Macroeconomic sources of FOREX risk

M.R.Wickens

and

P.N.Smith

University of York

Paper presented to the Metu-Erc conference
September 2002

Contents

- It aims to provide a general theory of asset pricing

  - affine factor pricing models are a special case

  - suitable for taking account of macroeconomic sources of risk

  - can be used for asset allocation.

  Smith and Wickens "Asset pricing with observable stochastic discount factors", Journal of Economic Surveys, 2002.

- Stochastic discount factor theory is used to provide the theoretical framework. This is capable of embracing most of the approaches in the literature, including general equilibrium theory.

- Three SDF models are used

  1. Benchmark model - based on joint distribution of traditional variables
  2. General equilibrium model - consumption based CAPM
  3. Partial equilibrium model based on pure currency risk

- Market structure needs to be added to this

  1. Complete and incomplete markets
  2. US investor model
  3. UK investor model
  4. Combined US & UK investor model.

- Modelling the exchange rate is key to much of monetary policy (eg the Bank of England's Monetary Policy Committee), and to testing FOREX market efficiency. The forward premium puzzle lies at the heart of the difficulty of doing this. The theoretical results of this paper are used to re-examine the distribution of exchange rate movements and to try to resolve this puzzle.

- Data are monthly for the sterling-dollar exchange rate 1975-1997.

Main findings

1. Many of the models used in the empirical literature of asset pricing have a fundamental flaw: they admit unlimited arbitrage opportunities. High profile suites of computer programs just produced and sold worldwide suffer the same problem, and hence should not be used.

2. The evidence is more consistent with the FOREX risk premium arising from traditional partial equilibrium models of currency risk that form the basis of hedging than with consumption-CAPM, a general equilibrium theory.

3. US and UK output appear to be important sources of FOREX risk.

Tests of FOREX market efficiency

- A stylised fact of the foreign exchange (FOREX) market is that it is not efficient

- Is this rejection due to

-the FOREX risk premium (being omitted or modelled inadequately)

- Or to other causes

- a peso effect
- non-rational expectations?

The Forward premium puzzle

Theory

Ex-post excess return to holding foreign bond

$$\begin{aligned} R(t+1) &= i^*(t) + \Delta s(t+1) - i(t) \\ &= s(t+1) - [s(t) + i(t) - i^*(t)] \\ &= s(t+1) - f(t) \\ &= \Delta s(t+1) - [f(t) - s(t)] \end{aligned}$$

$R(t+1)$ = the excess return to domestic investors from investing at time t in the foreign bond
$i(t)$ and $i^*(t)$ = the domestic and foreign one-period nominal interest rates
$s(t)$ = the logarithm of the domestic price of foreign exchange
$f(t)$ = the logarithm of the forward rate
$f(t) - s(t) = i(t) - i^*(t)$ = forward premium

Risk-neutral investors

$$E_t[R(t+1)] = 0$$

Risk-averse investors

$$E_t[R(t+1)] = \phi(t)$$

$\phi(t)$ = a risk premium.
$\phi(t) \geq 0$ if only domestic investors are exposed to exchange risk
(i.e. foreign investors only hold the foreign bond)
$\phi(t) \gtreqless 0$ in general, depending on the portfolio composition

Traditional tests of FOREX market efficiency

- Null hypothesis

$$E_t s(t+1) = f(t)$$

- Assumes

(i) risk-neutrality
(ii) rational expectations

$$s(t+1) - E_t[s(t+1)] = \varepsilon(t+1)$$

$$E_t[\varepsilon(t+1)] = 0$$

- The null hypothesis can be written

$$s(t+1) = f(t) + \varepsilon(t+1)$$

- Alternative hypothesis

$$s(t+1) = \alpha + \beta f(t) + e(t+1)$$

$H_0 : \alpha = 0$, $\beta = 1$, $e(t+1)$ is serially uncorrelated

- Null hypothesis also implies

$$\Delta s(t+1) = f(t) - s(t) + \varepsilon(t+1)$$

- The alternative hypothesis is now

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

Evidence

- Levels data

$s(t) \approx f(t)$
$s(t+1) \approx f(t) + e(t+1)$
$\Delta s(t+1) \approx e(t+1) \sim$ random walk
$s(t),\ f(t) \sim I(1)$

- Transformed data

$\Delta s(t+1) = -3[f(t) - s(t)] + e(t+1)$
$\Delta s(t),\ f(t) - s(t) \sim I(0)$

Implications

- $\beta = 1$ implies the dollar will *depreciate* if $i > i^*$
- $\beta < 0$ implies dollar *appreciates* if $i > i^*$.
- Hence instead of $i - i^*$ compensating for an expected exchange depreciation, it is accompanied by an appreciation
- The greater the interest differential to holding the foreign bond $i^* - i$, the greater also is the excess return to doing so
- 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 this suggests that there must be another explanation.

Explanation

- An omitted stationary risk premium $\phi(t)$ has caused

- no bias in the levels model as the estimates are super-consistent
- large negative bias in the transformed model

- Bias in β can be expressed as

$$\begin{aligned} bias &= cov[f(t) - s(t), \phi(t)]/var[f(t) - s(t)] \\ &= \rho \left[\frac{var[\phi(t)]}{var[f(t) - s(t)]} \right]^{\frac{1}{2}} \end{aligned}$$

- ρ is the correlation between $f(t) - s(t)$ and $\phi(t)$, (i.e. between the forward and risk premia).
- $bias < 0$ implies $\rho < 0$
- i.e. for US investors, the greater the expected depreciation of domestic currency, the lower is the required risk premium for holding foreign assets
- In effect the 45^0-line predicted by theory when $\Delta s(t+1)$ is used is shifting up or down due to changes in the risk premium.
- The greater the expected depreciation, the smaller the shift
- Hence get a negative (or vertical) scatter diagram instead of a positive one.

- The puzzle is deepened by the fact that general equilibrium models of the risk premium typically do not produce a risk premium that is capable of generating the sorts of bias observed in practice.
- Since the estimate of β is typically negative and $\rho < 1$ the variance of the risk premium would need to be considerably greater than the variance of the forward premium.
- Since the maximum value of $\rho^2 = 1$

$$
\begin{aligned}
var[\phi(t)] &= bias^2.var[f(t) - s(t)]/\rho^2 \\
&\geq bias^2.var[f(t) - s(t)]
\end{aligned}
$$

- equivalent of the Hansen-Jagannathan bound

Contribution of the paper

- Derives a general theory of asset pricing capable of pricing macroeconomic and other sources of risk
- Standard theories such as consumption CAPM are a special case
- Provides the basis for dynamic asset allocation, where the portfolio is tilted (hedged) to avoid specific risks
- Shows that conventional time-series of asset returns are misspecified as they are not arbitrage-free
- For example, univariate modeling of returns will in general be incorrect. Must use multi-variate models
- Standard ARCH and GARCH models are also inappropriate, including the new suite of GAUSS programs FANPAC.

Is the FOREX risk premium the problem?

- Lewis (1995) "no risk premium model with believable measures of risk aversion has yet been able to generate the variability in predictable excess returns that are observed in the data."
- Engel (1996) identified four general directions in which the literature might go forward. One was to extend the analysis of the risk premium.
- This paper examines whether the SDF model is able provide a measure of the foreign exchange risk premium that is consistent with FOREX market efficiency.
- We use an observable, not latent, factor SDF model.
- We assume that the SDF can be proxied by observable macroeconomic variables that are jointly distributed with the excess return on foreign exchange.
- The tests surveyed by Engel and Lewis are based on a special case of the SDF model with observable factors, the inter-temporal consumption-based capital asset pricing model.
- Our more general framework enables us to examine a broader range of macroeconomic variables in a theoretically consistent way.
- Instead of using the familar Cox-Ingersoll-Ross (CIR) model to describe the factors, we employ a vector GARCH-in mean model.

Perspective of this paper

- Risk averse investors require a risk premium
- The no-arbitrage condition for testing market efficiency must take account of this risk premium
- The SDF model provides a flexible, general way of model risk that is consistent with general equilibrium theory.
- To help identify the sources of risk, observable factors are used instead of latent factors.
- Since risk is due to the *covariance* of returns with other variables, the joint distribution of these variables is required, i.e. multi-variate models.
- Since the risk premium will be a *conditional* covariance, ARCH type models are an obvious choice.
- Since the *expected return* is affected by the risk premium, ARCH-in-mean effects are required.
- Taken together this leads to vector ARCH (or GARCH)-in-mean models.
- Because we want to use observable macro variables, we have to use low frequency finance data.
- An example of this approach to optimal asset allocation is provided by Flavin and Wickens (1999). This paper extends the methodology to the FOREX market.

Previous work on SDF models

- Usually based on the Duffie-Kan (1996) class of affine models. This involves the use of unobservable factors which are extracted from the asset returns.
- Term structure papers using unobservable affine factor models include:

  Duffie and Kan (1996)

  Duffie and Singleton (1997)

  Backus, Foresi and Telmer (1998)

  Remolona, Wickens and Gong (1999).
- Using SDF models for currency pricing creates new conceptual problems
- Aim is to price currency risk and hence derive the foreign exchange risk premium
- Pure currency risk arises when the underlying assets are risk free in terms of their domestic currencies.
- When investors are risk-neutral, the arbitrage condition is given by uncovered interest parity. In this case their is no risk premium and the domestic and foreign investor is treated symmetrically.
- When investors are risk-averse, the currency risk premium for domestic investors may be different from that of the foreign investor.
- The relative size of domestic and foreign investors may also matter. In other words, there may be portfolio effects, and these could reflect differences in attitudes to risk between investors.
- In the case of complete markets these complications do not arise.

Currency pricing studies

- Backus, Foresi and Telmer (1996) use the Duffie-Kan approach
- Hollifield and Yaron (1999) use a higher order expansion of the no–arbitrage condition with two observable factors (money and inflation) generated by a CIR model.
- They estimate the model using GMM on the moment condition.
- They conclude that:

- the model must have significant real risk
- the monetary shocks should result in small inflation risk
- but lead to volatility in the real pricing kernel.

- Mark (1985) based a test of efficiency on the Euler condition and used GMM estimation. Implausibly large estimates of the coefficient of relative risk aversion (CRRA) were obtained, and the restrictions of the theory were rejected.
- Kaminsky and Peruga (1990) adopted an approach to testing the general equilibrium model that is similar to the SDF model, and they employed a vector GARCH specification of the error structure. Their findings were similar to those of Mark in that they obtained an implausibly large estimate the CRRA, but they could not reject the theoretical restrictions. These two studies were based on monthly data.
- Baillie and Bollerslev (1991) used weekly data, allowed for moving average dynamics of the conditional mean of the excess return, and used a univariate GARCH model for each variable from which they derived an estimate of the risk premium. Their findings were similar to those of Kaminsky and Peruga in that all of the ARCH-in-mean effects were insignificant.

