}, h_{n,t+1}) = Cov_t(m_{t+1}, h_{n,t+1} - s_t) = -(n-1)Cov_t(m_{t+1}, R_{n-1,t+1})$, the term spread (the slope of the yield curve) for an n-period bond must comply with

$$\begin{aligned} R_{n,t} - s_t &= \frac{n-1}{n} E_t(R_{n-1,t+1} - s_{t+1}) + \frac{n-1}{n} E_t \Delta s_{t+1} \\ &\quad - \frac{1}{2} \frac{(n-1)^2}{n} V_t(R_{n-1,t+1}) - \frac{n-1}{n} Cov_t(m_{t+1}, R_{n-1,t+1}) \end{aligned} \tag{45}$$

The yield itself can be expressed as

$$\begin{aligned} R_{n,t} &= \frac{1}{n} \sum_{i=0}^{n-1} E_t s_{t+i} - \frac{1}{2n} \sum_{i=0}^{n-1} (n-i-1)^2 V_t(R_{n-i-1,t+i+1}) \\ &\quad - \frac{1}{n} \sum_{i=0}^{n-1} (n-i-1) Cov_t(m_{t+i+1}, R_{n-i-1,t+i+1}) \end{aligned}$$

This is a modification of the usual result that under risk neutrality the yield is the average of current and expected future short rates. From the Fisher equation $s_t = r_t + E_t \pi_{t+1}$, where r_t is the real return and π_{t+1} is the inflation rate, the shape of the yield curve is determined by average future value of real returns, the inflation rate and the risk premia on that bond until maturity.

4.1 Affine models of the term structure

4.1.1 The single factor affine model

In a single factor affine model we write $p_{n,t}$ as a linear function of the factor z_t:

$$p_{n,t} = -[A_n + B_n z_t]$$

The coefficients differ for each maturity but are related in such a way that there are no arbitrage opportunities across maturties. Below, we derive the restrictions implied by this. The yield to maturity is

$$R_{n,t} = -\frac{1}{n} p_{n,t} = \frac{A_n}{n} + \frac{B_n}{n} z_t$$

and the one-period risk-free rate is

$$s_t = -p_{1,t} = A_1 + B_1 z_t$$

In order to derive the restrictions on A_n and B_n we need to introduce an assumption about the process generating z_t. We consider first the CIR model.

The Cox-Ingersoll-Ross model (CIR) The CIR model assumes that

$$\begin{aligned} -m_{t+1} &= z_t + \lambda e_{t+1} && (46) \\ z_{t+1} - \mu &= \theta(z_t - \mu) + e_{t+1}. && (47) \end{aligned}$$

where $e_{t+1} = \sigma\sqrt{z_t}\varepsilon_{t+1}$ and $\varepsilon_{t+1}\tilde{}\, iid(0,1)$. Thus $V_t(e_{t+1}) = \sigma^2 z_t$. We now evaluate equation (43) using these assumptions. We note that

$$\begin{aligned} E_t[m_{t+1} + p_{n-1,t+1}] &= -[z_t + A_{n-1} + B_{n-1}E_t z_{t+1}] \\ &= -[z_t + A_{n-1} + B_{n-1}(\mu(1-\phi) + \phi z_t)] \end{aligned}$$

and

$$\begin{aligned} V_t[m_{t+1} + p_{n-1,t+1}] &= V_t[\lambda e_{t+1} + B_{n-1}e_{t+1}] = (\lambda + B_{n-1})^2 V_t(e_{t+1}) \\ &= (\lambda + B_{n-1})^2\sigma^2 z_t \end{aligned}$$

hence equation (41) becomes

$$\begin{aligned} -[A_n + B_n z_t] &= -[(1+\phi B_{n-1})z_t + A_{n-1} + B_{n-1}\mu(1-\phi)] + \frac{1}{2}(\lambda + B_{n-1})^2\sigma^2 z_t \\ &= -[A_{n-1} + B_{n-1}\mu(1-\phi)] - [1 + \phi B_{n-1} - \frac{1}{2}(\lambda + B_{n-1})^2\sigma^2)]z_t \end{aligned}$$

Equating terms on the left-hand and right-hand sides (i.e. the intercepts and the coefficients on z_t) gives the recursive formulae

$$A_n = A_{n-1} + B_{n-1}\mu(1-\phi)$$

$$B_n = 1 + \phi B_{n-1} - \frac{1}{2}(\lambda + B_{n-1})^2\sigma^2$$

Using $P_{0,t} = 1$ implies $p_{0,t} = 0$ and $A_0 = 0, B_0 = 0$. This enables us to use the two formulae to solve recursively for all A_n, B_n. For $n = 1$, $B_1 = 1 - \frac{1}{2}\lambda^2\sigma^2$ and $A_1 = 0$. Hence

$$\begin{aligned} s_t &= -p_{1,t} = A_1 + B_1 z_t \\ &= (1 - \frac{1}{2}\lambda^2\sigma^2)z_t && (48) \end{aligned}$$

The no-arbitrage condition, equation (43), is therefore

$$E_t[p_{n-1,t+1}] - p_{nt} + p_{1,t} = E_t[h_{n,t+1} - s_t] = \frac{1}{2}B_{n-1}^2\sigma^2 z_t + \lambda B_{n-1}\sigma^2 z_t$$

The first term is the Jensen effect and the second is the risk premium. Thus

$$\text{risk premium} = -Cov_t(m_{t+1}, p_{n-1,t+1}) = \lambda B_{n-1}\sigma^2 z_t$$

Both are linear functions of z_t. We note that if $\lambda = 0$ (ie if $m_{t+1} = -z_t$ and non-stochastic) then the risk premium is zero.

So far z_t has been treated as a latent variable and assumed to be observable. However, equation (48) shows z_t that it is a linear function of s_t which is observable. This implies that we can replace z_t everywhere by s_t. As a result $p_{n,t}$, $R_{n,t}$ and the risk premium can all be shown to be linear functions of s_t and the model becomes an SDF model with observable factors. Thus

$$p_{n,t} = -[A_n + B_n z_t] = -\left[A_n + \frac{B_n}{1 - \frac{1}{2}\lambda^2\sigma^2} s_t\right]$$
$$R_{n,t} = \frac{1}{n}[A_n + B_n z_t] = \frac{A_n}{n} + \frac{B_n}{n\left(1 - \frac{1}{2}\lambda^2\sigma^2\right)} s_t$$
$$\text{risk premium} = \lambda B_{n-1}\sigma^2 z_t = \frac{\lambda B_{n-1}\sigma^2}{1 - \frac{1}{2}\lambda^2\sigma^2} s_t$$

This implies that over time the shape of the yield curve is constant and the curve shifts up and down due to movements in the short rate. Since the shape of the yield curve actually varies over time, this single-factor CIR model is clearly not an appropriate model of the term structure.

The Vasicek model The Vasicek model is also based on equations (46) and (47), but it assumes that $e_t = \sigma\varepsilon_{t+1}$, i.e. e_t is homoskedastic. The analysis proceeds as for the CIR model. It can be shown that

$$V_t[m_{t+1} + p_{n-1,t+1}] = (\lambda + B_{n-1})^2\sigma^2$$

and equation (41) is

$$-[A_n + B_n z_t] = -[A_{n-1} + B_{n-1}\mu(1-\phi) + \frac{1}{2}(\lambda + B_{n-1})^2\sigma^2)] - [1 + \phi B_{n-1}]z_t$$

Thus $A_n = A_{n-1} + B_{n-1}\mu(1-\phi) + \frac{1}{2}(\lambda + B_{n-1})^2\sigma^2)$ and $B_n = 1 + \phi B_{n-1}$. Using $p_{0,t} = A_0 = B_0 = 0$ we obtain $B_1 = 1$ and $A_1 = \frac{1}{2}\lambda^2\sigma^2$. Thus

$$s_t = -p_{1,t} = A_1 + B_1 z_t = \frac{1}{2}\lambda^2\sigma^2 + z_t$$

Once again we can therefore re-write the yields as a linear function of s_t :

$$p_{n,t} = -[A_n + B_n z_t] = -[A_n - \frac{1}{2}\lambda^2\sigma^2 B_n] - B_n s_t$$
$$R_{n,t} = \frac{1}{n}[A_n + B_n z_t] = \frac{1}{n}[A_n - \frac{1}{2}\lambda^2\sigma^2 B_n] + \frac{B_n}{n} s_t$$

>From equation (43), and as $B_n = \frac{1-\phi^n}{1-\phi}$,

$$\begin{aligned} E_t[p_{n-1,t+1}] - p_{nt} + p_{1,t} &= E_t[h_{n,t+1} - s_t] \\ &= \frac{1}{2}B_{n-1}^2\sigma^2 + \lambda B_{n-1}\sigma^2 \\ &= \frac{1}{2}(\frac{1-\phi^n}{1-\phi})^2\sigma^2 + \frac{\lambda(1-\phi^{n-1})\sigma^2}{1-\phi} \end{aligned}$$

$$\text{risk premium} = \frac{\lambda(1-\phi^{n-1})\sigma^2}{1-\phi}.$$

Thus, like the CIR model, yields are linear functions of the short rate, and the shape of the yield curve is constant through time and shifts due to changes in the short rate. But unlike the CIR model, the risk premium depends only on the time to maturity and not on time itself. The Vasicek model is, therefore, an even more unsatisfactory way to model the term structure than the single-factor CIR approach. Clearly, both specifications should be rejected *a priori*.

4.1.2 Multi-factor affine models

One of the problems with these CIR and Vasicek models is that they are single factor models. A multi-factor affine CIR model may be better able to capture both changes in the shape of the yield curve over time and shifts. A two factor CIR model of the term structure has been used by Gong and Remolona (1997) and by Remolona, Wickens and Gong (1998). Dai and Singleton (2000) have used a three factor continuous-time CIR model. The evidence suggests that although two factors capture most of the behaviour of the yield curve, a third may still prove to be significant.

The multi-factor CIR model of the term structure proposed by Dai and Singleton can be written in discrete time as:

$$p_{n,t} = -[A_n + \mathbf{B}_n'\mathbf{z}_t] \tag{49}$$

where $\mathbf{z}_t$ is a vector of factors. The stochastic discount factor is

$$\begin{aligned}
-m_{t+1} &= \ell'\mathbf{z}_t + \boldsymbol{\lambda}'\mathbf{e}_{t+1} \\
\mathbf{z}_{t+1}-\boldsymbol{\mu} &= \boldsymbol{\theta}(\mathbf{z}_t-\boldsymbol{\mu}) + \mathbf{e}_{t+1} \\
\mathbf{e}_{t+1} &= \Sigma\sqrt{\mathbf{S}_t}\boldsymbol{\varepsilon}_{t+1} \\
\mathbf{S}_{ii,t} &= \nu_i + \phi_i'\mathbf{z}_t
\end{aligned}$$

where ℓ is a vector of ones, $\mathbf{S}_t$ is a diagonal matrix, $\boldsymbol{\varepsilon}_{t+1}$ is $i.i.d(0,I)$, and θ and Σ are both square matrices. Gong and Remolona and Remolona, Wickens and Gong use particular versions of this.

To illustrate, we consider the case where the factors are independent so that $\boldsymbol{\theta}$ and Σ are diagonal matrices. Also, we set $\mathbf{S}_{ii,t} = z_{it}$. The whole model is therefore additive. As result, it can be shown that

$$E_t[p_{n-1,t+1}] - p_{nt} + p_{1,t} = \frac{1}{2}\sum_i B_{i,n-1}^2\sigma_i^2 z_{it} + \sum_i \lambda_i B_{i,n-1}\sigma_i^2 z_{it}$$

$$\text{risk premium} = \sum_i \lambda_i B_{i,n-1}\sigma_i^2 z_{it}$$

Hence the risk premium is the sum of the risk effects associated with each factor. Further, the short rate is a linear function of both factors

$$s_t = -p_{1,t} = \frac{1}{2}\sum_i (\lambda_i^2 \sigma_i^2 + z_{it})$$

which implies that the yields and the term premia can no longer be written as a linear function of the short rate. Since every yield is a linear function of the factors, if there are n factors, it would require the short rate plus $n-1$ further yields to represent the factors. If there are more than n yields (including the short rate) then the factors would not be a unique linear function of the yields. A way would then be needed to project the yields onto the lower dimension factor space. A possibility is principal components. If there were less than n yields no observable representation of the factors would be possible, and the model could not be re-interpreted as an observable factor model. If each yield is a linear function of two or more reference yields then the shape of the yield curve can now change over time, although only with limited additional flexibility.

We may summarise our findings as far as the term structure is concerned in the following way. The SDF model commonly used is the affine term structure model factor. This requires at least two factors to adequately represent the yield curve. But having more factors than yields implies that the model cannot necessarily be interpreted as an observable factor model.

4.2 Empirical Evidence on Affine Models

4.2.1 Vasicek models

Estimation of the Vasicek model is typically carried out by Kalman filter methods in order to obtain maximum likelihood estimates. Hamilton (1994) presents a comprehensive treatement of these recursive methods which rely on assumption of Gaussian errors and are therefore not suitable for estimation of the more general CIR-type models. Jegadeesh and Pennacchi (1996) employ the Kalman filter to estimate the parameters of single and two-factor Vasicek models. The single factor model is given in (46 and 47) with $e_t = \sigma\varepsilon_t$, whilst the two factor model is a restricted version of (49). The motivation for the second factor in Jegadeesh and Pennacchi (1996) is as a time-varying mean, or 'target' for the short-term interest rate, the observable single factor of the Vasicek model. The structure of the model for the state variables is therefore:

$$z_{1t+1} = (1-\theta_1)\mu_1 + \theta_1 z_{1t} + \sigma_1 \varepsilon_{1t+1} \quad (50)$$

$$z_{2t+1} = (1-\theta_2) z_{1t} + \theta_2 z_{2t} + \sigma_2 \varepsilon_{2t+1} \quad (51)$$

where the short-term interest rate is z_{2t} and the 'target' or 'central tendency' is z_{1t}. The data set employed is monthly sampled three-month Eurodollar futures prices based on the 90-day LIBOR for contracts of four different maturities from 1992 to 1995 (ie n=1,..,4.). The maturities are concentrated at the short end but extend up to five years. This is an actively traded market and, the authors argue, less susceptible to the consequences of thin trading seen in some of the bond markets which have been used to construct term structure data. Whilst many of the parameters of the single factor model are significantly different from zero, the steady-state level of the state variable, the short interest rate is not. The point estimate is also counter-intuitively negative. The single state variable is found to mean-revert at a very slow rate. The estimate of 0.934 implies a half-life of 10.5 yars. This implied borderline unit-root behaviour of interest rates seems counter-intuitive. The estimate of the price of risk is positive and close to one. The implication is that the risk premium, which is constant over time, is positive and increases with maturity, and the yield curve is upward sloping.

In the two factor model, the coefficient is positive. A likelihood ratio test of $\theta_2 = \infty$, the restriction implied by the single factor Vasicek model, is rejected at less than 1% significance. Mean reversion in the short rate is somewhat closer to that found elsewhere in the literature, with a half-life closer to four years. The short rate reverts somewhat faster to its central tendency according to these estimates. Given these adjustment speeds, any interpretation of these factors as reflecting the conduct of monetary policy seem unreliable - contrary to the claims of Jegadeesh and Pennacchi. The estimate of the mean interest rate to which these state variables converge is positive but, as for the single-factor model, not well determined. This, taken together with the theoretical drawbacks of the Vasicek model discussed above, suggests that a more general specification may further improve the model.

4.2.2 General affine models with unobservable factors

As noted, the empirical literature has gravitated to the conclusion that three factor models probably contain enough flexibility to fit the yield curve to most term structure data sets. Heteroskedasticity in the term structure along with the theoretical considerations discussed above have motivated the analysis of multi-factor CIR models. Dai and Singleton (2000) provide the most complete statistical comparison of various three factor models. They provide tests of models with both one and two sources of conditional volatility amongst the three affine factors. Here we focus on the single source case, that preferred by Dai and Singleton. To be compatible with our earlier discussion, we present discrete-time versions of their models. The preferred structure of the model of the three state variables is

$$z_{1t+1} = (1-\theta_1)\mu_1 + \theta_1 z_{1t} + \sigma_1 \sqrt{z_{1t}} \varepsilon_{1t+1} \quad (52)$$

$$z_{2t+1} = (1-\theta_2)\mu_2 + \theta_2 z_{2t} + \sqrt{\nu_2 + z_{1t}} \varepsilon_{2t+1} + \sigma_{23} \sqrt{z_{1t}} \varepsilon_{3t+1} \quad (53)$$

$$z_{3t+1} = (1-\theta_3)z_{2t} + \theta_3 z_{3t} + \sqrt{z_{1t}}\varepsilon_{3t+1} + \sigma_{31}\sigma_1\sqrt{z_{1t}}\varepsilon_{1t+1} + \sigma_{32}\nu_2\varepsilon_{2t+1} \tag{54}$$

where the short-term interest rate is z_{3t}, the 'central tendency' is z_{2t}, and the 'volatility' factor z_{1t} is the sole source of conditional volatility. A similar three-factor model is proposed by Balduzzi et al. (1996) with the restrictions $\sigma_{23} = \sigma_{32} = 0$ (i.e. the conditional correlation between the short rate and the 'central tendency' is zero). This structure can be incorporated into the stochastic discount factor through equation (49).

Dai and Singleton (2000) estimate the three-factor model on a dataset of yields on plain-vanilla, fixed-for-variable rate US dollar swap contracts. The data are weekly from April 1987 -August 1996 on three yields; 6 month LIBOR and 2 and 10-year fixed-for-variable rate swaps. The continuity of swap yield data making them preferable to treasury rates. The estimation method employed is EMM. To capture the non-Gaussian dependencies in bond returns, Dai and Singleton choose a semi-non-parametric function as the auxiliary model. Interestingly, this includes forms of heteroscedasticity, such as ARCH, that are more general than those assumed in the strucutural affine model being estimated. Whether a more general specification not constrained to be affine would be preferable is still an open question.

4.2.3 Vasicek Model with Observable Macro Factors

Ang and Piazzesi (2000) use a modification of the multi-factor Vasicek model discussed above with additional observable macroeconomic factors, but with Gaussian errors.The structure of the model is

$$\begin{aligned}
-m_{t+1} &= \ell'\boldsymbol{\delta}'\mathbf{z}_t + \boldsymbol{\lambda}'\mathbf{e}_{t+1} \\
\mathbf{z}_{t+1} - \boldsymbol{\mu} &= \boldsymbol{\theta}(\mathbf{z}_t - \boldsymbol{\mu}) + \mathbf{e}_{t+1} \\
\mathbf{e}_{t+1} &= \Sigma\boldsymbol{\varepsilon}_{t+1} \\
s_t &= \boldsymbol{\delta}'\mathbf{z}_t
\end{aligned}$$

where $\mathbf{z}_t = (z_t^o, z_t^u)'$ includes observable factors z_t^o and unobservable factors z_t^u, and the short interest rate is assumed to be a linear function of the factors. Ang and Piazzesi choose a specification with two observable macroeconomic and three unobservable factors. The macroeconomic factors are inflation and real activity factors created as principal components from a larger set of macroeconomic data. The factors are assumed to follow ARMA processes, the parameterisation of which is dealt with in the construction of the θ matrix.