The Stochastic Discount Factor asset pricing model

The SDF model is based on the simple idea that P_t, the price of an asset in period t, is the discounted value of its pay-off X_{t+1} in period $t+1$:

$$P_t = E_t[M_{t+1}X_{t+1}],$$

where M_{t+1} is a stochastic discount factor.

The problem that occupies finance is how to choose M_{t+1}.

The pricing equation can also be written as

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

where $R_{t+1} = X_{t+1}/P_t = 1 + r_{t+1}$ is the gross return.

We can re-express this as

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

The expected return on the asset is therefore given by

$$E_t(R_{t+1}) = \frac{1 - Cov_t(M_{t+1}, R_{t+1})}{E_t(M_{t+1})}. \quad (2)$$

No-arbitrage condition

- a key condition in finance required for a self-financing portfolio (i.e. between asset returns) to eliminate profit opportunities

- commonly ignored in econometric work

The SDF pricing equation holds whether the asset is risky or risk-free. If it is risk-free, then its payoff 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 net risk-free rate of return. As a result the risk-free return must satisfy

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

This implies
(i)

$$E_t(M_{t+1}) = \frac{1}{1 + r_t^f}$$

(ii) 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. $E_t\xi_{t+1} = 0$.

A self-financing portfolio consisting of holding the risky asset and selling the risk-free has the expected return

$$E_t r_{t+1} - r_t^f = -(1 + r_t^f)Cov_t(M_{t+1}, R_{t+1}).$$

where $R_{t+1} = 1 + r_{t+1}$

The right-hand side of this equation is the risk-premium; it is the extra return over the risk-free return that is required to compensate investors for holding the risk-free asset.

This is the no-arbitrage condition relating the expected excess return on a risky asset over the risk-free to the covariance between the stochastic discount factor and the gross return on the risky asset. .

As r_t^f is known at time t,

$$V_t(R_{t+1}) = V_t(r_{t+1} - r_t^f)$$

$$Cov_t(M_{t+1}, R_{t+1}) = Cov_t(M_{t+1}, r_{t+1} - r_t^f)$$

and the no-arbitrage condition can be expressed as

$$E_t(r_{t+1} - r_t^f) = -(1 + r_t^f)Cov_t(M_{t+1}, r_{t+1} - r_t^f)$$

Thus the SDF asset pricing model yields both a general theory of asset pricing and a general theory of risk.

The assumption of log-normality

If $\ln x$ is $N(\mu, \sigma^2)$ then $\ln E(x) = \mu + \frac{\sigma^2}{2}$.

If gross returns and the stochastic discount factor are jointly distributed as lognormal then

(i) risky asset

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

(ii) risk-free asset

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

(iii) excess return

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

This is our key no-arbitrage condition under log-normality

The choice of SDF, M_{t+1}

(1) General equilibrium intertemporal consumption CAPM (CCAPM)
(2) CAPM
(3) Multi-factor models
(4) Latent affine factor models

(1) CCAPM

Asset prices derive their ultimate value from the expected consumption stream of the private sector. Thus

$$\max_{\{c_t, c_{t+1}, \dots\}} \mathcal{U}_t = U(C_t) + \beta E_t[U(C_{t+1})] + \beta^2 E_t[U(C_{t+2})] + \cdots$$

or

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

subject to the budget constraint

$$C_t + W_{t+1} = y_t + W_t(1 + r_t)$$

This gives the Euler equation

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

Hence,

$$\begin{aligned} M_{t+1} &\equiv \frac{\beta U'_{t+1}}{U'_t} \\ &\simeq \beta[\frac{U'_t + \Delta C_{t+1} U''_t}{U'_t}] = \beta[1 + \frac{\Delta C_{t+1}}{C_t} . \frac{C_t U''_t}{U'_t}] \\ &= \beta[1 - \sigma_t \frac{\Delta C_{t+1}}{C_t}] \end{aligned} \quad (3)$$

where

$$\sigma_t = -\frac{C_t U''_t}{U'_t} \geqslant 0 \text{ as } U''_t \leq 0$$

is the coefficient of relative risk aversion.

The no-arbitrage condition can therefore be written
(i) general

$$E_t r_{t+1} - r_t^f = \beta \sigma_t (1 + r_t^f) Cov_t(\frac{\Delta C_{t+1}}{C_t}, R_{t+1}).$$

(ii) log-normal

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

An asset is risky if for states of nature in which returns are low, the intertemporal marginal rate of substitution in consumption M_{t+1} is high.

Since M_{t+1} will be high if future consumption is low, a risky asset is one which yields low returns in states for which consumers also have low consumption. This is a common feature of business cycles. In recessions, consumption growth falls and so does the stock market, and hence stock returns. In booms, consumption growth and stock returns are high. To ensure that consumers are willing to hold a risky asset, it must have an expected return that is higher than that of the risk-free asset, which has the same return in *all* states of nature. Put another way, the returns on assets that are least affected by the business cycle will have the smaller risk premia because they have a lower correlation with consumption growth. Examples of such defensive assets are utility stocks and government bonds.

(2) CAPM

Static CAPM states that the risk premium on the risky asset can be explained in terms of the covariance of the return on the risky asset with the return on the market portfolio.

The key results are:

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

Hence,

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

The rate of return on the market is given by $(1 + r_{t+1}^m) = \frac{W_{t+1}}{W_t}$, which implies that $r_{t+1}^m = \frac{\Delta W_{t+1}}{W_t}$. Therefore,

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

Choice of M_{t+1}

(1) CCAPM: $M_{t+1} = \beta\left(1 - \sigma_t \frac{\Delta C_{t+1}}{C_t}\right)$

(2) CAPM: $M_{t+1} = \sigma_t \frac{\Delta W_{t+1}}{W_t}$

(3) Multi-factor models

For convenience we just consider the model under log-normality

(i) Single factor models

$$m_{t+1} = a + bz_{t+1}.$$

For $\sigma_t = \sigma$, a constant

$$z_{t+1} = \begin{cases} \frac{\Delta C_{t+1}}{C_t} & CCAPM \\ r^m_{t+1} & CAPM \end{cases}$$

(ii) Multi-factor models

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

or

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

Affine models

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

No-arbitrage condition for multi-factor models

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

where the f_{it} are known as *common factors.*

Multi-asset multi-factor models

Each asset $r_{j,t+1}$ will have a similar equation indexed by j

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

In matrix terms

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

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

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.

(4) Latent variable affine factor models

The most popular model in finance literature for asset pricing
Based on the idea that the factors are not directly observable.

In contrast, for CCAPM and CAPM the factors are observable

Two well-known examples
(i) Vasicek model
(ii) Cox, Ingersoll and Ross (CIR) model

Both assume that the factors are a mean-reverting AR(1) process

(i) Vasicek model

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

The no-arbitrage condition is

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

(ii) Cox, Ingersoll and Ross (CIR) model

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

The no-arbitrage condition is

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

Problems:
Both give a restrictive theory of risk
(i) for the Vasicek model the risk premium is constant
(ii) for the CIR model it is a linear function of z_t.

Macroeconomic sources of risk

An advantage of using observable, instead of unobservable, factor models is that we can identify and estimate the fundamental sources of risk.

In

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

the variables $z_{i,t+1}$ can be any number of observable (or unobservable) variables

The idea is to model the joint distribution of $\mathbf{x}_{t+1} = (r_{t+1} - r_t^f,\ z_{1,t+1},\ z_{2,t+1}, ...)'$

The no-arbitrage condition is then the mean of the conditional distribution of $r_{t+1} - r_t^f$

This should include terms from the time-varying conditional covariance matrix of the joint distribution in order to capture the risk premium.

The $z_{i,t+1}$ variables may, or may not be, "PRICED" sources of risk.

If they are priced then $b_{ij} \neq 0$.

If they are not priced then $b_{ij} = 0$.

Even if a factor is not priced it may still be jointly distributed with the excess return or with the priced factors - i.e. not be distributed independently of them - and should therefore be included in the model. For example, a non-priced variable could affect the conditional covariance matrix of the joint distribution.

Multivariate conditional heteroskedasticity model

A convenient way to implement the observable SDF model is to use the multivariate GARCH-in-mean model, but not of course the multivariate GARCH model.

Many practical problems arise in using the MGM model.

(i) 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.

(ii) 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 supply.

(iii) Problem of extreme values and the choice of distribution

Often excess returns exhibit extreme values that suggest Normality is not a good assumption.

Extreme values cause the GARCH model to display near integrated variances.

Solutions are to use non-Normal distributions and stochastic volatility models.

A problem is that a multivariate SV model with conditional covariances in the mean does not exist.

The MGM model can be written

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

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

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

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

The *vech* operator converts the the lower triangle of a symmetric matrix into a vector. A key feature of this approach is that the first equation of (**??**) must be restricted so that it satisfies the condition of no arbitrage. Thus the first row of $\mathbf{\Gamma}$ is zero and the first row of $\mathbf{\Phi}$ is $(-\frac{1}{2},\ -b_{11},\ -b_{12},\ -b_{13},...)$. Contrasting the MGM model with a VAR, we note that in a VAR $\mathbf{\Phi}$ is implictly zero and $\mathbf{H}_{t+1}$ is assumed to be homoskedastic.[1]

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

$$vech(\mathbf{H}_{t+1}) = \mathbf{\Lambda} + \Sigma_{i=0}^{p-1}\mathbf{\Phi}_i vech(\mathbf{H}_{t-i}) + \Sigma_{j=0}^{q-1}\mathbf{\Theta}_j vech(\varepsilon_{t-j}\varepsilon_{t-j}') \tag{4}$$

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):

$$\mathbf{H}_{t+1} = \mathbf{V}'\mathbf{V} + \mathbf{A}'(\mathbf{H}_t - \mathbf{V}'\mathbf{V})\mathbf{A} + \mathbf{B}'(\varepsilon_t\varepsilon_t' - \mathbf{V}'\mathbf{V})\mathbf{B}. \tag{5}$$

The first term on the right-hand side of equation (5) 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

[1] 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.

good performance in comparison with the more general BEKK model. It can be written

$$\mathbf{h}_{ij,t+1} = \rho_{ij}[\mathbf{h}_{ii,t+1} \times \mathbf{h}_{jj,t+1}]^{\frac{1}{2}} \tag{6}$$

$$\mathbf{h}_{ii,t+1} = v_i + a_i \mathbf{h}_{ii,t} + b_i {\boldsymbol{\varepsilon}_{it}}^2 \tag{7}$$

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 sometimes the constant correlation model may be the best that one can achieve. Further variants of these models can, of course, be considered.Using the

The implications of the SDF model for empirical finance

1. A risk premium should be included in the model

2. This is a (possibly linear) function of the conditional covariance between the factors and the excess return.

3. Hence, in general in empirical finance the model must be multi-variate, not univariate, as the joint distribution of the excess return and the factors is required to model the risk premium.