Ang and Piazzesi estimate a restricted version of the model in which the price of risk is zero. The model is estimated in two steps. The coefficients on the observed factors are first estimated by GMM. Those for the unobserved factors are estimated by maximum likelihood in a manner similar to that for the Vasicek models discussed above. Their term structure data are bond yields of maturities 1, 6, 12, 36 and 60 months. In order to identify the three unobservable factors, they assume that the 1, 12 and 60 month yields are measured without error.

The macroeconomic factors account for about 30% of of the forecast variance of the mid-range of the term structure, but somewhat less at the longer end. The estimates of the processes generating the state variable show strong persistence in two of the three unobservable state variables. These three factors are assumed to be independent of each other as well as of the two macroeconomic variables which by construction are independent of each other. Ang and Piazzesi report that the persistence in the two most persistent unobservables in a standard multi-factor Vasicek model is not significantly affected by adding the macroeconomic factors. But the third factor becomes more persistent. Both macroeconomic factors and all three unobservable factors are found to have a significant impact on the short interest rate and hence the term structure. They are not independent of the unobservable factors. There is a clear statistical relationship between the two most persistent of the unobservable factors from the Vasicek model and the macroeconomic factors of the full model suggesting that the macroeconomic factors are picking up part of the behaviour of the unobservable factors.

The incorporation into affine term structure models of observable macroeconomic factors in adition to the short interest rate is a clear step forward. However, the restrictions implied by the Vasicek model - even one that is multi-variate with independent factors - limit the credibility of the results.

4.3 Nominal and real term structures and inflation risk

In some countries data exist on indexed as well as on nominal yields. The UK has the longest data set of indexed yields and the few studies available on indexed bonds have used these data. An additional interest in these data is that, when combined with nominal yields, it is possible, from financial data alone, to extract an estimate of future inflation, and of the inflation risk premium. Evans (1998) and Remolona, Wickens and Gong (1998) have estimated SDF models of the real term structure. In the UK, indexation occurs with a lag of eight months, so indexed yields are not exactly the same as real yields. We illustrate the methodology by describing the approach of Remolona, Wickens and Gong (RWG). Evans takes a very different approach and does not use affine factor models.

RWG assume a two-factor CIR model for nominal yields and a one-factor CIR model for indexed yields. Thus, for the logarithm of nominal bond prices,

$$p_{n,t}^{N} = -[A_n^N + B_{1,n}^N z_{1t} + B_{2,n}^N z_{2t}]$$

and for the logarithm of indexed prices

$$p^r_{n,t} = -[A^r_n + B^r_n z_{1t}]$$

The pricing kernel for indexed bonds is

$$-m^r_{t+1} = z_{1t} + \lambda_1 e_{1,t+1}$$

and that for nominal bonds is

$$-m^N_{t+1} = -m^r_{t+1} + z_{2t} + \lambda_2 e_{2,t+1}$$

RWG assume that e_{1t} and e_{2t} are uncorrelated. The model is therefore additively separable. The model for the nominal yields is that for the indexed yields plus an independent factor that can be interpreted as an inflation factor. The factors are assumed to be generated by generalised CIR models

$$z_{i,t+1} - \mu_i = \theta_i(z_{it} - \mu_i) + e_{i,t+1}.$$

$e_{i,t+1} = \sigma_i\sqrt{1+\nu_i z_{it}}\varepsilon_{i,t+1}$ and $\varepsilon_{i,t+1}\tilde{}\,iid(0,1)$ for $i = \{1,2\}$. Recursive relations can be derived for the coefficients of the price equations, and expressions for the risk premia can be obtained as above. The risk premium for nominal yields is the sum of the risk premia for indexed yields and inflation. The pricing equations can be combined into a multi-variate model where the vector of nominal and indexed yields are linear functions of the two factors. As the factors are unobservable, this model together with the equations for the two factors, was estimated using the Kalman filter.

Information about inflation can be extracted from this model by noting that the nominal and indexed yields to maturity satisfy

$$R_{n,t} = \frac{1}{n}\sum_{i=0}^{n-1} E_t R_{1,t+i} + \frac{1}{n}\sum_{i=0}^{n-1} E_t \varphi^N_{1,t+i}$$

$$r_{n,t} = \frac{1}{n}\sum_{i=0}^{n-1} E_t r_{1,t+i} + \frac{1}{n}\sum_{i=0}^{n-1} E_t \varphi^r_{1,t+i}$$

where $R_{n,t}$ and $r_{n,t}$ are the yields on n-period nominal and indexed bonds, $\varphi^N_{n,t}$ and $\varphi^r_{n,t}$ are their combined risk premia and Jensen effects. Using the Fisher equation $R_{1,t} = r_{1,t} + E_t\pi_{t+1}$, where π_{t+1} is the inflation rate, and noting that if $\varphi^\pi_{n,t}$ is the sum of the Jensen effect and inflation risk premium, then $\varphi^\pi_{n,t} = \varphi^N_{n,t} - \varphi^r_{n,t}$, and hence the average expected inflation over the next n periods is

$$\frac{1}{n}\sum_{i=0}^{n-1} E_t\pi_{t+i} = R_{nt} - r_{nt} - \frac{1}{n}\sum_{i=0}^{n-1} E_t\varphi^\pi_{1,t+i} \tag{55}$$

To estimate future inflation, it is necessary to subtract the average expected value of the inflation risk premium plus the Jensen effect from the difference between the nominal and indexed yields, rather than use the difference itself. The inflation risk premium varies over time and is estimated to be around 1% for the UK in the 1990's.

Evans (1998) proceeds very differently. He estimates equation (55) directly using as the dependent variable data on inflation over the next n-periods. The yields on n-period nominal and indexed bonds are the explanatory variables and the inflation risk premium is not taken into account. A variant is based on using the VAR approach to proxy the risk premia in which the risk premium is taken into account. The VAR is defined in terms of all of these variables. Thus Evans does not specify the stochastic discount factors explicitly. From the results above, however, it may be noted that if the factors were specified and the CIR model used, then the risk premia would in fact be a linear function of nominal and indexed yields.

4.4 Observable factors using the MGM model

We complete our discussion of SDF models of the term structure by considering how one might go about using the approach based on observable factors and the MGM model. We have shown that the no-arbitrage condition for bond yields is equation (44) or equivalently, equation (45). If we also assume that the log discount factor is generated by

$$m_{t+1} = a + \sum_i b_i z_{i,t+1}.$$

then the no-arbitrage condition for bond yields can be written as

$$E_t(h_{n,t+1} - s_t) + \frac{1}{2}V_t(h_{n,t+1}) = -\sum_i b_i Cov_t(z_{i,t+1}, h_{n,t+1})$$

or as

$$\begin{aligned} R_{n,t} - s_t &= \frac{n-1}{n}E_t(R_{n-1,t+1} - s_{t+1}) + \frac{n-1}{n}E_t \Delta s_{t+1} \\ &\quad -\frac{1}{2}\frac{(n-1)^2}{n}V_t(R_{n-1,t+1}) - \frac{n-1}{n}\sum_i b_i Cov_t(z_{i,t+1}, R_{n-1,t+1}) \end{aligned}$$

This suggests that we model the joint distribution of $\mathbf{x}_{t+1} = (h_{n_1,t+1} - s_t, h_{n_2,t+1} - s_t, ..., z_{1,t+1}, z_{2,t+1}, ...)'$ as an MGM model defined by equations (30), (31) and (32), where we include the maturities $n_1, n_2, ...$ in the model.

As yields are nominal, using CCAPM with power utility would give the log discount factor

$$m_{t+1} = \theta - \sigma \Delta \ln C_{t+1} - \pi_{t+1}$$

. Thus the factors would consist of just consumption growth and inflation. There is also an argument for including Δs_{t+1} as one of the factors as this is an independent source of variation. Other possible factors are output, and money growth, the exchange rate, and for countries other than the US, perhaps the change in the US short rate. We are currently pursuing this approach.

5 SDF Models of the Foreign Exchange Rate Risk Premium

5.1 The Forward Premium Puzzle

The SDF approach has only recently been applied to modelling the FOREX. Surveys of the FOREX market have been written by Lewis (1995) and Engel (1996) following earlier coverage of futures as well as forward markets by Hodrick (1987). In this area modelling has become focused on potential solutions to what is known as the forward premium puzzle. Before analysing SDF models, we first outline the nature of the puzzle.

Consider two countries (domestic and foreign) each of which issues a one-period bond that is risk-free in terms of its own currency. Let R_{t+1} denote the excess return to domestic investors from investing at time t in the foreign bond. Thus

$$R_{t+1} = i_t^* + \Delta s_{t+1} - i_t \tag{56}$$

where i_t and i_t^* are the domestic and foreign one-period nominal interest rates, respectively and s_t is the logarithm of the domestic price of foreign exchange. If investors are risk neutral and rational then the expectation of R_{t+1} conditional on information at time t is $E_t[R_{t+1}] = 0$, the uncovered interest parity condition. But if investors are risk averse then $E_t[R_{t+1}] = \phi_t$, where, for the moment, ϕ_t will be given the interpretation of a risk premium. If only domestic investors are exposed to exchange risk (i.e. foreign investors hold only their own bond) then $\phi_t \geq 0$. The sign is reversed if only foreign investors are exposed to exchange risk. More generally, ϕ_t can be positive or negative depending on the relative magnitudes of these portfolio composition effects.

If the logarithm of the forward rate is denoted by $f_t = s_t + i_t - i_t^*$ then the expected excess return can be written

$$E_t R_{t+1} = E_t s_{t+1} - f_t = E_t \Delta s_{t+1} - [f_t - s_t] = \phi_t \tag{57}$$

where $f_t - s_t$ is the forward premium. If the rational expectations innovation is defined as $\varepsilon_{t+1} = R_{t+1} - E_t[R_{t+1}]$, then equation (57) can be written

$$R_{t+1} = \alpha + \beta[f_t - s_t] + e_{t+1} \tag{58}$$

where $e_{t+1} = \phi_t + \varepsilon_{t+1}$. Market efficiency implies that $\alpha = \beta = 0$ and rationality implies that $E_t[\varepsilon_{t+1}] = 0$. If, in addition, investors are risk neutral then $\phi_t = 0$, and so $E_t[e_{t+1}] = 0$. When all these assumptions hold the OLS estimators of α and β will be consistent.[5] Equation (58)

[5] A more familiar way of writing this model is in terms of $\Delta s(t+1)$ instead of $R(t+1)$ when equation (58) becomes

$$\Delta s(t+1) = \alpha + (\beta + 1)[f(t) - s(t)] + e(t+1)$$

- or more commonly a variant in which R_{t+1} is replaced by Δs_{t+1} - is used as the alternative hypothesis in most tests of the efficiency of the foreign exchange market.

Estimates of (58) assuming risk neutrality uniformly reject the parameter restrictions required for market efficiency. In particular, many studies find estimates of β which are negative. Although theory predicts that the dollar will depreciate if the forward premium is positive, the implication of a negative value of β is that it will appreciate. Thus, instead of the interest differential $i - i^*$ compensating for an expected future exchange rate depreciation, this evidence implies that it is accompanied by an exchange rate appreciation. Or, put differently, the greater the interest differential of the foreign over the domestic bond $i^* - i$, the larger will be the excess return to doing so. This is the forward premium puzzle. In general, therefore, the appropriate investment strategy would be to hold the bond with the higher interest rate; the subsequent exchange change will usually reinforce this advantage. In practice, this would be bound to lead to destabilizing FOREX speculation; investing in the bond with the higher domestic currency return is therefore a one-way bet.

The implausibility of the presence of such an arbitrage opportunity suggests that there must be another explanation for the puzzle. One explanation is that the estimate of β is biased downwards due to the presence of the risk premium in the error term of the regression. This is also consistent with the finding of significant serial correlation and heteroskedasticity in the residuals. Assuming that the FOREX market is efficient and investors are rational, Fama (1984) has shown that the bias in β can be expressed as

$$\begin{aligned} bias &= cov[f_t - s_t, \phi_t]/var[f_t - s_t] \\ &= \rho \left[\frac{var[\phi_t]}{var[f_t - s_t]} \right]^{\frac{1}{2}} \end{aligned} \tag{59}$$

where ρ is the correlation between $f_t - s_t$ and ϕ_t, (i.e. between the forward and risk premia). Negative bias implies, therefore, that $\rho < 0$. This can be interpreted as meaning that for US investors, the greater the expected depreciation of domestic currency, the lower is the required risk premium for holding foreign assets. Furthermore, the size of the bias (about -2 in many studies) implies that the variance of the risk premium must be a multiple factor (four for these studies) greater than the variance of the forward premium.

It may be noted that when the test is carried using a regression of s_{t+1} on f_t, the slope cofficient is close to its theoretical value of unity even though a risk premium is still being omitted. The probable explanation for the different outcome is that whereas s_{t+1} and f_t are non-stationary processes, the risk premium is stationary. Hence super-consistent estimates are obtained even when the risk premium is omitted. This test can therefore tell us nothing about whether there should be a risk premium, although it does suggest that $\beta = 0$ in (58).

5.2 SDF FOREX Models

We consider how to obtain an expression for the foreign exchange risk premium using the stochastic discount model. These factors are jointly distributed with the excess return on foreign exchange, the only asset we consider. The SDF model in this case can be expressed as

$$1 = E_t[M_{t+1}(1 + i_t^* + \Delta s_{t+1})] \tag{60}$$

Here, M_{t+1} is the discount factor required to make the present value of the total income $(1 + i_t^*)$ from holding a foreign bond and converted to domestic currency equal to one unit of domestic curency. The only source of uncertainty here is the one-period ahead spot exchange rate as both i_t and i_t^* are known at time t. Taking logarithms of equation (60) and assuming log-normality gives

$$E_t[m_{t+1} + i_t^* + \Delta s_{t+1}] + \frac{1}{2}V_t[m_{t+1} + i_t^* + \Delta s_{t+1}] = 0 \tag{61}$$

Replacing R_{t+1} in equations (60) and (61) by the risk free rate i_t gives

$$E_t[m_{t+1} + i_t] + \frac{1}{2}V_t[m_{t+1}] = 0 \tag{62}$$

Subtracting equation (62) from equation (61) gives

$$E_t[R_{t+1}] + \frac{1}{2}V_t[R_{t+1}] = -Cov_t[m_{t+1}, R_{t+1}] \tag{63}$$

The last term on the left hand-side of equation (63) is the Jensen effect. The term of the right hand-side is the FOREX risk premium for the US investor. Comparing equation (63) with equation (58) implies that ϕ_t is not in fact just the risk premium but is

$$\phi_t = -\frac{1}{2}V_t[R_{t+1}] - Cov_t[m_{t+1}, R_{t+1}] \tag{64}$$

This implies that ϕ_t will have a higher variance than the FOREX risk premium which will be of some assistance in helping to generate the additional variability required in ϕ_t.[6]

Because equation (63) involves conditional expectations, and given equation (57), it can be expressed in other ways, for example, as[7]

$$E_t[R_{t+1}] + \frac{1}{2}V_t[\Delta s_{t+1}] = -Cov_t[m_{t+1}, \Delta s_{t+1}] \tag{65}$$

[6] If logarithms are not taken, and the excess return is defined as $1 + R(t+1) = \frac{(1+i^*(t))S(t+1)}{(1+i(t))S(t)}$ then the arbitrage relation would be $E_t[R(t+1)] = -Cov_t[M(t+1), R(t+1)]$
which does not involve the Jensen effect.

[7] $\Delta s(t+1)$ could be replaced in equation (65) by $s(t+1)$.

This shows explicitly that uncertainty about the future spot exchange rate is a necessary element in the risk premium. The larger the predicted covariance between the rate of appreciation of domestic currency and the discount rate, the smaller the risk premium of domestic investors holding foreign denominated assets. Although these domestic investors only suffer a loss when domestic currency appreciates, the larger the discount rate, the less this loss is.

It has been implicitly assumed that the risk is being borne by domestic investors through their holding of foreign bonds. This would imply that the discount factor is that appropriate for domestic investors. In practice, of course, foreign investors are exposed to the same FOREX risk in reverse.[8] Measuring returns and the discount factor in foreign currency would give

$$E_t[R^*_{t+1}] + \frac{1}{2}V_t[R^*_{t+1}] = -Cov_t[m^*_{t+1}, R^*_{t+1}] \tag{66}$$

where $-m^*$ is the foreign investor's discount rate and is measured in foreign currency, and $R^* = -R$. This implies that for the UK investor the expected excess return is determind by

$$E_t[-R_{t+1}] + \frac{1}{2}V_t[R_{t+1}] = Cov_t[m^*_{t+1}, R_{t+1}] \tag{67}$$

where $Cov_t[m^*_{t+1}, R_{t+1}]$ is the risk premium.

Subtracting equation (67) from (63) gives

$$E_t[R_{t+1}] = -Cov_t[\frac{1}{2}(m_{t+1} + m^*_{t+1}), R_{t+1}] \tag{68}$$

Thus the Jensen effect disappears. The combined risk premium is the difference between the individual investor risk premia and is due to covariation between the average of the discount factors of the domestic and foreign investors and the excess return defined for the domestic investor (or, equivalently, Δs_{t+1}). Adding equations (68) and (63) gives

$$V_t[R_{t+1}] = Cov_t[m^*_{t+1}, \Delta s_{t+1}] - Cov_t[m_{t+1}, \Delta s_{t+1}] \tag{69}$$

This implies that $\Delta s_{t+1} = m^*_{t+1} - m_{t+1} + \eta_{t+1}$ with $Cov_t[\Delta s_{t+1}, \eta_{t+1}] = 0$. Equation (69) reveals that there is a linear relation between $V_t[\Delta s_{t+1}]$, $Cov_t[m^*_{t+1}, \Delta s_{t+1}]$ and $Cov_t[m_{t+1}, \Delta s_{t+1}]$ and only two terms are required, as in equation (68). There is an important proviso to this result. If, as is likely, there is measurement error in the proxy for the discount factor, then it will not hold in practice in the data. This is not a weakness of SDF theory *per se*, but an indication of the likely effects of modelling the discount factor incorrectly.

[8] It is also possible for domestic investors to hold short positions. In this case the source of risk would be the same as that of foreign investors, though the discount factor would still be that of the domestic investor. We ignore this complication in our discussion. It is probable that a relatively small proportion of FOREX transactions are of this type.

In the case of complete markets the two discount factors are identical when measured in the same currency, see Backus, Foresi and Telmer (2001). Hence $m^*_{t+1} = m_{t+1} + \Delta s_{t+1}$. This would imply that equations (65) and (67) are then identical.