4. 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.

5. There may be other relevant variables that should be included in the model of the excess return.

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

(i) 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. eg CIR model

In these cases it would be possible to use a VAR model.

(ii) It is often possible to estimate the parameters and to carry out statistical tests using the Euler equation itself through, for example, GMM estimation.

GMM estimation is best suited to testing. It doesn't provide a direct estimate of the risk premium.

FOREX Risk Premium

1. Market structure

Domestic investor model

Only domestic investors affect the exchange rate as a result of their purchases of the foreign asset which is denominated in foreign currency

- Return on the risky foreign asset expressed in domestic currency is $\Delta s_{t+1} + i_t^*$, the return on the foreign bond i_{t+1}^* which is risk-free in terms of foreign currency
- The risk-free return $r_t^f = i_t$ is the return on the domestic bond which is also expressed in domestic currency
- Recalling the notation used earlier that the excess return is

  $$R(t+1) = i^*(t) + \Delta s(t+1) - i(t)$$

  and assuming log-normality, the no-arbitrage condition is

  $$E_t[R(t+1)] + \frac{1}{2}V_t[R(t+1)] = -Cov_t[m(t+1), R(t+1)]$$

- Hence

  $$\phi(t) = -\frac{1}{2}V_t[R(t+1)] - Cov_t[m(t+1), R(t+1)]$$

- As $s(t+1)$ is the only part of $R(t+1)$ that is unknown at t,

  $$E_t[R(t+1)] + \frac{1}{2}V_t[\Delta s(t+1)] = -Cov_t[m(t+1), \Delta s(t+1)]$$

- Hence the risk premium arises from uncertainty about the future spot exchange rate and its correlation with the discount factor $-m(t+1)$
- The higher the rate at which foreign returns are discounted
  1. the greater they must be
  2. hence the greater the exchnage depreciation required

Foreign investor model

This assumes that only the foreign investor undertakes FOREX transactions. Hence

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

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

where m^* is measured in foreign currency and $R^* = -R$

Domestic and foreign investor model
In practice, both investors will be carrying out FOREX trades.

- To combine the two types of investror we add the domestic and foreign equations

  $$E_t[R(t+1)] = -Cov_t[\frac{1}{2}(m(t+1) + m^*(t+1)), R(t+1)]$$

i.e. there is no Jensen effect

- Subtracting gives

  $$V_t[R(t+1)] = Cov_t[(m^*(t+1) - m(t+1)), R(t+1)]$$

- Hence

  $$\Delta s(t+1) = m^*(t+1) - m(t+1) + \eta(t+1)]$$

  $$Cov_t[\Delta s(t+1), \eta(t+1)] = 0$$

- Implies a linear relation between
  $V_t[\Delta s(t+1)]$, $Cov_t[(m^*(t+1), \Delta s(t+1)]$ and $Cov_t[m(t+1)), \Delta s(t+1)]$
  Hence only two terms are required.

- But if there is measurement error in the proxy for the discount factor, then the data will not hold in practice.

Complete markets

$$m^*(t+1) = m(t+1) + \Delta s(t+1)$$

i.e. the discount factors are identical when expressed in the same currency

2. Choice of SDFs

General equilibrium model (C-CAPM)

- The general equilibrium model with power utility function implies

  $U[C(t)] = [C(t)^{1-\sigma} - 1]/(1-\sigma)$

- Hence

$$\begin{aligned} M(t+1) &= \delta \left[\frac{U'[C(t+1)]}{U'[C(t)]} \right] \frac{P(t)}{P(t+1)} \\ &= \delta \left[\frac{C(t+1)}{C(t)} \right]^{-\sigma} \frac{P(t)}{P(t+1)} \end{aligned}$$

- Taking logs

  $m(t+1) = \ln\delta - \sigma\Delta c(t+1) - \Delta p(t+1)$

- More generally can assume $m(t)$ is a linear function of observable macroeconomic variables $z(t)$, namely

  $m(t+1) = \beta' z(t+1) + \xi(t+1)$

$\xi(t)$ represents omitted factors. It is assumed that $\xi(t)$ is orthogonal to $z(t)$.

- Alternatively $z(t)$ is a single variable that measures $m(t)$ with error when $\xi(t)$ is correlated with $z(t)$ but not with $m(t)$.

- Can now write

$$\begin{aligned} & E_t[R(t+1)] + \frac{1}{2}V_t[R(t+1)] \\ = \ & -\beta' Cov_t[z(t+1), R(t+1)] - Cov_t[\xi(t+1), R(t+1)] \end{aligned}$$

- Aim is to proxy $Cov_t[z(t+1), R(t+1)]$.

- Need the conditional covariance from the joint conditional distribution of $\{R(t+1),\ z(t+1)\}$

CAPM & pure currency risk - monetary model

- 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

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

W_t = nominal financial wealth
R^W_{t+1} = the nominal return on wealth

- The discount factors can be obtained from the variables that explain this portfolio return.
- We assume that the portfolio consists of hedged and unhedged currency and so the element that is unknown in R^W_{t+1} is the future spot exchange rate
- We use the monetary model of the exchange rate to explain this, and hence to provide the macroeconomic factors.
- These are the rates of growth of the US and UK money supplies, and the output growth rates.

Econometric Model

- Use vector GARCH-in-mean: a mixture of a VAR and MGARCH-M
- Need the joint conditional distribution of the stationary vector $\mathbf{x}(t+1) = \{R(t+1), z(t+1)\}'$
- To be arbitrage-free this joint conditional distribution should satisfy the theoretical restrictions on $E_t[R(t+1)]$

$$\mathbf{x}(t+1) = \boldsymbol{\alpha} + \mathbf{\Gamma x}(t) + \mathbf{\Phi g}(t) + \varepsilon(\mathbf{t+1})$$

$$\boldsymbol{\varepsilon}(\mathbf{t+1}) \mid \Psi(t) \sim N[0, \mathbf{H}(t+1)]$$

$$\mathbf{g}(t) = vech\{\mathbf{H}(\mathbf{t+1})\} \tag{8}$$

Constant correlation MGARCH

- $$\mathbf{h}_{ij}(t+1) = \rho_{ij}[\mathbf{h}_{ii}(t+1) \times \mathbf{h}_{jj}(t+1)]^{\frac{1}{2}}$$

  $$\mathbf{h}_{ii}(t+1) = v_i + a_i\mathbf{h}_{ii}(t) + b_i\boldsymbol{\varepsilon}_i(t)^2$$

- where $\mathbf{h}_{ii}(t+1)$ has a GARCH(1,1) structure
- ρ_{ij} = correlation between $\boldsymbol{\varepsilon}_i(t+1)$ and $\boldsymbol{\varepsilon}_j(t+1)$

BEKK

$$\mathbf{H}(t+1) = \mathbf{V}'\mathbf{V} + \mathbf{A}'[\mathbf{H}(t)\mathbf{H}(t)' - \mathbf{V}'\mathbf{V}]\mathbf{A} + \mathbf{B}'[\varepsilon(t)\varepsilon(t)' - \mathbf{V}'\mathbf{V}]\mathbf{B}$$

- Possible restrictions: $\mathbf{V}$ lower triangular, $\mathbf{A}$ and $\mathbf{B}$ symmetric matrices.

Estimation

- BEKK is the most general model, but often fails to converge when there are a large number of variables
- Constant correlation is less general, but usually converges even when there are a large number of variables - used here.
- Single equation constant correlation model - a possible alternative when the system constant correlation fails to converge. Not used here

Single equation method
For the excess return equation estimate

$$\mathbf{E}_t[\mathbf{R}(\mathbf{t+1})]+\frac{1}{2}\mathbf{V}_t[\mathbf{R}(\mathbf{t+1})] = -\sum_{i=1}\boldsymbol{\theta}_i \hat{V}_t [\mathbf{z}_i(\mathbf{t+1})]^{\frac{1}{2}}\mathbf{V}_t[\mathbf{R}(\mathbf{t+1})]^{\frac{1}{2}} + \boldsymbol{\varepsilon}_R(\mathbf{t+1})$$

$$\boldsymbol{\varepsilon}_R(\mathbf{t+1}) \mid \boldsymbol{\Psi}(\mathbf{t}) \sim \mathbf{N}\{\mathbf{0}, \mathbf{V}[\mathbf{R}(\mathbf{t+1})]\}$$

$$\mathbf{V}[\mathbf{R}(\mathbf{t+1})] = \mathbf{v}_R + \mathbf{a}_R\mathbf{V}[\mathbf{R}(\mathbf{t})] + \mathbf{b}_R\boldsymbol{\varepsilon}_R(\mathbf{t})^2$$

$\hat{V}_t [\mathbf{z}_i(\mathbf{t+1})]^{\frac{1}{2}}$ = prior estimate obtained by univariate GARCH estimation

$$\boldsymbol{\theta}_i = \boldsymbol{\Phi}_{Ri}\boldsymbol{\beta}_i\boldsymbol{\rho}_{Ri} \qquad i = 1, ..., n$$

Empirical models

1. Market structure

US investor

$$R(t+1) + \frac{1}{2}V_t[R(t+1)] = \phi^{us\prime} C_t^{us}(t+1) + \boldsymbol{\varepsilon}_1(t+1)$$

UK investor

$$R(t+1) - \frac{1}{2}V_t[R(t+1)] = \phi^{uk\prime} C_t^{uk}(t+1) + \boldsymbol{\varepsilon}_1(t+1)$$

US and UK investors

$$R(t+1) = \phi^{us\prime} C_t^{us}(t+1) + \phi^{uk\prime} C_t^{uk}(t+1) + \boldsymbol{\varepsilon}_1(t+1)$$

General alternative model

$$\begin{aligned} R(t+1) &= \gamma_1 R(t) + \gamma_2[f(t) - s(t)] + V_t[R(t+1)] \\ &\quad + \phi^{us\prime} C_t^{us}(t+1) + \phi^{uk\prime} C_t^{uk}(t+1) + \boldsymbol{\varepsilon}_1(t+1) \end{aligned}$$

$C_t(t+1)$ is a vector with j^{th} element $[\mathbf{H}_{11}(t+1) \times \mathbf{H}_{jj}(t+1)]^{\frac{1}{2}}$.