5.3 Affine Models of FOREX

The affine models of the term structure discussed above can be applied to the FOREX problem. Backus et al (2001) assess the ability of two affine models to match the properties of the risk premium set out by Fama (1984). The first model they consider is a two factor CIR model where the factors are assumed to be independent and country-specific. As we have seen, this is a model that was rejected by Dai and Singleton (2000) as a good description of the term structure. However, it can be used to make the issues with multi-country affine models clear. Using the notation employed earlier, the structure of the affine models of the SDF for each country and the processes for the unobservable state variables are

$$\begin{aligned}
-m_{t+1} &= z_{1t} + \lambda_1 \sigma_1 \sqrt{z_{1t}} \varepsilon_{1t+1} \\
-m^*_{t+1} &= z_{2t} + \lambda_2 \sigma_2 \sqrt{z_{2t}} \varepsilon_{2t+1} \\
z_{1,t+1} - \mu &= \theta(z_{1t} - \mu) + \sigma_1 \sqrt{z_{1t}} \varepsilon_{1t+1}. \\
z_{2,t+1} - \mu &= \theta(z_{2t} - \mu) + \sigma_2 \sqrt{z_{2t}} \varepsilon_{2t+1}
\end{aligned}$$

It can be shown that the independent state variables, the short interest rate for each country, are proportional to the country-specific state variable

$$r_t = (1 - \frac{1}{2}\lambda_1^2 \sigma_1^2) z_{1t} \tag{70}$$

$$r^*_t = (1 - \frac{1}{2}\lambda_2^2 \sigma_2^2) z_{2t} \tag{71}$$

Backus et al. work in a complete markets environment. This gives the forward premium

$$f_t - s_t = (1 - \frac{1}{2}\lambda^2 \sigma^2)(z_{1,t} - z_{2,t}) \tag{72}$$

for the simplified case where the prices of risk and unconditional variances are symmetric $\lambda_1 = \lambda_2 = \lambda$ and $\sigma_1 = \sigma_2 = \sigma$. The expected rate of depreciation of the spot rate is

$$E_t \Delta s_{t+1} = z_{1,t} - z_{2,t} \tag{73}$$

and the implied risk premium is

$$rp_t = -\frac{1}{2}\lambda^2 \sigma^2 (z_{1,t} - z_{2,t}) \tag{74}$$

In this case the first of Fama's requirements is satisfied, namely, the risk premium and the expected rate of depreciation are negatively correlated. The implied slope coefficient in a regression of the expected rate of depreciation on the forward premium is $1/(1 - \frac{1}{2}\lambda^2\sigma^2)$ which will be positive and greater than one for plausible calibrations. Hence this fails to provide an explanation for the empirical results discussed above. Backus et al. show that this slope coefficient can be negative if a common independent factor is added to the model. However, in this case there is a finite probability that interest rates could be counter-factually negative.

Backus et al. also use a version of the two-factor affine model that is closer to that preferred by Dai and Singleton (2000), namely a two-factor model with interdependent factors. Again, in theory, this model is capable of matching both the negative correlation between the risk premium and the expected rate of depreciation, and the negative regression slope. But this only happens if the home country factor has a larger impact on the interest rate in the foreign country, and vice versa.

Backus et al. also present GMM estimates of the parameters of the unobserved state variable processes based on matching the first two moments and the autocorrelation coefficient of the home currency short rate, the variance and autocorrelation of the forward premium, the variance of the depreciation rate and the slope parameter of the forward premium regression. Using monthly data on the US dollar - sterling exchange rate from July 1974 - November 1994, their estimates imply extreme distributional properties for the unobserved factors and the forward premium for both the CIR independent factors and for the interdependent factor models. In particular, if these factors were interpreted in the term structure context, interdendent factors imply high prices of risk and counterfactually steep term structure slopes. Backus et al. therefore conclude that whilst the more general model with interdependent factors can in theory be consistent with the features of the risk premium that were laid out by Fama, the estimates suggest the model has severe shortcomings. A more direct comparison with observed factor models could be made if other moments were used in the estimation, or the direct estimation approach of Dai and Singleton were followed.

An extended version of the approach followed by Backus et al. is presented in Hollifield and Yaron (2001). They extend the modedl to allow for the risk premium to be decomposed into real and nominal components for each country, and for their interaction. Using GMM estimation, Hollifield and Yaron find that all of the predictable components in FOREX returns can be ascribed to predictable variation in real risk and none to nominal risk, or to the interaction between the two. An interesting conclusion is that the effects of monetary policy on predictable returns are not observable in inflation rates, but only in interest rate differentials.

5.4 The SDF FOREX model with observable factors

5.4.1 CCAPM

An early study of the SDF model of the FOREX market is that of Mark (1985). Based on CCAPM with power utility, assuming complete markets and taking the US as the home country, he uses GMM to estimate the Euler equation for a number of exchange rates. This can be written

$$\beta E_t[\beta \left(\frac{C_{t+1}}{C_t} \right)^{-\sigma} \frac{P_t}{P_{t+1}} \frac{S_{t+1}}{S_t} \frac{(1+i_t^*)}{(1+i_t)}] = 1$$

where P_t is the domestic price level and S_t is the domestic price of foreign exchange. Mark finds estimates of the coefficient of relative risk aversion σ which are both large and not well determined. He also rejects the over-identifying restrictions implied by CCAPM. Also using GMM, Hodrick (1989) subsequently confirmed this result, but for a wider set of currencies and for additional home countries. Although the model was not rejected for one or two configurations, the overall conclusion is negative.

Wickens and Smith (2001) compare the use of two models: CCAPM and CAPM. Both are expressed as observable stochastic discount factor models. They also note the relevance of the market structure: whether it is assumed that investors are domestic residents, foreign residents or both. They consider the sterling dollar exchange rate, where s_t is the logarithm of the number of dollars per pound. For the US domestic investor the no-arbitrage condition for the excess return is

$$E_t R_{t+1} + \frac{1}{2} V_t[R_{t+1}] = \sigma^{us} Cov_t[c_{t+1}^{us}, R_{t+1}] + Cov_t[\Delta p_{t+1}^{us}, R_{t+1}] \tag{75}$$

where $R_{t+1} = \Delta s_{t+1} + i_t^* - i_t$. Thus, the larger the predicted covariation of the depreciation of domestic currency with the rate of growth of domestic consumption and with domestic inflation, the greater the risk premium for domestic investors in foreign bonds. For the UK investor the no-arbitrage condition is

$$E_t R_{t+1} - \frac{1}{2} V_t[R_{t+1}] = -\sigma^{uk} Cov_t[c_{t+1}^{uk}, R_{t+1}] - Cov_t[\Delta p_{t+1}^{uk}, R_{t+1}] \tag{76}$$

When both investors are in the market these two equations can be added to obtain

$$R_{t+1} = \frac{1}{2}\{\sigma^{us} Cov_t[c_{t+1}^{us}, R_{t+1}] + Cov_t[\Delta p_{t+1}^{us}, R_{t+1}] - \sigma^{uk} Cov_t[c_{t+1}^{uk}, R_{t+1}] - Cov_t[\Delta p_{t+1}^{uk}, R_{t+1}]\} \tag{77}$$

Under the assumption of complete markets the US and UK discount factors are identical when expressed in the same currency. It can be shown that in this case the three equations (75), (76) and (77) are identical.

5.4.2 CAPM

The discussion in Section 2 shows that in traditional CAPM the value function is defined in terms of the mean and variance of financial wealth rather than consumption. For the two-period problem this gives

$$M_{t+1} = \sigma_t \frac{W_{t+1}}{W_t} = \sigma_t(1 + R^W_{t+1}) \tag{78}$$

where W_t is nominal financial wealth and R^W_{t+1} is the nominal return on wealth. The discount factors can be obtained from the variables that explain this portfolio return.

The relevance of CAPM to FOREX is the widespread use of mean-variance analysis in hedging FOREX risk. The uncertainty about the pay-off to the possibly partly-hedged portfolio then arises from the future return on the portfolio. In the case of FOREX this is just pure currency risk and anything correlated with this - such as tomorrow's domestic and foreign money supplies and output - could be used to help reduce it. Wickens and Smith use the monetary model of the exchange rate to generate observable factors that determine the exchange rate. In the monetary model, the exchange rate is determined by future expected relative money supplies and output levels, see for example Frenkel (1976) and Obstfeld and Rogoff (1998). This would suggest that exchange risk might be due to forecast covariation between today's exchange rate and tomorrow's domestic and foreign money supplies and output.

The index of industrial production is used as the monthly output measure, and narrow money as the measure of money. For the US investor model, only US industrial production and the US money supply are used, and for the UK investor model, only UK industrial production and the UK money supply. The combined investor model includes all four variablesWickens and Smith then use the monetary model of the exchange rate to explain this, and hence to provide the macroeconomic factors.

5.4.3 Empirical evidence

Wickens and Smith estimate all of the CCAPM and CAPM models using constant correlation MGM for the vector of stationary variables $x_{t+1} = \{R_{t+1}, z'_{t+1}\}'$ where the z_{t+1} is the vector of observable factors appropriate for each specification. As explained in section 3.2, the equation for the excess return must be constrained to avoid arbitrage possibilities. For comparison purposes, a general alternative model is also estimated. This can be written

$$R_{t+1} = \gamma_1 R_t + \gamma_2[f_t - s_t] + V_t[R_{t+1}] + \phi^{us\prime} C^{us}_{t+1} + \phi^{uk\prime} C^{uk}_{t+1} + \varepsilon_{1t+1}$$

where C^{us}_{t+1} and C^{uk}_{t+1} denote the macroeconomic factors associated with the US and UK investor models, respectively. ε_{1t+1} is used to refer to the error term in the excess return equation no matter which model is chosen. In practice the error terms in each model would be different.

The data used are monthly from 1975.1 - 1997.12. As consumption data are not available monthly, deflated retail sales data are used. The output series are the indices of industrial production. The money supply is the money base for the US and M0 for the UK. At a monthly frequency there are significant ARCH effects for all of the variables. Estimates are presented for the dollar-based investor holding sterling assets, the sterling-based investor and for both types of investor. For CCAPM, the conditional covariance terms in the two single investor models are significant at the 10% level. In the two investor model only the conditional covariance with US consumption is significant. In the general model none of the conditional covariances is significant, but the covariance with UK consumption is significant. The implied estimate of the coefficient of relative risk aversion for the US and UK investors is -289 and -283. For the two-investor model it is -410 for US investors. All of these estimates therefore have the wrong sign and are very large. Moreover, in the general model none of the conditional covariances is significant and, as would be expected in view of this, the lagged excess return and forward premium retain their significance. In other words, the forward premium puzzle is not resolved. The broad conclusion that emerges from these results about CCAPM is similar to those of Mark and Wu (1998), Lewis (1995) and Engel (1996). Estimates based on power utility are not consistent with the theory.

The theoretical predictions for the monetary model are that the coefficient on conditional covariances should be positive for US money and negative for US output, and these signs should be reversed for the UK variables. The estimates for the UK investor have the correct sign and are significant. For the two investor model all the signs are correct, but the covariances with UK money are not significant. In the general model the output covariances are the most significant. These results therefore show considerable support for the monetary model and hence for the traditional models of currency risk. However, the continued significance of the lagged excess return and forward premium in the general model indicates, however, that the forward premium puzzle is not resolved even if output, and in some cases money, seem to be significant sources of FOREX risk.

6 SDF models of equity returns

A great deal of empirical evidence has been accumulated on equity returns based on the capital asset pricing model (for surveys of this see, for example, Ferson (1995) who refers to them as beta models, and Campbell, Lo and MacKinlay (1997)) and on the use of VAR models to test the present value model of equity (see, for example, the surveys of Campbell, Lo and MacKinlay (1997), Cuthbertson (1996) and Bollerslev and Hodrick (1995)). Although we have shown that CAPM can be interpreted as an SDF model, and the present value model of equity is an SDF model with a constant discount factor, for space reasons, we will not include this material in this survey. The beta versions of the CAPM as implemented in many empirical studies of portfolio returns assume that the risk premium is constant; this rules these models out *a priori*.[9] Instead, we will concentrate on the more explicit use of the SDF model to price equity.

The theory is straightforward. We can use the SDF models, equations (6) or (10) using $1 + r_{t+1} = \frac{P_{t+1}+D_{t+1}}{P_t} . \frac{1}{1+\pi_{t+1}}$ as our measure of real equity returns.[10] The remaining issues are the choice of discount factor model and the method of estimation. Nearly all of the research on SDF models for pricing equity are based on CCAPM or its variants. For recent surveys see Ferson (1995) and Campbell (1999). Almost no work has been carried out using other methods. A notable exception is the explicit use of the affine factor model by Bekaert and Grenadier (1999).

6.1 CCAPM

GMM estimation The widely cited paper by Hansen and Singleton (1982) was the first to estimate the Euler equation directly. Their SDF model was CCAPM based on power utility and their method estimation was GMM. Using equations (13) and (18) their Euler equation can be written

$$E_t[\beta \left(\frac{C_{t+1}}{C_t} \right)^{-\sigma} (1 + r_{t+1}) - 1] = 0 \tag{79}$$

This is a particular case of equation (40). The instruments used were lags of $\frac{C_{t+1}}{C_t}(1 + r_{t+1})$. The estimates were found to give some, but not strong, support for the theory.

The consensus of a large number of studies, using a wide variety of usually less formal empirical methods, including calibration, is that CCAPM equity returns are not well explained by CCAPM based on power utility. As noted above, the general conclusion is that equity returns are too large to be explained by the model and implausibly large values of the coefficient of relative risk aversion, σ_t, would be required to remedy this. This is the equity premium puzzle. Attention therefore turned to reformulating the utility function to produce a larger risk premium. We now examine these attempts in more detail.

---

[9] An interesting debate has recently begun over whether more precise estimates can be obtained from estimation of factor models such as the CAPM by traditional beta methods or by treating the model as an SDF. Kan and Zhou (1999) find in favour of beta methods in a very restrictive case. Whilst Cochrane (2000) and Jagannathan and Wang (2002) show how restrictive this case is, it may be relevant in particular for linear or linearised models such as those reviewed by Campbell (1999) and thus be of more general importance.

[10] It may be noted that when the discount factor M_{t+1} is the constant M, equation (1) becomes $P_t = E_t[M_{t+1}X_{t+1}] = ME_t[P_{t+1} + D_{t+1}] = \sum_{s=1}^{\infty} M^s E_t D_{t+s}$ if $\lim_{n\to\infty} M^s P_{t+s} = 0$, implying that the price of equity is the present value of discounted expected future dividends. And since $E_t(M_{t+1}) = M = \frac{1}{1+r_t^f}$ equation (6) becomes $E_t r_{t+1} = r_t^f = \frac{1}{M} - 1$.

6.1.1 Habit persistence

One of the most successful alternatives is the habit persistence model in which instantaneous utility is assumed to depend on current consumption and past consumption habits (or the sustainable level of consumption) X_t to give $U(C_t, X_t)$. We consider two examples. Abel (1990) assumes that the utility function can be written

$$U_t = \frac{\left(\frac{C_t}{X_t}\right)^{1-\sigma} - 1}{1-\sigma}$$

where X_t is a function of past consumption, for example, $X_t = C_{t-1}^{\delta}$. The stochastic discount factor becomes $M_{t+1} = \beta \left(\frac{\frac{C_{t+1}}{X_{t+1}}}{\frac{C_t}{X_t}} \right)^{-\sigma}$ and the no-arbitrage condition is

$$E_t(r_{t+1} - r_t^f) = \beta(1 + r_t^f)\sigma_t Cov_t\left(\Delta \ln C_{t+1} - \Delta \ln X_{t+1}, r_{t+1}\right) \tag{80}$$

It appears from this that there is additional flexibility in the risk premium arising from the conditional covariance between $\Delta \ln X_{t+1}$ and r_{t+1}. But if X_t is a function of past consumption then the conditional covariance is zero, and equation (80) simply reverts to (16), an unhelpful outcome.

As noted earlier, Constantinides (1990) has proposed a utility function of the form

$$U_t = \frac{(C_t - \lambda X_t)^{1-\sigma} - 1}{1-\sigma}$$

This has been used by Campbell and Cochrane (1999) with the restriction that $\lambda = 1$. They introduce the concept of surplus consumption, defined as $S_t = \frac{C_t - X_t}{C_t}$. Maintaining the framework above, the no-arbitrage condition can be written

$$E_t(r_{t+1} - r_t^f) = \beta(1 + r_t^f)\sigma_t Cov_t\left(\Delta \ln C_{t+1}, r_{t+1}\right) + \beta(1 + r_t^f)\sigma_t Cov_t\left(\ln S_{t+1}, r_{t+1}\right)$$

This has the added flexibility provided by the conditional covariance of $\ln S_{t+1}$ and r_{t+1}. Campbell and Cochrane assume that $\ln S_t$ is generated by an AR(1) process with a disturbance term whose variance depends on $\ln S_t$. They do not estimate the resulting model but, like Constantinides, use calibration methods. By calibrating the variance of the error term of the AR process suitably, it is possible to force the conditional covariance between $\ln S_{t+1}$ and r_{t+1} to be of the necessary size. By this means success is virtually guaranteed. A final judgement about habit persistence models must await the use of empirical methods based on classical statistical inference that are less biased in favour of the model.

6.1.2 Time non-separable preferences

Empirical work on time non-separable utility has concentrated on the Epstein-Zin utility function. This has been implemented by, for example, Epstein and Zin (1991). They estimate the Euler equation

$$E_t\left\{\left[\beta\left(\frac{C_{t+1}}{C_t}\right)^{-\frac{1}{\gamma}}\right]^{\frac{1-\sigma}{1-\frac{1}{\gamma}}}(R^m_{t+1})^{1-\frac{1-\sigma}{1-\frac{1}{\gamma}}}R_{t+1}-1\right\}=0 \tag{81}$$

by GMM using as instruments current and lagged consumption growth and R^m_t which is measured by the value-weighted index of NYSE shares. R_t is an individual share price. Using the parameter definitions above, a typical example of their estimates based on the Epstein-Zin results is $\beta = 0.997$, $\sigma = -0.94$, $\gamma = -0.97$. Since the estimates of σ and γ are negative, and not positive, this is not encouraging.

Campbell and Viceira (1999) also use the Epstein-Zin model in an examination of portfolio choice, but they calibrate the model and do not estimate it. They assume log-normality and use the no-arbitrage condition, equation (19).