2. Alternative SDF models

Benchmark model
$x(t+1)' = \{R(t+1),\ f(t+1) - s(t+1),\ \Delta i^{us}(t+1)\}$

C-CAPM
US investor: $x(t+1)' = \{R(t+1),\ \Delta c^{us}(t+1),\ \Delta p^{us}(t+1)\}$
UK investor: $x(t+1)' = \{R(t+1),\ \Delta c^{uk}(t+1),\ \Delta p^{uk}(t+1)\}$
Alternative hypothesis:
$x(t+1)' = \{R(t+1),\ \Delta c^{us}(t+1),\ \Delta p^{us},\ \Delta c^{uk}(t+1), \Delta p^{uk}(t+1)\}$

Monetary model (CAPM)
US investor: $x(t+1)' = \{R(t+1),\ \Delta m^{us}(t+1)\}$
UK investor: $x(t+1)' = \{R(t+1),\ \Delta m^{uk}(t+1)\}$
Alternative hypothesis:
$x(t+1)' = \{R(t+1),\ \Delta m^{us}(t+1),\ \Delta m^{uk}(t+1)\}$

General combined model
US investor: $x(t+1)' = \{R(t+1),\ \Delta c^{us}(t+1),\ \Delta p^{us}(t+1),\ \Delta m^{us}(t+1)\}$
UK investor: $x(t+1)' = \{R(t+1),\ \Delta c^{uk}(t+1),\ \Delta p^{uk}(t+1),\ \Delta m^{uk}(t+1)\}$
Alternative hypothesis:
$$x(t+1)' = \{R(t+1), \Delta c^{us}(t+1), \Delta p^{us}, \Delta m^{us}(t+1), \Delta c^{uk}(t+1), \Delta p^{uk}(t+1), \Delta m^{uk}(t+1)\}$$

3. Switching conditional variance structure

- US and UK investors may have different attitudes to risk
- A dollar-based investor holding sterling assets faces losses from exchange risk when the dollar unexpectedly appreciates, i.e. when $\Delta s(t+1)-E_t\Delta s(t+1) < 0$; the interest differential is supposed to compensate for any expected appreciation.
- Sterling-based investors face losses from exchange risk when $\Delta s(t+1)-E_t\Delta s(t+1) > 0$.
- This suggests that a different model of the conditional variance of the excess return - i.e. of $\Delta s(t+1)$ - should be used depending on whether $\Delta s(t+1) - E_t\Delta s(t+1) = \boldsymbol{\varepsilon}_1(t) \lesseqgtr 0$.

  $$\mathbf{h}_{11}(t+1) = v_1 + a_1\mathbf{h}_{11}(t) + b_1\boldsymbol{\varepsilon}_1(t)^2 + \delta b_1^*\boldsymbol{\varepsilon}_1(t)^2$$

- $\delta = 0$ if $\Delta s(t+1) \leqslant 0$
- $\delta = 1$ if $\Delta s(t+1) > 0$
- In order to obtain estimates of the conditional variance for the general model we found it necessary to add the further restriction that $a_1 + b_1 + \delta b_1^* < 1$.

4. Excess kurtosis

- It is well known that exchange rates exhibit excess kurtosis relative to the Normal distribution
- One way of trying to take account of this is to use the t-distribution instead of the Normal
- Technical problem: If the logs of the excess return and the discount rate have a multi-variate t-distribution, then taking logs of the SDF model has the problem that the moment-generating function of the t-distribution doesn't exist
- For all of our data we estimate there are at least 9 degrees of freedom, and not 3 as for the Normal.

Data

Monthly from 1975.1 to 1997.12 for the US and the UK.
Consists of the US dollar-sterling exchange rate
One month Eurocurrency interest rates
Real retail sales (the nearest we can get to monthly real consumption)
The CPI for the US and the RPI for the UK
The monetary base for the US and M0 for the UK.
All data are expressed in annualized rates.
The forward premium is calculated as the US minus the UK interest rate.

- Since a unit root cannot be rejected for each series, we use stationary transformations of the data either in the form of first differences or spreads.

- All except the US monetary base also showed strong ARCH. This provides some justification for the use of monthly data in GARCH-based models of risk premia.

Estimates

- In columns one and two the coefficient of the conditional variance of the excess return is imposed as $\frac{1}{2400} = 0.0004166$, and not 0.5
- Column three is the two investor model and so this coefficient is set to zero
- Column four is the general unrestricted version
- Columns three and four have no switching effect
- Columns five and six repeat three and four, but also allow for switching.

Benchmark model

- Tables IIa and IIb - little qualitative difference between a nd b
- None of the conditional covariances with the excess return is significant in any of the models or with either distribution
- One measure of being able to successfully measure the FOREX risk premium is that it eliminates the forward premium puzzle. In other words, the biases in the estimate of the forward premium should be removed so that the estimate is insignificant from zero.
- Having failed to provide a significant model of the FOREX premium, it is not surprising that the forward premium retains its significance
- We also find that the lagged excess return is significant
- Thus, the model used in traditional tests of the FOREX market when reformulated so that the conditional heteroskedasticity in the joint distribution of the variables is taken into account, is unable to provide a significant measure of the FOREX risk premium.

CCAPM

- Tables IIIa and IIIb
- The conditional covariance terms in the two single investor models (columns 1 and 2) are now significant at the 10% level for the Normal distribution.
- In the two investor model (column 3) only the conditional covariance with US consumption is significant
- In the general model (column 4) none of the conditional covariances is significant for the Normal distribution, but the covariance with UK consumption is significant
- Tests of the restriction on the coefficient of the own conditional variance do not reject the restriction
- The outcome of this test is the same for all of the models. This is mainly because the theoretical value is so small and the unrestricted coefficient is not estimated precisely enough
- The estimates in Table III are for the coefficients of the product of the conditional standard deviations (i.e. of $\rho_{1,j}^{us}\beta_j^{us}$), not for the conditional covariances required by the theory (i.e. of β_j^{us}). But by using the unconditional correlations with the excess return reported in Table I, it is possible to recover estimates of the β_j^{us}.
- The coefficient of the covariance of the excess return with consumption growth is σ^{us}, the coefficient of relative risk aversion, and that with inflation is unity.
- The implied estimates of these coefficients for the US investor are -289 and 43800, respectively
- For the UK investor σ^{uk} is -283 and of the coefficient of the covariance with inflation is 10320
- For the combined model only the coefficient of the conditional covariance with US consumption is significant and this is -410

- All of the estimates of the coefficient of relative risk aversion therefore have the wrong sign and are very large
- The signs for the covariance with inflation are correct in the single investor and combined models, but are significant only in the single investor models but the size of the coefficients is far too large
- In the general model none of the conditional covariances is significant.
- And the lagged excess return and forward premium retain their significance
- In other words, the forward premium puzzle is not resolved
- Conclusion: the estimates are not consistent with the theoretical predictions and the theory does not seem able to provide a satisfactory model of the FOREX risk premium.

CAPM - Monetary model

- Tables IVa and IVb
- Expect 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
- The single US investor model performs best assuming a t-distribution, when the estimates are bordering on significance and also have the correct sign
- 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
- Conclusion: these results show considerable support for the monetary model and hence for the traditional models of currency risk
- But the continued significance of the lagged excess return and forward premium in the general model indicates that the forward premium puzzle is not resolved even if output, and in some cases money, seem to be significant sources of FOREX risk.

Combined model

- This model has no explicit theoretical justification
- Tables Va and Vb
- In the two single investor models and the two investor model the conditional covariances with consumption and output are the only significant variables
- But only the output terms have the theoretically correct sign
- This suggests that the factors may come from a mixture of the C-CAPM and CAPM models
- In the general model the lagged excess return and the forward premium retain their significance, showing once more that the forward premium puzzle remains.

Switching model

- The aim with the switching model is to allow investors to have different attitudes to risk.
- Columns 5 and 6 of Tables II-V are a re-estimate of columns 3 and 4 and include an additional term in the expression for the conditional variance of the excess return to allow for a shift in the impact of last period's error.
- The switching term is significant in some of the models. It is most important for the two investor case within the monetary and combined models, and slightly more significant for the Normal than the t-distribution estimates
- Including the switching term has a major impact on the estimates of the coefficients of the conditional covariances in the monetary model
- In the two investor model, all of these coefficents are significant and have the correct sign
- In the general model, the estimates are similar to those without switching effects
- These results indicate that US and UK investors may have different attitudes to risk
- They also lend strong support for the monetary model, but still without eliminating the forward premium puzzle.

Conclusions

- In this paper we have considered the problem of measuring macroeconomic sources of financial risk
- We have used the stochastic discount factor model to provide a general theory of asset pricing
- We have shown how to extend this to take account of potential macroeconomic sources of risk
- We have described in detail how this can be implemented empirically using the multivariate GARCH-in-mean model.
- We have argued that ARCH-in mean effects must be included in order that the empirical model may satisfy the no-arbitrage condition
- And that as the risk premium is a conditional covariance, the model must also be multivariate
- In analysing the FOREX market, we have shown that it is important to take account of the fact that market participants may be dollar or sterling based
- The empirical results provide no support for either the benchmark model or the inter-temporal consumption-based CAPM
- One of the most interesting results of the paper is the support provided for the monetary model.
- This performs best is the two investor monetary model which includes switching effects to allow for different attitudes to risk among US and UK investors
- The zero restriction on the own conditional variance is satisfied in all of the estimates of the two investor model
- There is little to choose between the Normal and t-distribution estimates, but the former are to be preferred on the grounds that they are slightly better, and there is a logical difficulty with using the t-distribution in the SDF framework

- The main problem that remains is the continued presence of the forward premium puzzle. Even our preferred model does not eliminate this
- Modelling risk using observable factors within the SDF framework is, in our view, a considerable advance on existing work
- Discovering the potential usefulness of the monetary model to capture the FOREX premium is a further advance in our knowledge
- It has the interesting implication that the FOREX risk premium may be more associated with pure currency risk than general equilibrium considerations
- A number of problems still remain

  1. . Once general equilibrium models fail it is not clear how to choose the variables from which to measure the discount factor. The SDF model itself provides no guidance
  2. Mis-measuring the discount factor, for example, by using the wrong variables, may greatly impair the usefulness of the SDF model
  3. The use of observable sources of macroeconomic risk makes severe data demands. Ideally high frequency macro data is needed, but most macroeconomic data are not widely available and then mainly at monthly intervals. This curtails the amount of heteroskedasticity in the explanatory variables
  4. In order to provide an adequate representation of the theory, the VGARCHM model must be highly parameterized. This, together with the lack of heteroskedasticity in the data makes the numerical convergence difficult and the optimization a lengthy procedure
  5. Further advances in the use of this general approach will depend in large part in finding satisfactory solutions to these problems.