6.2 Affine models of equity

Bekaert and Grenadier (1999) have proposed a class of affine factor models for pricing equity that involves observable factors. Starting with CCAPM, they argue that in equilibrium the consumption process C_t must equal the exogenous aggregate real dividend process D_t and hence, from equation (18), the log stochastic discount factor can be written

$$m_{t+1} = \ln\beta - \sigma\Delta d_{t+1}$$

where $\ln D = d$. They then assume that the dividend process is driven by real productivity shocks x_t so that

$$\Delta d_{t+1} = \frac{\sigma}{2}\sigma_d^2 + \frac{\ln\beta}{\sigma} + \frac{1}{\sigma}x_t + \sigma_d\xi_{t+1}$$

and they assume that x_t has a CIR process. We write this as

$$x_{t+1} - \mu = \theta(x_t - \mu) + \sigma_x\sqrt{x_t}\varepsilon_{t+1}$$

ξ_t and ε_t are asumed to be $iid(0,1)$ processes. They do not derive the no-arbitrage equation for pricing equity that would be implied by this, but it is

$$E_t(r_{t+1} - r_t^f) + \frac{1}{2}V_t(r_{t+1}) = \sigma Cov_t(\Delta d_{t+1}, r_{t+1}) = 0.$$

Hence the equity risk premium is zero. If the dividend process had been written instead as

$$\Delta d_{t+1} = \frac{\sigma}{2}\sigma_d^2 + \frac{\ln\beta}{\sigma} + \frac{1}{\sigma}x_{t+1} + \sigma_d\xi_{t+1}$$

then the no-arbitrage condition would become

$$E_t(r_{t+1} - r_t^f) + \frac{1}{2}V_t(r_{t+1}) = \sigma_x\sqrt{x_t}Cov_t(\varepsilon_{t+1}, r_{t+1}).$$

This would be a more useful formulation.

Bekaert and Grenadier model equity and bond togther. They therefore introduce another factor, the inflation rate, which is also assumed to be generated by a CIR process. The whole model is estimated by GMM. Given the complexities involved, and the space that would be required to explain these, we are unable to go into further detail.

6.3 Observable factors using the MGM model

As for the term structure, we complete our discussion of SDF models of equity by proposing a new approach based on observable factors and the use of the MGM model. This is based on a discussion paper currently in preparation, see Smith, Sorensen and Wickens (2002). A problem with the affine factor approach of Bekaert and Grenadier is the assumption that the macroeconomic variables are generated by CIR processes may be incorrect. In fact, it is unlikely that they are. Ideally, we would like to model the macroeconomic factors using appropriate macroeconomic theory and econometric models, and not necessarily models that satisfy finance conventions, however convenient this may be. On the other hand, to be useful for asset pricing, the macroeconomic variables must exhibit time-varying volatility, something very few macroeconometric models allow.

Our starting point is the assumption that the log discount factor is generated by

$$m_{t+1} = a + \sum_i b_i z_{i,t+1}.$$

and hence the no-arbitrage condition is

$$E_t(r_{t+1} - r_t^f) + \frac{1}{2}V_t(r_{t+1}) = -\sum_i b_i Cov_t(z_{i,t+1}, r_{t+1}) \qquad (82)$$

Instead of assuming that the factors are generated by CIR processes we suppose that the joint distribution of $\mathbf{x}_{t+1} = (r_{t+1} - r_t^f,\, z_{1,t+1},\, z_{2,t+1},...)'$ is the MGM model defined by equations (30), (31) and (32). Next we must choose the factors. We could do so on the basis of any (or all) of the equity pricing models discuss in this section. The alternative hypothesis could be formulated from an inclusive choice, and the null hypothesis from any specific model. The previous discussion suggests that possible variables are the growth in consumption, output and dividends. we could also include, for example, bond yields, inflation, the exchange rate, money growth and earnings growth. It would, of course, be more satisfactory if our choice had sound theoretical underpinnings. Essentially, we are aiming to exploit any volatility contagion between equity returns and other variables. For many stock markets, contagion from the US stock market is likely to be important. There may also be contagion from the domestic bond market and the foreign exchange market. There is, therefore no shortage of potential factors. The main practical problem is that the more factors we include, the larger the number of parameters to be estimated, and the more difficult it is to achieve numerical convergence. As shown above, the parameters in the unrestricted BEKK model grow at the rate n^4, and for the other models they grow at the rate n^2.

Among the models considered by Smith, Sorensen and Wickens is CCAPM with power utility. For real returns, this generates just one stochastic discount factor, the rate of growth of consumption. The model does not perform well. Although the estimate of the coefficient on the conditional covariance in mean (ie the coefficient of relative risk aversion) is positive and significant, it is implausibly large, like previous studies based on CCAPM.

7 Conclusion

Empirical work on financial markets has recently become more conscious of using a well-specified no-arbitrage condition to price assets, rather than the more more traditional expectations hypothesis. The stochastic discount factor model provides a general way of deriving this condition that encompasses most standard asset pricing models. One of the main problems that remain is how to implement the SDF model. In this paper we have provided a selective survey of the methods that have been used in the literature, and we have shown how they been implemented in the three principal financial markets: the bond market, the foreign exchange market and the stock market.

We have outlined the main features of the SDF approach and shown why a large number of standard asset pricing methods are implicitly SDF models. A key issue in selecting an SDF model is the choice of pricing kernel, or discount factor. Both latent and observable factors have been used. In general, the SDF model is a highly nonlinear. This greatly complicates the estimation of the model. One response is to use affine factor models. Another is to find better non-linear estimators. We discuss both approaches. We have considered in some detail the implementation of latent variable and observable affine factors models, especially in the bond market. And we have described several non-linear econometric methods and their application.

In addition to surveying the literature, we have proposed a new way to formulate and estimate the SDF model using observable factors. This provides a general formulation convenient for comparing and testing different types of SDF model. We show how the model can be estimated by multivariate GARCH-in-mean and we include empirical examples of this approach.

8 References

Abel, A.B. (1990) Asset prices under habit formation and catching up with the Joneses, *American Economic Review*, 80, 38-42.

Ang, A. and M. Piazzesi (2000) A no-arbitrage vector autoregression of term strcuture dynamics with macroeconomic and latent variables, Columbia University Working Paper.

Backus,D., S.Foresi and C.Telmer (1996) Affine models of currency pricing, NBER Working Paper No. 5623.

Backus,D., S.Foresi and C.Telmer (1998) Disrete-time models of bond pricing, NBER Working Paper No. 6736.

Backus,D., S.Foresi and C.Telmer (2001) Affine term structure models and the forward premium anomaly, *Journal of Finance*, 56, 1, 279-304.

Baillie, R.T. and T.Bollerslev (1990), A multivariate generalized ARCH approach to modeling risk premia in forward foreign exchange rate markets, *Journal of International Money and Finance*, 9, 309-324.

Balduzzi, P., S.R. Das, S. Foresi and R.K. Sundaram (1996) A simple approach to three factor affine term strcuture models, *Journal of Fixed Income*, 6, 43-53.

Bekaert, G. and S.R. Grenadier (1999) Stock and bond pricing in an affine economy, NBER Working Paper 7346.

Bekaert,G. and R.J. Hodrick (2001) Expectations hypothesis tests, *Journal of Finance*, 56, 1357-1393.

Black, F. (1972) Capital market equilibrium with restricted borrowing, *Journal of Business*, 45, 444-454.

Bollerslev,T.(2001) Financial econometrics: past developments and future challenges, *Journal of Econometrics*, 100, 41-51.

Bollerslev,T. Engle,R.F. and D.B Nelson(1994) ARCH models, in Engle,R.F. and D.McFadden (eds), *Handbook of Econometrics*, vol IV, North-Holland, Amsterdam.

Bollerslev,T., R.Y.Chou and K.F.Kroner(1992), "ARCH modeling in finance: a selective review of the theory and selective evidence", *Journal of Econometrics*, 52, 5-59.

Bollerslev,T. and R.J.Hodrick (1995), Financial market effciency tests, in *Handbook of Applied Econometrics*, eds M.H.Pesaran and M.R.Wickens, Blackwell.

Campbell, J.Y. (1987) Stock returns and the term structure, *Journal of Financial Economics*, 18, 373-399.

Campbell, J.Y. (1993) Intertemporal asset pricing without consumption data, *American Economic Review,* 83, 487-512.

Campbell, J.Y. (1996) Understanding risk and return, *Journal of Political Economy*, 104, 298-345.

Campbell, J.Y. (1999), Asset pricing, consumption, and the business cycle, in *Handbook of Macroeconomics*, eds J.B.Taylor and M.Woodford, vol 1, Elsevier.

Campbell, J.Y. and J.H.Cochrane (1999) By force of habit: a consumption-based explanation of aggregate stock market behaviour, *Journal of Political Economy*, 107, 205-251.

Campbell, J.Y., A. Lo and A. MacKinlay (1997) *The Econometrics of Financial Markets*, Princeton University Press, Princeton, New Jersey.

Campbell, J.Y. and R. Shiller (1987) Cointegration and tests of present value models, *Journal of Political Economy*, 95, 1062-1087.

Campbell, J.Y. and R. Shiller (1988) Stock prices, earnings and expected dividends, *Journal of Finance*, 43, 661-676.

Campbell, J.Y. and L.M. Viceira (1999) Consumption and portfolio decisions when expected returns are time varying, *Quarterly Journal of Economics*, 114, 433-495.

Cochrane, J.H.(2000), "A rehabilitation of stochastic discount factor methodology", mimeo, University of Chicago.

Chen, N., R. Roll and S. Ross (1986) Economic forces and the stock market, *Journal of Business*, 59, 383-404.

Chen, R.R. and L. Scott (1993) Maximum likelihood estimation for a multi-factor equilibrium model of the term structure of interest rates, *Journal of Fixed Income*, 3, 14-31.

Cochrane, J. (2000) A resurrection of the stochastic discount factor / GMM methodology, mimeo, University of Chicago.

Cochrane, J. (2001) *Asset Pricing*, Princeton University Press, Princeton, New Jersey.

Connor, G. and R.A.Korajczyk (1995) The arbitrage pricing theory and multifactor models of asset returns, in *Finance, Handbooks in Operational Research and Management Science*, vol 9, eds R.A. Jarrow, V. Maksimovic and W.T.Ziemba, North-Holland, Amsterdam.

Constantinides, G.M. (1990) Habit formation: a resolution of the equity premium puzzle, *Journal of Political Economy*, 98, 519-543.

Cuthbertson, K.(1996), *Quantitative Financial Economics*, Wiley, London.

Cox, J.C., J. Ingersoll and S. Ross (1985) A theory of the term structure of interest rates, *Econometrica*, 53, 385-408.

Dai, Q. and K. Singleton (2000) Specification analysis of affine term structure models, *Journal of Finance*, 55,1943-1978.

Ding, Z. and R.F. Engle (1994) Large scale conditional covariaqnce matrix modelling, estimation and testing, mimeo, University of California, San Diego.

Duffie,D. and R.Kan (1996) A yield factor model of interest rates, *Mathematical Finance* 6,379-406.

Duffie,D and K.Singleton (1993) Simulated method of moments estimation of markov models of asset prices, *Econometrica*, 61, 929-952.

Duffie,D and K.Singleton (1997) An econometric model of the term structure of interest rate swap yields, *Journal of Finance*, 52, 1287-1321.

Engel, C. (1996) The forward discount anomaly and the risk premium: a survey of recent evidence, *Journal of Empirical Finance*, 3, 123-192.

Engle, R.F. and K.K.Kroner (1995) Multivariate simultaneous generalised GARCH, *Econometric Theory*, 11, 122-150.

Epstein, L.G. and S.E. Zin (1989) Substitution, risk aversion and the temporal behaviour of consumption and asset returns: a theoretical framework, *Econometrica*, 57, 937-968.

Epstein, L.G. and S.E. Zin (1990) First order risk aversion and the equity premium puzzle, *Journal of Monetary Economics*, 26, 387-407.

Epstein, L.G. and S.E. Zin (1991) Substitution, risk aversion and the temporal behaviour of consumption and asset returns: an empirical investigation, *Journal of Political Economy*, 99, 263-286.

Evans, M.D.D. (1998) Real rates, expected inflation and inflation risk premia, *Journal of Finance*, 53, 187-218.

Fama, E.(1984), Forward and spot exchange rates, *Journal of Monetary Economics*, 14, 319-338.

Ferson, W.E. (1995) Theory and empirical testing of asset pricing models, in *Finance, Handbooks in Operational Research and Management Science*, vol 9, eds R.A. Jarrow, V. Maksimovic and W.T.Ziemba, North-Holland, Amsterdam.

Ferson, W.E. and S.R. Foerster (1994) Small sample properties of the generalized method of moments in tests of conditional asset pricing, *Journal of Financial Economics*, 36, 29-36.

Flavin, T.J. and M.R.Wickens (1998) A risk management approach to optimal asset allocation, mimeo, University of York.

Frenkel, J.A.(1976) A monetary approach to the exchange rate: doctrinal aspects and empirical evidence, *Scandanavian Journal of Economics*, 78, 169-191.

Gallant, R. and J. Long (1997) Estimating stochastic differential equations efficiently using minimum chi-squared methods, *Biometrika*, 84, 125-141.

Gallant, R. and G. Tauchen (1996) Which moments to match?, *Econometric Theory*, 12, 657-681.

Gallant, R. and G.Tauchen (2001) Simulated score methods and indirect inference for continuous-time models, mimeo, University of North Carolina.

Gong, F.F. and E.M. Remolona, (1997) Two factors along the yield curve, *Papers in Money, Macroeconomics and Finance: The Manchester School Supplement*, 65, 1-31.

Gourieroux, C., A. Monfort and E. Renault (1993) Indirect inference, *Journal of Applied Econometrics*, 8, S85-S199.

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

Hall, A. (1992) Some aspects of generalized mthod of moments estimation, in G. Maddala, C. Rao and H. Vinod (eds) *Handbook of Statistics, Vol 11: Econometrics*, North-Holland, Amsterdam.

Hamilton, J. (1994) *Time Series Analysis,* Princeton University Press, Princeton, New Jersey.

Hansen, L.P. (1982) Large sample properties of the generalised method of moments estimators, *Econometrica*, 55, 587-613.

Hansen,L. and R.Jagannathan(1991), "Implications of security market data for models of dynamic economies", *Journal of Political Economy*, 99, 225-262.

Hansen, L.P. and K. Singleton (1982) Generalized instrumental variables estimation of non-linear rational expectations models, *Econometrica*, 50, 1269-1286.

Hollifield, B. and A.Yaron (2001) The foreign exchange risk premium: real and nominal factors, mimeo, Carnegie Mellon University.

Hodrick, R. (1987) *The Emprical Evidence on the Efficiency of the Foward and Futures Foreign Exchange Markets*, Harwood, Chur.

Hodrick, R. (1989) Risk uncertainty and exchange rates, *Journal of Monetary Economics*, 23, 433-459.

Ingram, B.F. and B.S. Lee (1991) Simulation estimation of time series models, *Journal of Econometrics*, 47, 197-250.

Jagannathan, R. and Z. Wang (2002) Empirical evaluation of asset pricing models: a comparison of the SDF and Beta methods, *Journal of Finance*, Oct.

Jegadeesh, N. and G. Pennacchi (1996) The behavior of interest rates implied by the term structure of Eurodollar futures, *Journal of Money, Credit and Banking*, 28, 3, 426-446.

Kaminisky, G. and R. Peruga (1990) Can a time varying risk premium explain excess returns in the forward market for foreign exchange?, *Journal of International Economics,* 28, 47-70.

Kan, R. and G. Zhou (1999) A critique of the stochastic discount methodology, *Journal of Finance*, 54, 1221-1248.

Kreps, D and E. Porteus (1978) Temporal resolution of uncertainty and dynamic choice theory, *Econometrica*, 46, 185-200.

Lewis, K.K. (1995) Puzzles in international financial markets, in G. Grossmand and K. Rogoff (eds), *Handbook of International Economics*, Vol 3. Ch 37, North Holland, Amsterdam.

Lintner, J. (1965) The valuation of risk assets and teh selection of risky investments in stock portfolios and capital budgets, *Review of Econmomics and Statistics*, 47, 13-37.

Lucas, R.E (1978) Asset prices in an exchange economy, *Econometrica*, 46, 1429-1446.

Mark, N. (1985) On time-varying risk premia in the foreign exchange market: an econometric analysis, *Journal of Monetary Economics*, 16, 3-18.

Mark, N. and Y. Wu (1998) Rethinking deviations from uncovered interest parity: the role of covariance risk and noise, *Economic Journal,* 108, 1686-1706.

Marsh, T.A. (1995) Term structure of interest rates and the pricing of fixed income claims and bonds, in *Finance, Handbooks in Operational Research and Management Science*, vol 9, eds R.A. Jarrow, V. Maksimovic and W.T.Ziemba, North-Holland, Amsterdam

Mehra, R. and E. Prescott (1985) The equity premium puzzle, *Journal of Monetary Economics*, 15, 145-161.

Newey, W. and K. West (1987) A simple, positive definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica*, 55, 703-708.

Obstfeld,M. and K.Rogoff (1998) Risk and exchange rates, NBER Working Paper No. 6694

Pagan,A.R.(1996) The econometrics of financial markets, *Journal of Empirical Finance*, 3, 15-102.

Remolona,E.M., M.R.Wickens and F.F.Gong (1998) What was the market's view of UK monetary policy? Estimating inflation risk and expected inflation with indexed bonds, Federal Reserve Bank of New York Discussion Paper.

Ross, S.A. (1976) The arbitrage theory of capital asset pricing, *Journal of Economic Theory*, 13, 341-360.

Ross, S.A. (1977) Risk, return and arbitrage, in I. Friend and J. Bicksler (eds) *Risk and Return in Finance I,* Ballinger, Cambridge, MA.

Rubinstein, M (1976) The valuation of uncertain income streams and the pricing of options, *Bell Journal of Economics and Managements Science*, 7, 407-425.

Sharpe, W. (1964) Capital asset prices: a theory of market equilibrium under conditions of risk, *Journal of Finance*,19, 425-422.

Shiller, R. (1990) The term structure of interest rates, in B. Friedmand and F. Hahn (eds) *Handbook of Monetary Economics*, North Holland, Ansterdam.

Singleton, K.(1990) Specification and Estimation of inter-temporal Asset Pricing Models, in B. Friedman and F. Hahn (eds), *Handbook of Monetary Economics*, North Holland Vol. 1, Ch 12.

Smith, A.A. (1993) Estimating non-linear time series models using simulated vector autoregressions, *Journal of Applied Econometrics*, 8, S63-S84.

Smith, P.N. (1993) Modelling risk premia in international asset markets, *European Economic Review*, 37, 159-176.

Smith, P.N., S.Sorenson and M.R. Wickens (2002) Modelling the equity risk premium using the SDF model with observable macroeconomic factors.

Staiger, D. and J. Stock (1997) Instrumental variables regression with weak instruments, *Econometrica,* 65, 557-586.

Tzavalis, E. and M.R. Wickens (1997) Explaining the failures of term spread models of the rational expectations hypothesis of the term structure, *Journal of Money, Credit and Banking*, 29, 364-380.

van der Sluis (1998) Computationally attractive stability tests for the efficient method of moments, *Econometrics Journal*, 1, C203-C227.

Vasicek, O. (1977) An equilibrium characterization of the term structure, *Journal of Financial Economics*, 5, 177-188.

Wickens, M.R. and P.N. Smith (2001) Macroeconomic sources of FOREX risk, mimeo University of York; http://www-users.york.ac.uk/~pns2/forex.pdf