Main Issues

1. Is FOREX correctly priced?
2. Is the FOREX market efficient?

- In effect, the efficiency of FOREX market follows from whether or not FOREX is correctly priced.

Problems to be addressed

1. Why is FOREX an asset price?
2.How should FOREX be priced?
3. Is UIP (uncovered interest parity) a n-arbitrage condition?
4. How should FOREX market efficiency be tested?
5. What is the foorward premium puzzle?
6. How can the FOREX premium be measured?
7. Are there serious anomalies in pricing FOREX?
8. How good are the professionals at pricing FOREX and predicting exchange rate movements?

FOREX Market

Foreign exchange is required for transactions on
current account - goods and services
capital account - financial assets (purchases of bonds, bonds, PFI, M&A etc, reserves)

The exchange rate is determined by the total demand and supply of a currency - as in most markets

It used to be widely assumed that exchange rates were determined primarily by current account transactions

In fact since the widespread removal of capital controls, the capital account is the main determinant.

About 95% of all FOREX transactions are associated with the capital account

EVERY DAY nearly half of the world's FOREX activity takes place through London

It amounts to about HALF on the UK's ANNUAL GDP

Conclusion: need to focus on a capital account explanation

Uncovered Interest Parity (UIP)

This is the key no-arbitrage condition in international bond markets. We consider one-period bonds (bills)

An investor has two choices:

1. Invest in a domestic bond, RISK-FREE in the domestic currency in nominal terms,

1-period return in domestic currency $= i_t$.

Pay-off in period $t+1$ is $1+i_t$

2. Invest in a foreign bond risk-free in terms of foreign currency

1-period return in foreign currency $= i_t^*$

To compare the two investments we need to measure their pay-offs in the same currency.

If the domestic investor invests in the foreign bond, there are three steps to carry out

(i) convert X units of domestic currency into foreign currency at the current (or spot) rate S_t

S_t= domestic price of foreign exchange, or the number of units of domestic currency required to purchase one unit of foreign currency.

An increase in S_t implies a depreciation of domestic currency.

The US dollar depreciates against sterling if the exchange rate moves from \$1.5 to \$2 to the £.

Hence for the UK investor $£X \rightarrow \$\frac{X}{S_t}$

(ii) Investing $\$\frac{X}{S_t}$ in a US bond at the rate i_t^* gives the pay-off in period $t+1$ of $\$\frac{X}{S_t}(1+i_t^*)$

(iii) Converting the proceeds into domestic currency at S_{t+1}, the spot rate prevailing in period $t+1$, gives the pay-off $£\frac{S_{t+1}}{S_t}X(1+i_t^*)$

As S_{t+1} is unknown at time t, so is the pay-off. The expected pay-off (conditional on information available at time t) is $E_t[£\frac{S_{t+1}}{S_t}X(1+i_t^*)]$

In order for the investor to be indifferent between the two investments the expected pay-offs when expressed in the same currency (domestic or foreign currency) must be the same. In terms of domestic currency, the pay-off from the domestic investment is certain. Thus

$1+i_t = E_t[\frac{S_{t+1}}{S_t}(1+i_t^*)]$

This is called the UNCOVERED INTEREST PARITY condition.

It can be expressed differently as

$1 + i_t = E_t[(1 + \frac{S_{t+1} - S_t}{S_t})(1 + i_t^*)]$

or approximately as

$i_t = i_t^* + E_t[\frac{\Delta S_{t+1}}{S_t}] = i_{*t} + E_t[\Delta s_{t+1}]$

where $s = \ln S$.

This can be interpreted as saying that if the domestic exchange rate is expected to depreciate ($E_t[\Delta s_{t+1}] > 0$), then investors need to be compensated for holding the domestic bond by receiving a higher rate of return on the domestic than the foreign bond.

Thus UIP is a theory of bond prices across currency zones (i.e. country interest differentials), but it can be converted into a theory of exchange rates by re-writing it as

$E_t[s_{t+1}] = s_t + i_t - i_t^*$

This gives the market's expected future (log) spot rate as linear function of variables known at time t.

Covered Interest Parity (CIP)

If you asked the market for a forecast of next period's spot rate, you would be quoted something called the FORWARD rate. This is

$F_t = S_t \frac{1+i_t}{1+^*_t}$

or, if $f = \ln F$,

$f_t = s_t + i_t - i_t^*$

In other words, the market assumes that UIP holds and so sets

$f_t = E_t[s_{t+1}]$

Thus f_t is assumed to be an unbiased predictor of s_{t+1}. This is the basis of standard tests of FOREX market efficiency.

To avoid the uncertainty associated with not knowing s_{t+1} at time t, investors typically hedge a proportion of their portfolio by taking out a forward contract. This fixes the exchange rate that the foreign bond proceeds at time $t+1$ will be converted into domestic currency at. The guaranteed exchange rate is f.

Although this removes uncertainty, it does not remove risk.

The predicted depreciation of the domestic exchange rate between t and $t+1$ is

$$f_t - s_t = i_t - i_t^*.$$

If $s_{t+1} - s_t > i_t - i_t^*$ then it would have been more profitable to have converted at the actual exchange rate in $t+1$.

Implications of UIP for the exchange rate

Can also write UIP as the forward-looking difference equation

$s_t = E_t[s_{t+1}] + i_t^* - i_t$

Hence, in period $t + 1$,

$s_{t+1} = E_{t+1}[s_{t+2}] + i_{t+1}^* - i_{t+1}$

Thus, taking expectations E_t, and noting that $E_t\{E_{t+1}[x_{t+n}]\} = E_t[x_{t+n}]$ - the law of iterated expectations.

$E_t[s_{t+1}] = E_t E_{t+1}[s_{t+2}] + E_t[i_{t+1}^* - i_{t+1}] = E_t[s_{t+2}] + E_t[i_{t+1}^* - i_{t+1}]$

Hence,

$s_t = E_t[s_{t+2}] + i_t^* - i_t + E_t[i_{t+1}^* - i_{t+1}]$.

or, continuing to substitute forwards,

$s_t = \sum_{k=0}^{\infty} E_t[i_{t+k}^* - i_{t+k}]$.

This says that the spot exchange rate is the sum of all expected future interest differentials, not just the current differential $i_t^* - i_t$.

If the foreign interest rate increases, or is expected to increase at any time in the future, then domestic currency depreciates.

When the central bank raises domestic interest rates, the currency appreciates. But by how much depends on the market's view on how long the interest differential will last.

An increased differential lasting one year will cause the SPOT rate to increase by 12 times more than if the differential is expected to last one month!

Expected future differentials will also cause exchange rate to change.

If when the time comes there is no change to the differential then the change in the exchange rate is called a Peso effect. Subsequently it might look as though the market had behaved irrationally, but actually it had behaved perfectly rationally throughout - given the expectation which turned out to be false.

FOREX Market Efficiency

If it assumed that bonds are priced on the basis of UIP then a prediction of the theory is that

$$f_t = E_t[s_{t+1}]$$

This is the basis of standard tests of FOREX market efficiency.

It is not, however, the only theory of country interest differentials.

A crucial implicit assumption of UIP is that investors are risk neutral, not risk averse.

As a result, there is no risk premium associated with the risky investment of holding the foreign bond.

The risk is due to s_{t+1} being unknown at time t.

If, as seems far more likely, given the prevalence of currency hedging, investors are risk averse then they would require a higher return on the foreign investment to compensate for the FOREX risk. i.e. the risk premium and the foreign interest rate should be positively correlated.

If ρ_t = risk premium for holding the foreign asset then UIP would be replaced by

$$i_t + \rho_t = i_t^* + E_t[\frac{\Delta S_{t+1}}{S_t}] = i_{*t} + E_t[\Delta s_{t+1}]$$

i.e. i_t^* would need to incorporate the risk premium ρ_t and therefore be greater by the amount ρ_t.

Hence, an alternative description of FOREX market efficiency is

$$f_t + \rho_t = E_t[s_{t+1}]$$

Later we consider how to obtain ρ_t.

Testing FOREX market efficiency

1. Levels test

First, we consider basing the test on the unbiasedness hypothesis

$H_0 : f_t = E_t[s_{t+1}]$

The first problem is how to form conditional expectations $E_t[s_{t+1}] = E[s_{t+1}|I_t]$ where I_t is the information set at time t.

It is common to assume weakly rational expectations when the information set consists of current and past values of s_t and f_t.

$I_t = \{s_t, s_{t-1}, ...; f_t, f_{t-1}, ...\}$

Thus $H_0 : E[s_{t+1}|s_t, s_{t-1}, ...; f_t, f_{t-1}, ...] = f_t$

If the innovation (forecasting) error is ε_{t+1} then

$\varepsilon_{t+1} = s_{t+1} - E_t[s_{t+1}]$

Rationality implies that $E_t[\varepsilon_{t+1}] = 0$, i.e. ε_{t+1} is serially uncorrelated.

Hence, under the unbiasedness hypothesis, FOREX market efficiency implies that

$H_0 : s_{t+1} = f_t + \varepsilon_{t+1},$ with $E_t[\varepsilon_{t+1}] = 0$

To test this we need an alternative hypothesis. Consider

$H_1 : s_{t+1} = \alpha + \beta f_t + e_{t+1}$

This gives the testable restrictions under H_0 that $\alpha = 0, \beta = 1$ and $E_t[e_{t+1}] = 0$, i.e. e_{t+1} is serially uncorrelated.

A more general alternative hypothesis that we could consider instead consistent with I_t is

$H_1 : s_{t+1} = \alpha + \sum_{k=0} \beta_k f_{t-k} + \sum_{k=0} \gamma_k s_{t-k} + e_{t+1}$

2. Differences test

An alternative way of writing H_0 is in terms of the ability of the forward spread to predict the change in the spot rate:

$$H_0 : f_t - s_t = E_t[\Delta s_{t+1}]$$

$$H_0 : \Delta s_{t+1} = f_t - s_t + \varepsilon_{t+1}, \text{ with } E_t[\varepsilon_{t+1}] = 0$$

The alternative hypothesis could then be written

$$H_1 : \Delta s_{t+1} = \alpha + \beta(f_t - s_t) + e_{t+1}$$

with $H_0 : \alpha = 0, \beta = 1$ and $E_t[e_{t+1}] = 0$.

In transpires that these two formulations of the test of FOREX market efficiency give very different empirical results.

At first sight this is very surprising. But it turns out to be very informative.

Empirical Evidence

Model:
$s_{t+1} = \alpha + \beta f_t + e_{t+1}$
The estimation sample is: 1975 (2) to 1997 (11)
Dependent variable s(t+1)

Variable	Coefficient	Std.Error	t-value	t-prob	Part. R^2
Constant	0.0416891	0.01796	2.32	0.021	0.0194
f(t)	0.975731	0.01028	94.9	0.000	0.9707

sigma = 0.0464938
RSS = 0.587975039
R^2= 0.970692
F(1,272) = 9009 [0.000]**
log-likelihood = 452.966
DW = 1.2
no. of observations 274
no. of parameters 2
mean(s) = 1.72533, var(s) = 0.0732198

Model:
$\Delta s_{t+1} = \alpha + \beta(f_t - s_t) + e_{t+1}$

The estimation sample is: 1975 (2) to 1997 (11)

Dependent variable Δs(t+1)

Variable	Coefficient	Std.Error	t-value	t-prob	Part. R^2
Constant	-0.00915492	0.003726	-2.46	0.015	0.0217
f(t)-s(t)	-2.94782	1.096	-2.69	0.008	0.0259

sigma = 0.0458865
RSS = 0.572715603
R^2 = 0.0258954
F(1,272) = 7.231 [0.008]**
log-likelihood = 456.569
DW = 1.23
no. of observations 274
no. of parameters 2
mean(Δs) = -0.00245839, var(Δs) = 0.00214577

Implications of Empirical Findings

1. $f_t \approx s_t$

Implies that f_t does not give a good forecast of s_{t+1}.

Hence if $E_t[s_{t+1}] = f_t \approx s_t$
and $s_{t+1} = E_t[s_{t+1}] + \varepsilon_{t+1}$
then $s_{t+1} \approx s_t + \varepsilon_{t+1}$
or $\Delta s_{t+1} \approx \varepsilon_{t+1}$.

Thus s_t is approximately a random walk.

Hence changes in s_t are not predictable.

2. $f_t - s_t$ has the wrong sign and explains little of the variation in Δs_{t+1}.

Thus Δs_{t+1} is negatively correlated with $f_t - s_t$,and NOT positively correlated as the efficient market hypothesis predicts

Implies that the innovations (shocks) ε_{t+1} are the biggest factor affecting s_{t+1} and Δs_{t+1}.

3. The two estimates of β are very different in the two models.

In the levels model $\beta \approx 1$, but in the other model $\beta < 0$, an significantly different from zero - let alone *unity*.

If H_0 were correct this could not happen.

Possible explanation for the estimates of β

If s_t is approximately a random walk then it is a non-stationary I(1) process and not stationary - i.e. I(0).

If $s_t \approx f_t$ then f_t is also I(1).

The levels model regresses s_{t+1} on f_t.

If there are no omitted I(1) variables the estimates are super consistent, i.e. very accurate.

The absence of serial correlation in the residuals is consistent with not omitting and I91) variable.

A further check is whether s_t and f_t are cointegrated. The contegrating vector is (1, -1.002).

The levels model is therefore a cointegrating regression. It shows the long-run relation between s_t and f_t.

It follows that Δs_{t+1} and $f_t - s_t$ are both I(0) variables.

The other model therefore involves stationary variables.

Standard results on omitted variable bias tell us that when the data are stationary and a stationary variable is omitted from the model, the estimates will be biased, possibly badly.

The bias in the estimate of β from OLS estimation of $\Delta s_{t+1} = \alpha + \beta(f_t - s_t) + e_{t+1}$ is

$$E(b_{OLS}) = \frac{cov(\Delta s_{t+1}, f_t - s_t)}{var(\Delta s_{t+1})} = \beta + \frac{cov(e_{t+1}, f_t - s_t)}{var(\Delta s_{t+1})}$$

Why might $cov(e_{t+1}, f_t - s_t) < 0$?

Explanations for these results

1. It is all due to omitting a risk premium
2. FOREX|market anomalies

1. Is it all due to an omitted risk premium?

If investors are risk averse then $f_t + \rho_t = E_t[s_{t+1}]$

Hence $f_t - s_t + \rho_t = E_t[\Delta s_{t+1}]$

and so $s_{t+1} = \alpha + \beta f_t + \rho_t + \varepsilon_{t+1}$
i.e. $e_{t+1} = \rho_t + \varepsilon_{t+1}$

Thus ρ_t is an omitted variable from the model.

Since s_t and f_t are cointegrated, ρ_t must be a stationary variable.

Hence $E(b_{OLS}) = \beta + \frac{cov(\rho_t, f_t - s_t)}{var(\Delta s_{t+1})}$

i.e. the bias is due to $cov(\rho_t, f_t - s_t) < 0$

The risk premium and the forward premium must be negatively correlated.

We noted before that if investors are risk averse then they would require a higher return on the foreign investment to compensate for the FOREX risk

i.e. the risk premium and the foreign interest rate should be positively correlated

Hence the risk premium should be negatively correlated with $f_t - s_t = i_t - i_t^*$, which is what we seem to be observing.

The evidence is therefore consistent with an omitted risk premium.

2. FOREX market anomalies.

According to the theory previously discussed, anomalies could arise from the formation of expectations.

This does not necessarily imply that expectations are not rational.

(i) Peso effect

The basic idea is that investors attach a non-zero probability to a regime switch, but that switch does not occur.

Investors expectations are therefore based on a weighted average of two regimes: the old which in fact persists, and the new which doesn't occur.

The data only reflect what actually happened - which is no regime switch.

So it is not obvious after the event why expectations seem to have been wrong. It looks like irrationality but it isn't.

Thus

Actual expectations $= E[s_{t+1}|Old\ regime] \times prob(Old\ regime) + E[s_{t+1}|New\ regime] \times prob(New\ regime)$

As the old regime didn't change $ex-post$ (i.e. the econometrician) would use $E[s_{t+1}|Old\ regime]$ which would be wrong.

Thus actual expectations $= E[s_{t+1}|Old\ regime] + prob(New\ regime) \times \{E[s_{t+1}|New\ regime] - E[s_{t+1}|Old\ regime]\}$

The econometric model is only correct if $prob(New\ regime) = 0$, i.e. there is no peso effect.

In this in effect $\rho_t = prob(New\ regime) \times \{E[s_{t+1}|New\ regime] - E[s_{t+1}|Old\ regime]\}$ and not necessarily a risk premium.

(ii) Noise traders

In this case there is an expectational error as some FOREX traders are using the wrong model. As FOREX dealing is a specialised activity this explanation is unlikley to be as useful as it might be for other assets such as equity which are traded by a large number people.

Hence actual expectations $= E[s_{t+1}|\text{Correct model}] \times prob(\text{using correct model}) + E[s_{t+1}|\text{wrong model}] \times prob(\text{using wrong model})$

Suppose, that a proportion θ use the correct model (i.e. UIP) and the rest assume there will be no change in the exchange rate then.

actual expectation $= \theta(s_t + i_t - i_t^*) + (1-\theta)s_t = s_t + \theta(i_t - i_t^*) = s_t + i_t - i_t^* - (1-\theta)(i_t - i_t^*)$

Thus in effect $\rho_t = -(1-\theta)(i_t - i_t^*)$

(iii) Learning

Another example is where learning about the regime is taking place. This would be give rational expectations. Suppose for example that monetary policy followed a rule, but it is not clear what rule is being followed. Again investors would form expectations using a weighted average of different regimes or rules.

Evidence from Survey Expectations

In order avoid the problem of having to form expectations, some researchers have used the FOREX forecasts of professional market participants, called survey expectations data.

eg Mark and Wu, Economic Journal (1998).

Let s^e_{t+1} = the forecast or survey expectation of s_{t+1} made at time t.

Questions:
1. Are these expectations rational?
2. Are they better than UIP or no arbitrage theories at explaining s_{t+1}?

According to our earlier theory
$\Delta s_{t+1} = f_t - s_t + \rho_t + \varepsilon_{t+1}$ with $E_t[\varepsilon_{t+1}] = 0$

And hence $E_t[\Delta s_{t+1}] = f_t - s_t + \rho_t$

If the survey expectations are well founded then $E_t[s_{t+1}] = s^e_{t+1}$

and so $s^e_{t+1} - s_t = f_t - s_t + \rho_t$

Consider the following models:

1. $\Delta s_{t+1} = \alpha + \beta(f_t - s_t) + e_{t+1}$ $\quad H_0 : \alpha = 0, \beta = 1$

2. $s_{t+1} - s^e_{t+1} = \alpha_1 + \beta_1(f_t - s_t) + e_{1,t+1}$ $\quad H_0 : \alpha_1 = 0, \beta_1 = 0$

3. $s^e_{t+1} - s_t = \alpha_2 + \beta_2(f_t - s_t) + e_{2,t+1}$ $\quad H_0 : \alpha_2 = 0, \beta_2 = 1$

All of the errors should be serially uncorrelated.

Mark and Wu examine survey expectations data from various different sources: including MMS, AMEX and the Economist and for forecast horizons of 3, 6 and 12 months.

The main findings are:

1. The estimates of β and β_1 are with one exception negative and significantly so.
2. The estimates of β_2 are always positive.
3. Tests of $H_0 : \beta_2 = 1$ vary according to the survey data used and the forecast horizon. The shorter the horizon, the greater the t-statistic. For longer horizons the estimates of β_2 are close to unity.

The estimates of β_1 suggest that the survey expectations are not rational as the forward premium seems to be able to partly explain the expectations error. The greater forward premium, the greater under-estimation of s_{t+1} by the survey expectations.

The estimates of β_2 suggest that survey expectations are based on the forward premium, especially over longer horizons. Thus investors seem to be using UIP to form expectations. Nothing about the efficiency of the FOREX market can be learned from this equation, just about the survey expectations are formed.

The absence of a risk premium in the model could still be affecting the results.

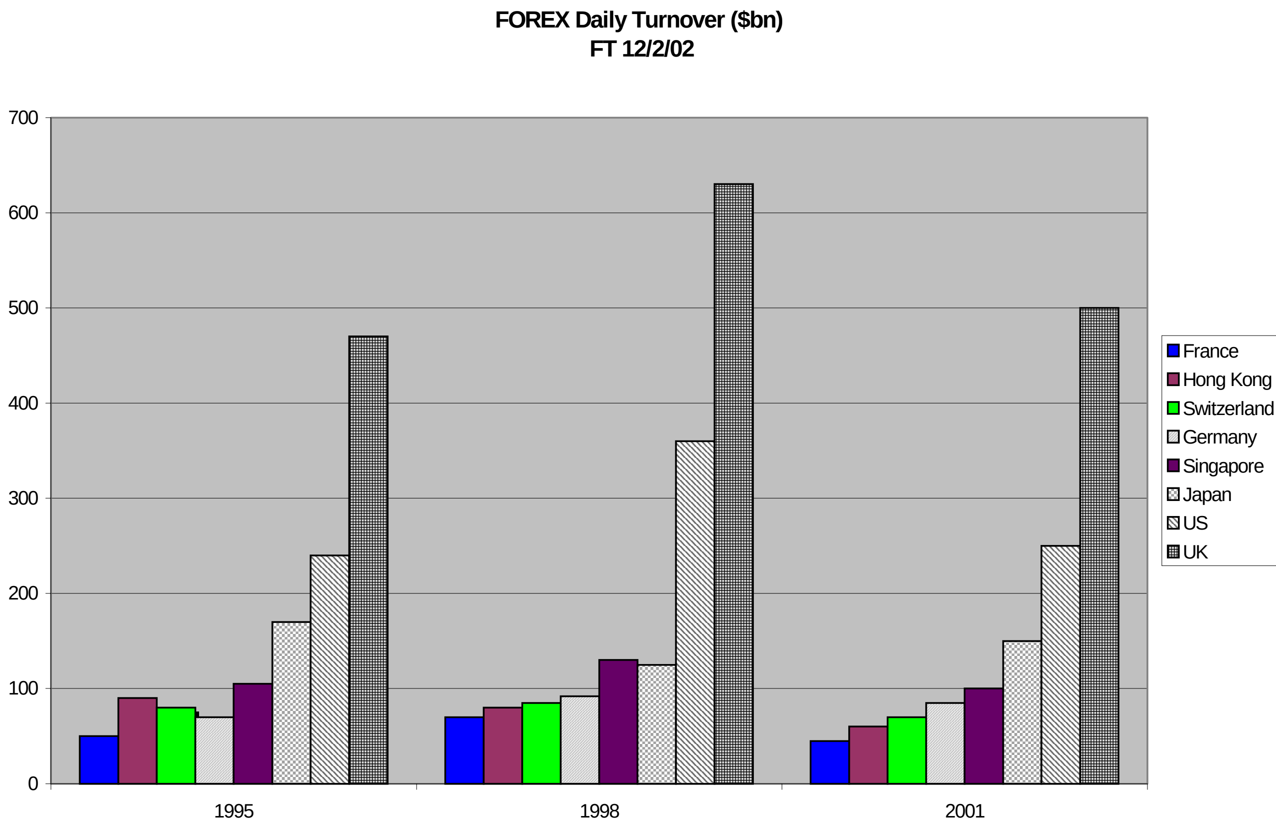
FOREX Daily Turnover ($bn)
FT 12/2/02
700
600
500
400
300
200
100
0
1995
1998
2001
France
Hong Kong
Switzerland
Germany
Singapore
Japan
US
UK

Figure 1:

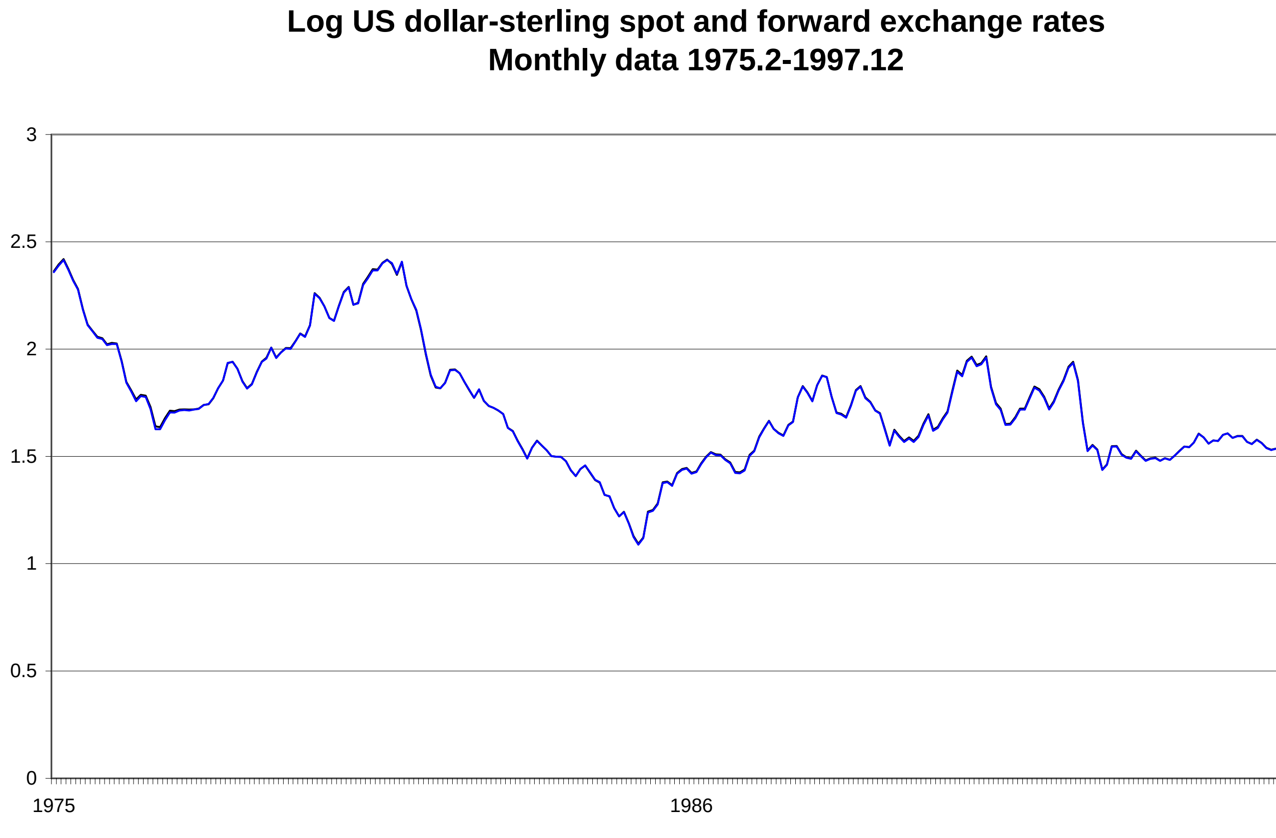
Log US dollar-sterling spot and forward exchange rates
Monthly data 1975.2-1997.12
3
2.5
2
1.5
1
0.5
0
1975
1986

Figure 2:

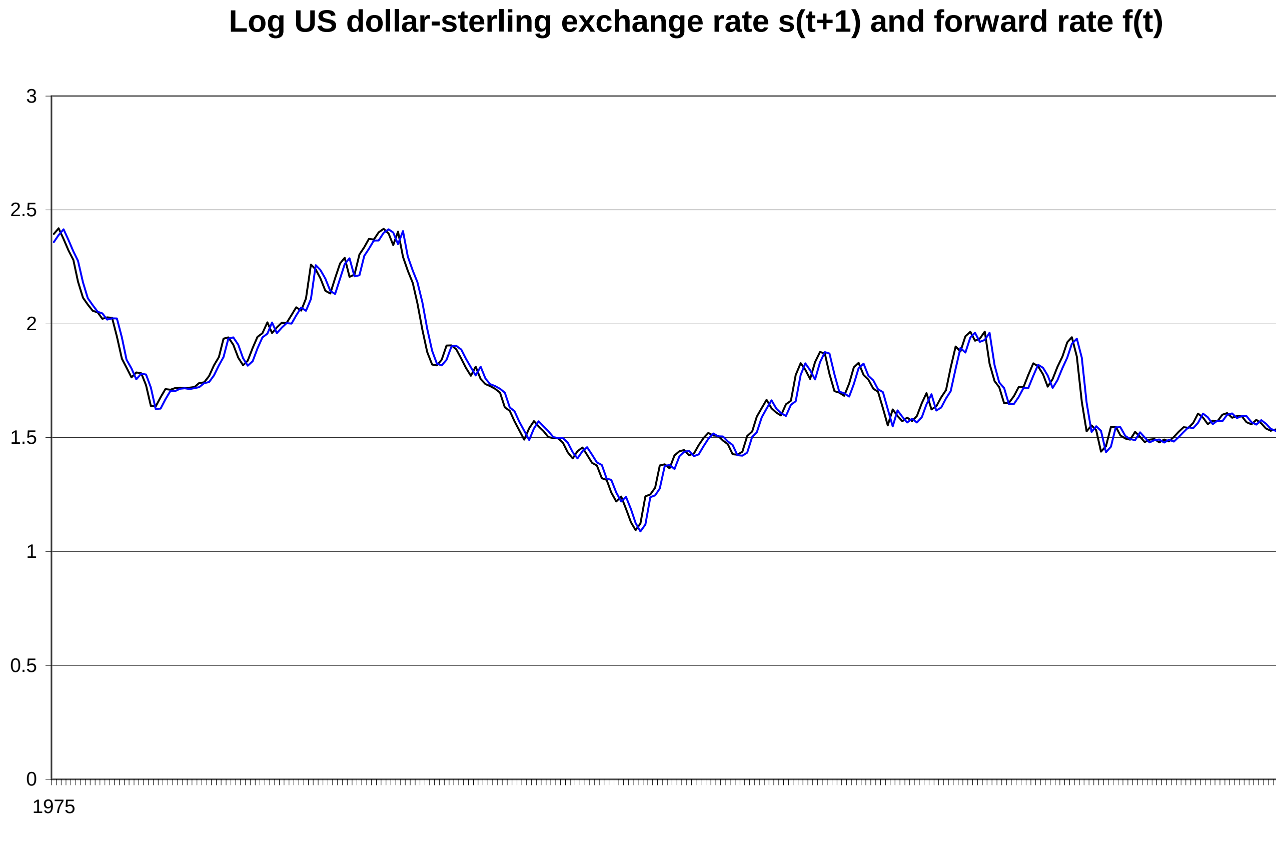
Log US dollar-sterling exchange rate s(t+1) and forward rate f(t)
3
2.5
2
1.5
1
0.5
0
1975

Figure 3:

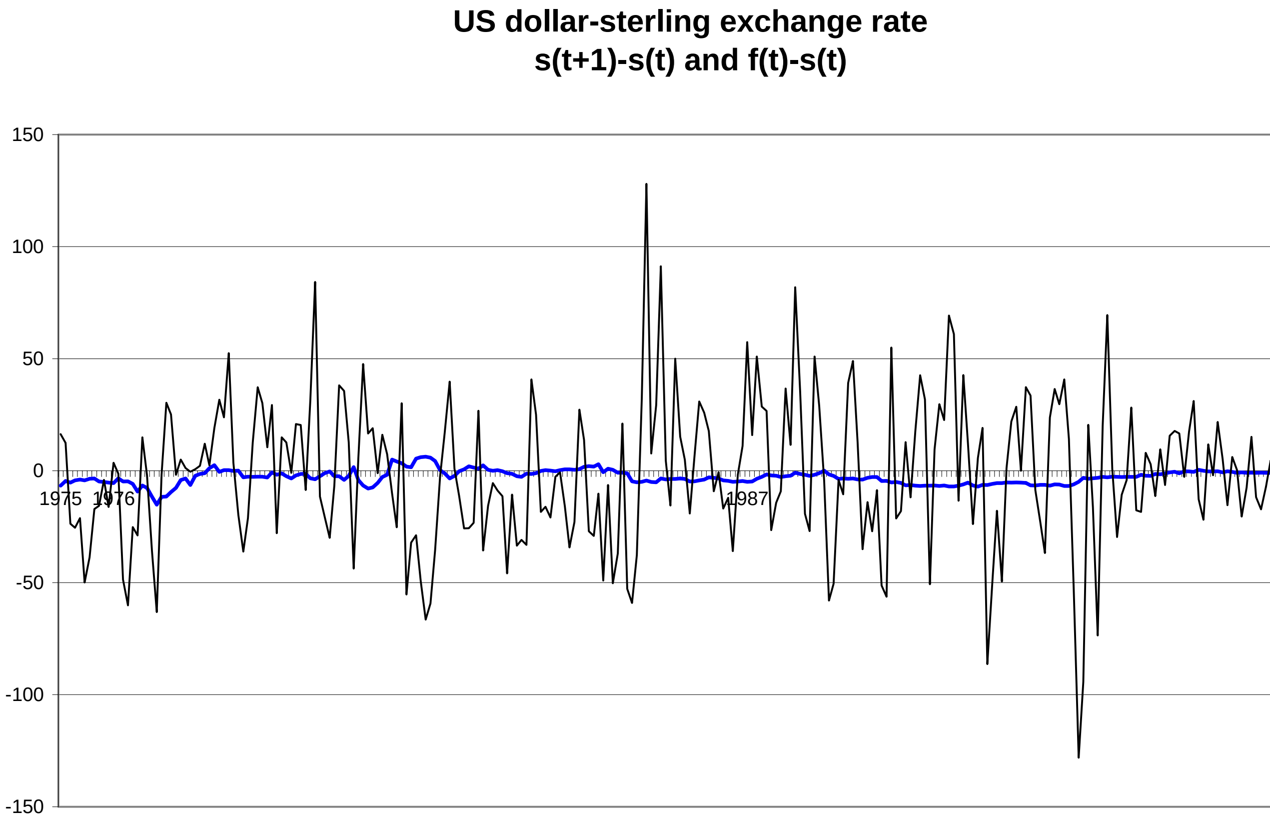
US dollar-sterling exchange rate
s(t+1)-s(t) and f(t)-s(t)
150
100
50
0
-50
-100
-150
1975
1976
1987

Figure 4:

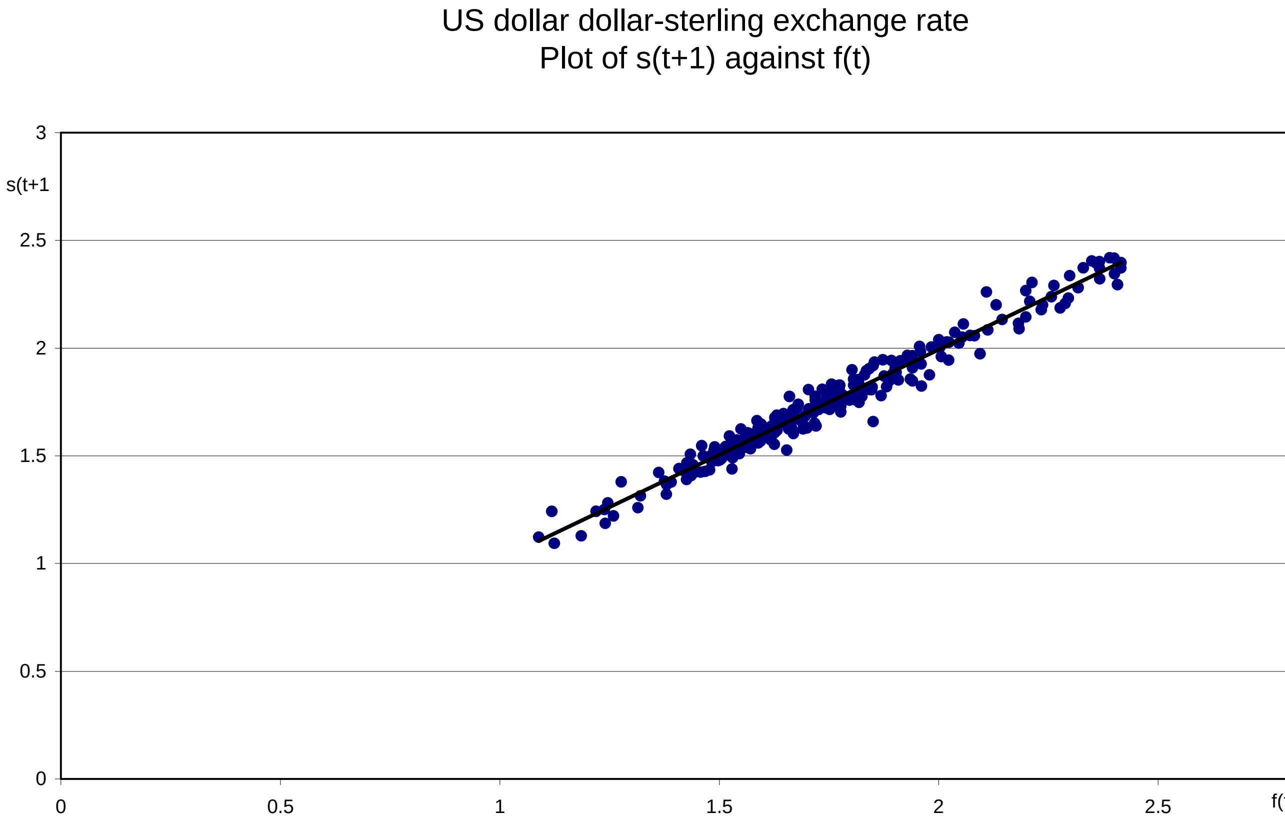
US dollar dollar-sterling exchange rate
Plot of s(t+1) against f(t)
s(t+1
3
2.5
2
1.5
1
0.5
0
0
0.5
1
1.5
2
2.5
f(

Figure 5:

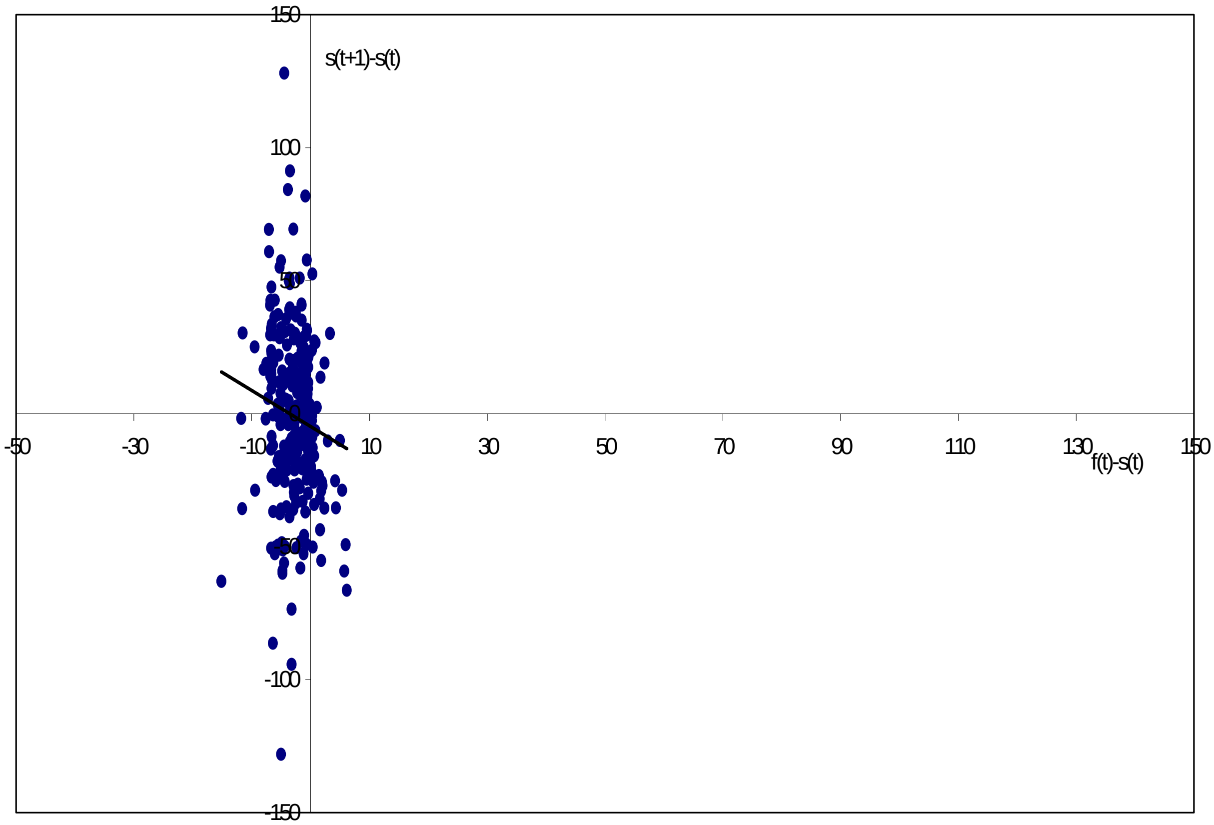
150
s(t+1)-s(t)
100
50
0
-50
-30
-10
10
30
50
70
90
110
130
150
f(t)-s(t)
-50
-100
-150

Figure 6: