Time varying forex market inefficiency

SERIE RESEARCH MEMORANDA

Time varying forex market inefficiency

Camiel de Koning
Stefan Straetmans

vrije **Universiteit**   *ams terdam*

Time varying forex market inefficiency *

Camiel de Koning
CMG Finance Amsterdam

Stefan Straetmans
Vrije Universiteit Amsterdam

Abstract

Researchers gathered abundant evidence on foreign exchange market inefficiency by regressing excess returns on lagged forward premia but they rarely investigated coefficient instability and its consequences for market efficiency testing. We allow for endogenous changes in the parameters when estimating by using rolling regressions and a Kalman Filter algorithm. Time variation in the regression coefficients is found to be statistically significant. If the regression parameters have changed over time, estimation methods that assume constant parameters may be inappropriate. We argue that the observed time variation in the forward premium slope is so large that a negative OLS slope for the post-Bretton Woods sample size is not improbable.

1

1 Introduction

In his renowned work on the rate of interest Fisher (1930, p. 39) discusses "the exact theoretical relation between the rates of interest measured in any two diverging standards of value and the rate of foreseen appreciation or depreciation of one of these standards relatively to the other . ..". And Fisher concludes that "the two rates of interest in the two diverging standards will, in a perfect adjustment, differ from each other by an amount equal to the rate of divergence between the two standards". Let s_t denote the log spot foreign exchange rate at time t. Under the hypothesis of rational expectations the 'foreseen depreciation' is equated with the expected depreciation rate $E_t [s_{t+1} - s_t]$. When applied to currencies, the interest differential can be replaced by the forward premium, and hence the Fisher hypothesis specializes to:

$$E_t[s_{t+1} - s_t] = f_t - s_t, \tag{1}$$

where f_t is the log forward foreign exchange rate at time t with maturity at $t + 1$. Let v_{t+1} be a mean zero innovation and consider the following equation:

$$s_{t+1} - s_t = f_t - s_t + v_{t+1}. \tag{2}$$

Within the rational expectations framework eq. (2) implies the forward market efficiency condition (1). Eq. (2) lends itself easily to a regression test. In an OLS regression of the realized spot return on a constant and the lagged forward premium (the 'forward premium' regression), the constant should be close to 0 and the slope β is expected to be close to 1. Doing just this for the dollar exchange rates the typical finding is a nonzero intercept and a slope coefficient that is significantly negative, often in the order -1 or -2. As a benchmark for the rest of the paper, Table 1 replicates this stylized fact for six currencies against the USdollar.

The original exchange rates are end-of-the-month nonoverlapping spot and forward middle rates vis-a-vis the £sterling. We calculate cross-currency US$ rates by exploiting the no triangular arbitrage condition. The series start in January 1976 and end in November 1995 except for the forward rate of the Japanese Yen which begins in June 1978.[1] One can see that the slope estimate is more often negative than positive (5 out of 6), and on average it is -1.223. When testing the null hypothesis that the slope equals 1, against the (two-sided) alternative

---

[1]Datastream also provides dollar rates but these are highly suspect. Using Datastream dollar rates we were unable to reproduce the negative slope estimates when testing the Fisher hypothesis. Presumably a too low degree of accuracy has been applied in calculating the dollar rates by the triangular arbitrage condition.

Table 1: OLS results (1976.01-1995.11)

$s_{t+1} - s_t = \alpha + \beta \left(f_t - s_t \right) + v_{t+1}$				
Currency	β	$s.e.(\beta)$	R^2	$D.W.$
DMark	-0.666	0.751	0.033	2.042
UK Pound	-1.849	0.799	0.022	1.854
Can. Dollar	-1.281	0.561	0.021	2.129
Fr. Franc	0.348	0.667	0.001	2.005
Jap. Yen	-2.563	0.899	0.038	1.998
Sw. Franc	-1.345	0.713	0.014	1.946
$\sum \beta_i / 6$	-1.223			

$(H_1: \beta \neq 1)$, the Fisher hypothesis can be rejected for all currencies except the French Franc at the 5 percent significance level.

The evidence for a less than unitary slope coefficient in Table 1 accords well with the abundant literature on the topic, see the surveys by Hodrick (1987), Lewis (1995) and Engel (1996). Nevertheless, financial markets seem to pay no attention to this result. We are not aware of any financial analyst using this result to beat the market. Perhaps this explains why the apparent downward bias continues to be investigated so heavily by the research community.

Two textbook econometrics explanations for the downward bias are the omitted variable bias and the failure of the innovation v_{t+1} to adhere to the standard OLS assumptions. Lewis (1995) uses this classification in her review of the premium puzzle. For long a risk premium has been the candidate omitted variable. Fama's (1984) seminal study showed that if the slope estimates are below $1/2$, this implies a risk premium which is more volatile than the variance of the spot returns. Later work showed that the implied coefficient of risk aversion is implausibly high. Moreover, identification of a time varying risk premium is not without its difficulties, see Nijman et al. (1993). Turning to the other explanation, there is some evidence from panel survey data that forecast errors are not in line with the rational expectations hypothesis, i.e., are correlated with lagged information (see e.g. Frankel and Froot (1989), Cavaglia et al. (1994)). Another possibility is the influence of infrequent policy shifts which are discounted by the public, but which are not properly accounted for by the regression residual. This failure to capture the 'peso phenomenon' is due to the very low frequency, possibly out of sample, nature of these events. As Lewis (1995) demonstrates such out of sample events can induce a downward bias, but the bias cannot be so large so as to render $\widehat{\beta}$ negative.

A third and rather novel way to interpret the estimation results in Table 1 is that the parameters α and β are time varying. Note that under rational expectations eq. (2) is not the only admissable stochastic specification consistent

3

with (1). Specifically, consider the following more general specification:

$$s_{t+1} - s_t = (1 + \varepsilon_{t+1})(f_t - s_t) + v_{t+1}, \qquad (3)$$

where both ε_{t+1} and v_{t+1} are conditionally zero mean innovations. It is easily seen that the Fisher hypothesis (1) follows from taking expectations with respect to time t information on both sides of (3).

Traditionally the additive innovation v_{t+1} is interpreted as forex news, see Frenkel (1981). In a similar vein, ε_{t+1} constitutes a multiplicative news factor. This factor expresses that there is no economic intuition as to why the ex post realized spot returns across different countries should be aligned along lines with a slope of 45 degrees with respect to the forward premium. It gives the direction and magnitude by which the interest differential is propagated through the forex spot market. The unconditional slope, however, should still be equal to one. But ex ante these disturbances are unknown and hence the forward premium is the only available indicator.

The purpose of this paper is to identify this slope variation and to investigate its relationship to the univariate results in Table 1. Both stability tests, rolling estimates and the Kalman Filter are employed to investigate the temporal behavior of ε_t. Anticipating on our results we show that the time variation is statistically significant across the different testing and modelling strategies. The Kalman Filter estimation procedure can also be used to take care of the fact that the intercept may be time varying due to the presence of a time varying risk premium. We show that the the time variation in the intercept term, although present, barely alters the slope dynamics (section 2). As regarding the consequences of the high slope variation we argue that it might induce a small sample bias in fixed parameter estimates such as OLS. This may partly explain the forward premium anomaly (section 3). We end with a summary and conclusions (section 4).

2 Slope variation as a stylized fact

2.1 Coefficient stability tests

Table 2 shows OLS slope estimates of the forward premium regression for subsamples of the original dataset used in Table 1.

While the case against unbiasedness is strong over the entire sample, it does not seem nearly as strong as it seems if one only were to look at specific subsample periods. For example, consider estimated slopes $\hat{\beta}_1$ over the sample period September 1977 to June 1990 in the first column of the Table.[2] For all currencies

---

[2]This sample period is interesting because it coincides with McCallum's (1994) sample period. This notorious study reports extremely negative point estimates for β for three major currencies against US$.

4

Table 2: OLS results for different sample splits

$$s_{t+1} - s_t = \alpha + \beta \left(f_t - s_t\right) + v_{t+1}$$

Currency	β_1	β_2	β_3	β_4	β_5
DMark	-4.372	-1.167	-0.896	-3.636	0.318
s.e.	1.670	2.234	0.830	1.573	1.165
s.e.	-4.201	0.481	-2.743	-2.349	-0.206
Can. Dollar.	-2.025 1.079	-0.302 1.413	-1.519 0.971	-1.331 0.895	-0.857 1.827
s.e.	0.671	1.356	0.615	0.726	1.033
s.e.	0.217	-0.42	0.457	-0.173	1.871
Jap. Yen.	-3.149 0.804	-6.300 1.496	-2.234 0.747	-3.239 0.779	-2.574 1.321
s.e.	1.187	6.085	0.971	1.226	1.775
Sw. Franc	-3.542	-3.523	-1.560	-4.037	-0.809
s.e.	1.166	2.249	0.786	1.271	1.426

$\widehat{\beta}_3$:1980:02-1995:11

$\widehat{\beta}_1$:1977:09-1990:06 $\quad$ $\widehat{\beta}_4$:1976:01-1987:01

$\widehat{\beta}_2$:1976:01-1980:01 $\quad$ $\widehat{\beta}_5$:1987:02-1995:11

we find even lower slope estimates than in the full sample case.

Table 2 also reports point estimates for nonoverlapping subsamples. Columns 2 and 3 contain slope estimates that arise from choosing January 1980 as a break-point. This date coincides with Bilson's (1981) sample, and his paper is often credited with the first published estimates of a statistically significant, negative β. This point is also close to the October 1979 change in US monetary policy. Comparing $\widehat{\beta}_2$ and $\widehat{\beta}_3$, the apparent deviation from market efficiency has increased in the 1980s. Columns 4 and 5 report slope estimates for the time periods before and after the Louvre accord (January 1987). Engel (1996) noted that such a split is reasonable because the long swings in the USdollar that appeared before the Louvre accord appear to have disappeared since. Presumably, the apparent deviation from market efficiency has again diminished in the post-Louvre period $\left(\widehat{\beta}_4 < \widehat{\beta}_5\right)$.

A myriad of test statistics has been applied to subsample regression results such as in Table 2 in order to test for temporal 'stability' or 'structural breaks' in the forward premium slope, see e.g. Barnhart and Szakmary (1991) and Bekaert and Hodrick (1993). The null hypothesis of equal slopes has been rejected time and time again. Although these results are supportive for our starting point, i.e., significant time variation, we think that these coefficient stability test results are not very informative, the main reason being that the economic interpretation of the chosen breakpoints is often unclear. One usually applies some kind of statistical search method (e.g. the Goldfeldt-Quandt likelihood ratio) in order

5

to determine the datapoint that makes a rejection of the null of equal slopes most likely. These breakpoints, however, most often differ from breakpoints exogenously chosen on the basis of international monetary history. Moreover, economically interpretable breakpoints do most often *not* lead to a rejection of parameter stability. As an illustration, we apply the popular Chow breakpoint test statistic to the subsample regression results in
Table 2.

$$H_0^{(1)} : \beta_2 = \beta_3 \text{ (Breakpoint: January 1980).}$$
$$H_0^{(2)} : \beta_4 = \beta_5 \text{ (Breakpoint: January 1987).}$$
$$H_0^{(3)} : \beta_6 = \beta_7 \text{ (Breakpoint : September 1985).}$$

The breakpoint of the third null refers to the so-called Plaza accord at which the G-7 countries decided to undertake a coordinated effort to reduce the value of the USdollar in world markets. [3] The Chow test's alternative hypothesis is two-sided and P-levels are reported between parentheses, see Table 3.

Table 3: F-statistics and corresponding P-levels for the Chow breakpoint test

Currency			
DMark	1.273	2.466	3.193
	(0.282)	(0.087)	(0.043)
UK Pound	1.665	0.593	0.472
	(0.191)	(0.584)	(0.624)
Can. Dollar	0.811	0.439	0.419
	(0.445)	(0.645)	(0.658)
Fr. Franc	0.264	1.467	3.517
	(0.768)	(0.233)	(0.010)
Jap. Yen	0.446	0.572	0.056
	(0.640)	(0.566)	(0.945)
Sw. Franc	2.391	3.330	1.395
	(0.094)	(0.037)	(0.249)

The Table shows that we are only able to reject $H_0^{(2)}$ for the Swiss Franc and $H_0^{(3)}$ for the German Mark and the French Franc at the 5 percent significance level. Moreover, neither of the null hypotheses can be rejected at the one percent significance level.

Another problem with parameter stability tests is that most of them are de-rived and applied in a regression framework assuming normally distributed inno-vations. However, when applying the Jarque-Bera normality test on the residuals

---

[3] Point **estimates for** β_6 **and** β_7 **have not been reported in Table 2.**

of the forward premium regression we were able to reject normality for five out of six currencies. Nonnormality in exchange rate returns and the resulting 'outliers' in the regression residuals may induce poor small sample properties of both point estimates of β and corresponding stability tests. Robust estimation techniques cum stability tests may provide a solution, see e.g. Kramer and Schotman (1994) or Lucas (1996).

Finally, note that the coefficient stability tests' alternative hypothesis, i.e., $H_1 : \beta_1 \neq \beta_2$ is rather restrictive. From a statistical point of view, the alternative hypothesis' vagueness results in a low power of the test statistic. From an economic point of view, a once-and-for-all jump in the forward premium slope would imply permanent deviations from forward market equilibrium. We would rather prefer to assume a continuously changing coefficient value in the short run that eventually may return to some economically meaningful value in the long run (e.g. implying market efficiency). We therefore turn our attention to time series techniques that explicitly model the time variation.

2.2 Rolling regressions

Apart from statistical testing for coefficient stability, researchers also tried to model coefficient time variation using sequential estimation methods such as rolling regressions, see e.g. Chiang (1988) or Barnhart and Szakmary (1994). This procedure consists of running a regression over $(T - k)$ subsamples with length k. The regression coefficients are estimated by shifting the sample period ahead one month at the time, i.e., by adding the most recent observation and at the same time deleting the most distant one. In this manner, the estimated coefficients reflect the impact of new information on the markets. If the coefficients display significant time variation when the subsample is 'rolled over', this is a strong indication of instability. Coefficient plots sometimes show dramatic jumps as the postulated equation tries to absorb a structural break. Chiang investigates the 'level' analog of the forward premium regression, i.e., he regresses the spot rate on the lagged one-month forward rate:

$$s_{t+1} = \alpha_{t+1} + \beta_{t+1} f_t + v_{t+1}. \tag{4}$$

This testing regression is equivalent to eq. (2) under the null hypothesis of market efficiency $(H_0 : \alpha = 0; \beta = 1)$. However, Chiang overlooks the fact that spot and forward rates are nonstationary time series. This implies that standard errors and corresponding t-ratios do not behave as in the standard linear regression model which renders Chiang's testing results rather suspect.

In order to provide a valid comparison to Barnhart and Szakmary (1991), we run the forward premium regression in a rolling way by OLS over successive 48-month subperiods. The estimation begins with the 1976:01-1980:01 subperiod and ends with the 1991.12-1995.11 subperiod. Figure 1 plots the rolling slope

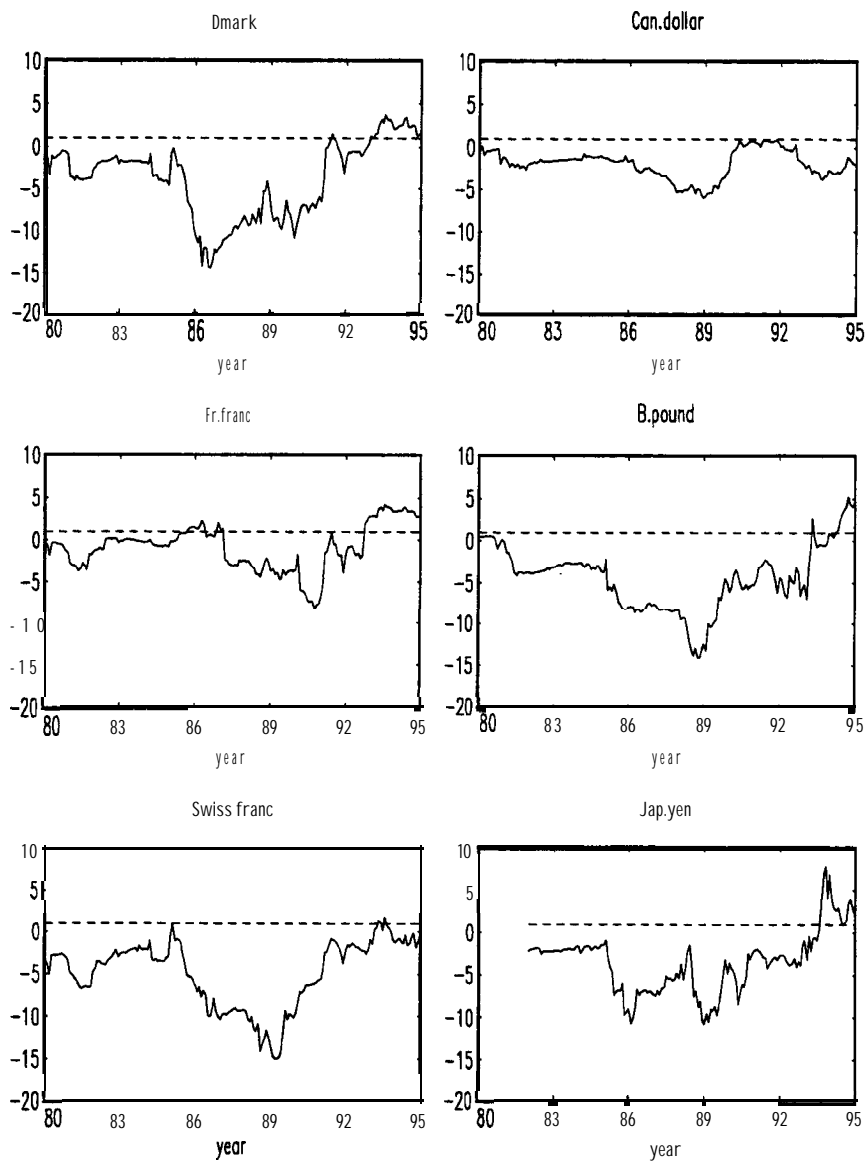Figure 1: Rolling USdollar slopes (1976:01-1995:11)

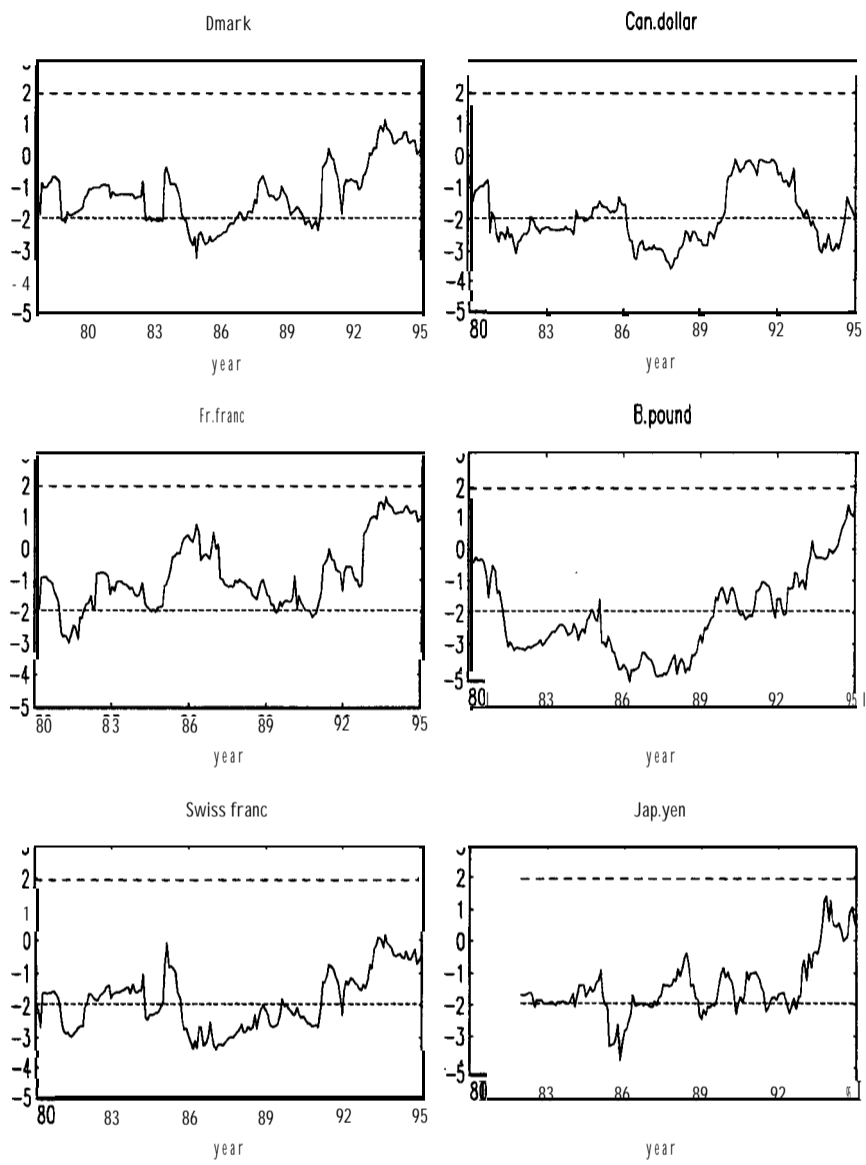Figure 2: Rolling USdollar t-ratios (1976:01-1995:11)

9

Table 4: Rejection percentages using rolling regressions

Currency	$k = 25$	$k = 50$	$k = 100$
DMark	15.42	23.38	59.0
UK Pound	28.97	56.08	73.38
Can. Dollar	26.17	57.14	79.14
Fr. Franc	6.54	12.70	1.44
Jap. Yen	14.05	30.62	53.64
Sw. Franc	20.09	45.51	66.19

parameters for the six currencies whereas Figure 2 shows the corresponding t-values under $H_0 : \beta = 1$. This null is represented by the dotted lines in Figure 1. The pairs of dotted lines in Figure 2 represent the critical values under the normal distribution for testing against $H_1 : \beta \neq 1$ at the 5 percent significance level.

First and foremost, one can see that the variability of the slope estimate is undeniable for all six currencies. Although the pattern is somewhat different for each currency, the bulk of the evidence shows that for all six currencies, β_{t+1} has a V-shape, i.e., it first drops and then begins to rise again towards the end of the sample. Note the relatively low degree of time variation for the Canadian-US$ premium slope. This should not come as a surprise given the high degree of monetary integration between both countries. Secondly, and conform to Barnhart and Szakmary, note the downward jump that all the β_{t+1} coefficients take around the beginning of 1985. This period is just prior to the Plaza accord of September 1985, at which the G-7 countries decided to coordinate their efforts to reduce the value of the USdollar in world markets. During 1986, however; the USdollar depreciated so quickly that the G-7 countries decided that a stabilization of the dollar became necessary (the Louvre accord of February 1987). This might explain the upward swing in the second half of the sample period.

Note that the t-ratios in Figure 2 may be interpreted as measuring the deviation from market efficiency over time. Clearly, the forward discount bias puzzle reappears in a rolling regression framework but less striking than in the fixed coefficient case. Note that t-values frequently lie in between the 95 percent confidence band indicating that the forex market may be efficient, at least temporarily. Moreover, efficiency seems to increase towards the end of the sample period. Table 4 reports statistical rejection rates for six currencies and for different window lengths $(k = 25, 50, 100)$.

To be more precise, the numbers denote the percentages of all rolling slope estimates for which the 5 percent confidence region contains the null. Clearly, all rates are bigger than 5 percent and increasing with the window size except for the French Franc. Hence the rolling regressions do not lead to an overall conclusion that is different from the fixed coefficient case.

2.3 State Space Models

The problem with sequential estimation methods such as rolling regressions is that they produce slowly varying estimates due to the high degree of sample overlap. Fitting a simple $AR(1)$ model to the estimated slope coefficients reveals serial correlations close to unity that are increasing with the window size. This falsely suggests that shocks to the slope may have a permanent character. Indeed, if markets are close to efficiency one rather expects that deviations from market efficiency are temporary and that economic forces like arbitrage drive the market back to some 'long run' equilibrium. The choice of the window size k also constitutes a factor of arbitrariness. The biggest problem with rolling regressions, however, is that they do not allow for an explicit a priori specification of the coefficients' stochastic process.

Time varying coefficient models that do nest a priori specifications of the coefficients' stochastic process are said to be dynamic. They represent a generalization of models in which the coefficients are *random,* i.e., independent over time. See e.g. Judge et al. (1985) for a discussion of this latter class of models. Two classes of dynamic coefficient models can be basically distinguished (see e.g. Harvey and Phillips (1982) and Harvey (1989)). In the first class, the coefficients are generated by stochastic processes which are nonstationary. An important example of this kind of behavior is when the parameters follow random walks (Cooley and Prescott (1976)). In the second class of models coefficients are generated by *stationary* stochastic processes around a fixed, but unknown, mean. Because the regression coefficients move around fixed means one can use the term 'return to normality' in order to describe the model. Schaefer (1975) found evidence for this kind of parameter variation in modelling a share's market risk; see also Rosenberg (1973). We used the latter type of state space model in order to describe the dynamics in the forward premium slope. Using the shorthand notations $x_t \equiv f_{t-1} - s_{t-1}$ and $y_t \equiv s_t - s_{t-1}$, the state space version of eq. (2) boils down to:

$$Y_t = \alpha + \beta_t x_t + v_t, \tag{5}$$

and

$$\beta_{t+1} - \mu_\beta = \rho_\beta \left(\beta_t - \mu_\beta \right) + \varepsilon_{t+1}. \tag{6}$$

We keep the intercept term constant because we want to concentrate on the temporal behavior of the forward premium slope. We will relax the constancy of the intercept in subsection 2.5. The first equation, the so-called 'measurement' equation in state space terminology, stands for a time varying version of eq. (2). T he other equation specifies a driving process for the 'state variable' β_{t+1}. The parameters of the model, ρ_β, σ_ε^2, σ_v^2, μ_β and α are assumed to be time

11

invariant .[4] The innovations v_{t+1} and ε_{t+1} are assumed to be mutually and serially uncorrelated. The parameter μ_β may be interpreted as the value to which β_{t+1} reverts in the mean.. There may be a *priori* economic reasons for wishing to regard the slope variation as being stationary. If $\mu_\beta = 1$, the exchange rate may be thought of as deviating from a situation of long run market equilibrium. Arbitrage probably ensures that these deviations from market efficiency are mean reverting. Moreover, taking into account the massive daily speculative flows in forex, it seems unlikely that 'short-run' market inefficiencies may persist for a long time. Thus we expect a relatively low value for μ_β.

Under these assumptions the system of equations (5)-(6) forms a state space model that can be recursively estimated by means of a Kalman filter. For some general discussions on the use of the Kalman filter see e.g. Kalman (1960), Harvey (1989) or Hamilton (1994).

The Kalman filter is essentially an algorithm that allows us to compute the mean and variance of the state variable $b_t = \beta_t - \mu_\beta$ on a period-by-period basis. We assume that the disturbance terms ε and v are normally distributed and that b_t has a normal prior distribution with mean $b_{0|0}$ and variance $V_{0|0}$. At every point in time, we want to update our prior distribution of the unknown state variable b_t by using the history of the processes $X_t \equiv \{x_\tau\}_{\tau=1}^t$ and $Y_t \equiv \{y_\tau\}_{\tau=1}^t$. The Kalman filter enables us, conditional upon knowledge of $b_{0|0}$, $V_{0|0}$, σ_ε^2, σ_v^2, ρ_β, μ_β and α to compute recursively the mean and variance of b for each point in time. Denote the conditional distribution of b_t given X_t and Y_t by $f(b_t | X_t, Y_t)$. Given the normality assumptions above, $f(b_t | X_t, Y_t)$ and $f(b_{t+1} | X_{t+1}, Y_{t+1})$ are also normal and thus completely characterized by their first two moments. If we represent the conditional mean and the conditional variance by $b_{t|t}$ and $V_{t|t}$, respectively, and those of $f(b_{t+1} | X_t, Y_t)$ by $b_{t+1|t}$ and $V_{t+1|t}$, then the Kalman filter recursions for $t = 1, \ldots . T$, where T denotes the sample size, are given by equations (7) through (11):

$$b_{t+1|t} = \rho_\beta b_{t|t} , \tag{7}$$

$$y_{t+1|t} = \alpha + b_{t+1|t} x_t, \tag{8}$$

$$V_{t+1|t} = \rho_\beta^2 V_{t|t} + \sigma_\varepsilon^2, \tag{9}$$

$$V_{t+1|t+1} = V_{t+1|t} - x_{t+1}^2 V_{t+1|t}^2 \left(x_{t+1}^2 V_{t+1|t} + \sigma_v^2 \right)^{-1} , \tag{10}$$

$$b_{t+1|t+1} = b_{t+1|t} + \left[y_t - (\mu_\beta + b_{t+1|t}) x_t \right] \left[x_t V_{t+1|t} (x_t^2 V_{t+1|t} + \sigma_v^2)^{-1} \right]. \tag{11}$$

---

[4]We distinguish between coefficients and *parameters* in the model. While coefficients are randomly fluctuating, parameters are fixed.

In order to start the Kalman filter recursions, the starting values $b_{0|0}$ and $V_{0|0}$ need to be specified. Since the process for b_t is given by eq. (6) such starting values are automatically available in the unconditional mean and the unconditional variance of b:

$$b_{0|0} = 0 \text{ and } V_{0|0} = \frac{\sigma_\varepsilon^2}{1 - \rho_\beta^2}.$$

Until now, we assumed that the recursive calculations of the Kalman filter take place for known values of the model parameters. In practice, however, this is not the case. Maximum likelihood techniques are available to estimate the model parameters. The sample log likelihood of our state space model boils down to:

$$\ln L = -\frac{T}{2}\ln 2\pi - \frac{1}{2}\sum_{t=1}^{T}\ln\left(x_t^2 V_{t|t-1} + \sigma_v^2\right)$$
$$- \frac{1}{2}\sum_{t=1}^{T}\frac{\left(y_t - \alpha - \mu_\beta x_t - b_{t|t-1}x_t\right)^2}{x_t^2 V_{t|t-1} + \sigma_v^2}. \tag{12}$$

Maximization of (12) is started by making an initial guess as to the numerical values of the model parameters. We can now iterate on (7) through (11) once they are calibrated with these initial values. The resulting sequences $\left\{b_{t|t-1}\right\}_{t=1}^{T}$ and $\left\{V_{t|t-1}\right\}_{t=1}^{T}$ can then be substituted in eq. (12) to calculate the value of the log likelihood function that results from these initial parameter values. Numerical optimization is employed to make better guesses as to the values of the unknown parameters until the likelihood function is maximized.

2.4 State Space results with fixed intercept

Table 5 reports Maximum Likelihood estimates for the return to normality model *(5)-(6)*. The data set coincides with the previously used currencies. Intercept estimates are all insignificantly different from zero and are thus not reported. First, the slope dynamics seem to be purely random because neither of the serial correlations $\hat{\rho}_\beta$ significantly differs from zero. Moreover, the slope estimates $\hat{\beta}_t$ fluctuate around a long run mean that is still significantly negative.

One should not be surprised that the $\hat{\mu}$-values are close to the previously reported OLS results for the forward premium slope. Indeed, it can be easily shown (see e.g. Harvey and Phillips (1982)) that Full Maximum Likelihood estimation of μ within the Kalman filter framework boils down to Generalized Least Squares (GLS) estimation of the forward premium slope:

$$s_{t+1} - s_t = \alpha + \mu_\beta\left(f_t - s_t\right) + w_{t+1}, \tag{13}$$

13

Table 5: Constant intercept with time varying slope (1976:01-1995:11)

$$y_t = \alpha + \beta_t x_t + v_t$$

$$\beta_{t+1} - \mu_\beta = \rho_\beta (\beta_t - \mu_\beta) + \varepsilon_{t+1}$$

Currency	$\widehat{\sigma}_\varepsilon^2$	$\widehat{\sigma}_v^2 (\times 10^3)$	$\widehat{\rho}_\beta$	$\widehat{\mu}_\beta$	$LM\ (\sigma_\varepsilon^2 = 0)$
DMark	32.107	8.261	-0.006	-0.326	43.72
s.e.	12.983	0.119	0.195	0.756	
UK Pound	24.848	8.202	0.331	-1.544	3.961
s.e.	8.569	0.111	0.186	0.796	
Can. Dollar	0.133	0.152	0.475	-0.35	3.951
s.e.	0.186	0.012	0.511	1.313	
Fr. Franc	11.185	0.9396	0.181	-0.233	15.53
s.e.	6.744	0.098	0.433	0.922	
Jap. Yen	0.421	1.179	0.511	-0.609	5.587
s.e.	0.123	0.115	0.421	1.343	
Sw. Franc	12.639	1.197	-0.091	-1.113	23.557
s.e.	5.692	0.141	0.269	0.848	

with heteroskedastic disturbance term

$$w_{t+1} = \varepsilon_{t+1} (f_t - s_t) + v_{t+1}. \tag{14}$$

As concerns the distribution of exchange rate news over the multiplicative news factor ε and the additive new factor v, most of the exchange rate news seems to be captured by the multiplicative news term $(\widehat{\sigma}_\varepsilon^2 > \widehat{\sigma}_v^2)$.

We also performed an explicit test of the slope variation's statistical significance. Testing for significant coefficient time variation in state space models is nontrivial because the variance parameter lies at the boundary of the parameter space under the null $(H_0 : \sigma_\varepsilon^2 = 0)$. When the remaining 'regularity' conditions for asymptotic normality of the ML estimator are fulfilled, the LR test statistic has a mixture of χ^2 distributed variables as asymptotic distribution. However, this testing approach cannot be applied either in our case because some of the remaining regularity conditions are also violated by the return to normality model, see Harvey (1989, p. 237). The use of a likelihood-based test, however, may be circumvented by noting that the estimated return to normality model in Table 5 is statistically indistinguishable from a random coefficient model. Thus, a standard test for heteroscedasticity may be used for testing the significance of the slope variation. From eq. (14) it follows that the residual variance is time dependent due to the squared forward premium[5]:

$$\sigma_{w,t}^2 = \sigma_v^2 + \sigma_\varepsilon^2 (f_t - s_t)^2 \tag{15}$$

---

[5] For sake of convenience we assume that $cov(\varepsilon, v) = 0$.

14

Using the variance estimate $\hat{\sigma}_\varepsilon^2$ from Table 5, the null hypothesis of no time variation or homoscedasticity ($H_0 : \sigma_\varepsilon^2 = 0$) can be tested by the *LM* statistic

$$LM = \frac{\hat{\sigma}_\varepsilon^4 \sum_t (f_t - s_t)^4}{2\hat{\sigma}_w^2} \tag{16}$$

that has a χ^2 (1) distribution in large samples. We were able to reject the null hypothesis of zero time variation (no heteroskedasticity) at the 5 precent significance level for all currencies and for 4 out of six currencies at the 1 percent significance level (see the last column in Table 5).[6]

2.5 State space results with time varying intercept

In the preceding state space model we identified the slope variation by assuming the intercept term α to be constant. This implies that we may have omitted a rational risk premium from eq. (5) possibly biasing the long run slope estimate $\hat{\mu}_\beta$. The proper inclusion of a risk premium, if present, might also lead to a drop in the observed time variation. To check these presumptions, we augment the model (5)-(6) with an additive news term α_t that may contain a rational risk premium p_t :

$$Y_t = \alpha_t + \beta_t x_t \tag{17}$$

$$\alpha_{t+1} = p_{t+1} + v_{t+1} \tag{18}$$

$$p_{t+1} - \mu_p = \rho_p (p_t - \mu_p) + \zeta_{t+1} \tag{19}$$

$$\beta_{t+1} - \mu_\beta = \mathbf{Pa} (\beta_t - \mu_\beta) + \varepsilon_{t+1} \tag{20}$$

Eq. (18) expresses that the risk premium is 'buried' into the additive news term α_t by the spot rate's rational forecast error v_t. Assuming that the risk premium exhibits a certain degree of persistence ρ_p in eq. (19) is consistent with both theoretical and empirical risk premium models, see e.g. Domowitz and Hakkio (1986), Taylor (1988) or Fraser and Taylor (1990).

The use of state space models in order to identify foreign exchange risk is not new. For example, Wolff (1987), Cheung (1993) and Nijman et al. (1993) apply Kalman Filter algorithms to identify risk. Our approach encompasses these attempts because the mentioned authors restricted the state variable β_t to be equal to one at each point in time.

Estimation of the augmented system (17)-(20) proceeds in fairly the same way as in the fixed intercept case, the only difference being that we have two

---

[6]The critical 5 % and 1% values for the used LM test are equal to 3.84 and 6.63, respectively.

15

state variables to identify instead of one. Conditional upon knowledge of the parameters of the α_t process the question remains whether and how the risk premium parameters are identified.

Substituting (19) into (18) yields the following model for α_t :

$$(1 - \rho_p L)(\alpha_t - \mu_p) = \zeta_t + (1 - \rho_p L) v_t, \tag{21}$$

which is an ARMA(1 ,1) model[7] because the right side of model (21) can be expressed as an MA(1) process with innovations η_t yielding

$$(1 - \rho_p L)(\alpha_t - \mu_p) = (1 - \theta L) \eta_t. \tag{22}$$

A comparison of structural form eq. (21) and reduced form eq. (22) shows that the well-known order condition for identification is satisfied because both equations contain the same number of unknown parameters (4).

Identification of the risk premium's mean and persistence parameter is trivial because the risk premium has the same mean and serial correlation as the state variable α_t. The other risk premium parameters may be identified by exploiting the equality of the second moments of the right side of (21) and (22). The first-order autocovariance and variance are given by, respectively:

$$-\rho_p \sigma_v^2 = -\theta \sigma_\eta^2, \tag{23}$$

and

$$\sigma_\zeta^2 + (1 + \rho_p^2) \sigma_v^2 = (1 + \theta^2) \sigma_\eta^2. \tag{24}$$

The Table shows ML results for the 'full' time varying state space model (17)-(20). The risk premium mean μ_p is found to be very small and insignificantly different from zero and is thus not reported in the Table. The additive news term α_t is found to be insignificantly different from white noise: the serial correlation parameter ρ_p is insignificantly different form zero for all currencies and the MA parameter θ is only significant for the Canadian Dollar. Thus the empirical evidence for a persistent forex risk premium is quite weak in our model. Moreover, the insignificance of ρ_p and θ also implies that the risk premium volatility σ_ζ^2 and the rational forecast variance σ_v^2 are not properly identified because the use of identifying eqs. (23) and (24) requires that we fill in significant values of these parameters.

The time series properties of the slope are only affected to a minor extent by the specification of an $\mathrm{ARMA}(1,1)$ intercept term which should not surprise taking into account the fact that we were unable to find a risk premium buried . into additive news. The estimated long run mean $\widehat{\mu}_\beta$ slightly increases in the full

---

[7]The summation theorem for moving averages in Ansley, Spivey and Wrobleski (1977) states that the summation of two uncorrelated MA processes of orders a and b respectively has a MA(c) representation where c $\leq$ max (a, b) .

Table 6: Full time varying model (1976:01-1995:11)

$$Y_t = \alpha_t + \beta_t x_t$$
$$\alpha_{t+1} = p_{t+1} + v_{t+1}$$
$$p_{t+1} - \mu_p = (p_t - \mu_p) + \zeta_{t+1}$$
$$\beta_{t+1} - \mu_\beta = \rho_\beta (\beta_t - \mu_\beta) + \varepsilon_{t+1}$$

	(1)	(2)	(3)	(4)	(5)	(6)
A	-0.337	0.232	-0.048	-0.297	0.003	0.300
s.e.	0.524	0.579	0.062	0.731	(0.927)	0.467
$\widehat{\theta}$	0.199	-0.339	152.45	0.228	0.002	-0.167
s.e.	0.581	0.568	3.161	0.747	0.925	0.492
$\widehat{\mu}_\beta$	0.173	-1.200	-0.272	-0.077	-0.604	-2.065
s.e.	0.617	0.814	2.943	0.872	1.346	1.038
$\widehat{\sigma}^2_\varepsilon$	31.836	22.629	0.005	11.135	0.406	9.712
s.e.	12.755	8.506	0.017	6.751	0.568	5.286
$\widehat{\sigma}^2_\eta(\times 10^{-4})$	7.929	8.061	1.369	9.252	11.76	11.88
s.e.	1.161	1.113	0.135	0.997	1.116	1.353

(1) Dmark	(3) Cn. Dollar	(5) Jap. Yen
(2) UK Pound	(4) Fr. Franc	(6) Fr. Franc

time varying model compared to the fixed intercept case but stays negative and significantly smaller than one for most currencies. Concerning the distribution of exchange rate news the movements in the multiplicative news term still highly dominate the volatility of the additive news term $(\widehat{\sigma}^2_\varepsilon > \widehat{\sigma}^2_\eta)$.

3 Coefficient time variation and small sample bias

In this section we discuss in a qualitative manner how the time variation documented in the previous subsections may be related to the anomalous OLS results for the forward premium slope in Table 1.

If the parameters of a regression relationship have changed over time, estimation methods, such as OLS, that presuppose constant underlying parameters may be 'inappropriate' for moderate sample sizes, i.e., they might exhibit small sample bias. To address this issue, suppose that the true underlying model of forward market equilibrium is given by eq. (3). If the forward premium slope is estimated by OLS then the latter can be written as:

$$\beta - 1 = \frac{\sum_t \varepsilon_{t+1} (f_t - s_t)^2}{\sum_t (f_t - s_t)^2} + \frac{\sum_t v_{t+1} (f_t - s_t)}{\sum_t (f_t - s_t)^2}. \tag{25}$$

17

In a non time varying framework ε would be fixed and the first term immediately reduces to ε. But with ε_{t+1} variable the speed of convergence of both terms towards zero depends on the time series properties of the news components (ε and v) and the forward premium. From the empirical subsections we know that the multiplicative news term ε is much more volatile than its additive counterpart v. Thus the first term in eq. (25) might be dominated by a few outliers which makes it quite plausible that $\hat{\beta}$ taked on a value of -1 or \$2 for samples of the standard Bretton Woods size.

In order to sustain this claim we set up a small Monte Carlo experiment. For a more theoretical and thorough treatment of this 'small sample argument', see e.g. Schotman et al. (1996). The benchmark model we consider for simulation is the following:

$$s_{t+1} - s_t = (1 + \varepsilon_{t+1})(f_t - s_t) + v_{t+1}. \tag{26}$$

The right hand side variables are sampled from the following distributions:

$$
\begin{aligned}
v &\sim cN_1, \\
(f - s) &\sim \frac{aM_t}{\sqrt{(N_2^2 + N_3^2 + N_4^2)/3}}, \\
M_t &= \rho M_{t-1} + N_5, \\
\varepsilon &\sim bN_6/M_t,
\end{aligned}
\tag{27}
$$

where all N_i, $i = 1, ..., 6$ are independent standard normal distributed random variables, and a, b, c and ρ are scaling constants which one has to calibrate in order to mimic the relative orders of magnitude of the real data series. More precisely the simulated spot return and forward premium should obey the stylized fact of 'news dominance', i.e. the variance of the spot return is approximately 100 times greater than the variance of the forward premium, see de Vries (1994). In order to attain this 'signal-to-noise' ratio we calibrated the parameter vector as follows: (a, b, c, p) = $(1/1500; 10; 1/10; 0.7)$.

Besides their relative orders of magnitude, we also want to mimic the distributional characteristics of the variables in (26). The additive noise term is sampled from a standard normal distribution because we did not find that this variable is heavy tailed. In contrast we find that the forward premium is highly fat tailed. Therefore we sample this term from a Student (3) distribution constructed as a ratio of a standard normal a M_t and the square root of a $\chi^2(3)$ variable divided by 3. Forward premia are highly dependent: one typically finds 1st order auto-correlations between 0.6 and 0.9.[8] To reproduce this feature in the data M_t is drawn out of an AR(l) process with normally distributed innovations and where ρ = 0.7. Multiplicative news shocks are modelled as Cauchy-distributed innovations in order to create some large outliers ('big news') in the time varying

---

*See also McCallum (1994) for this feature.
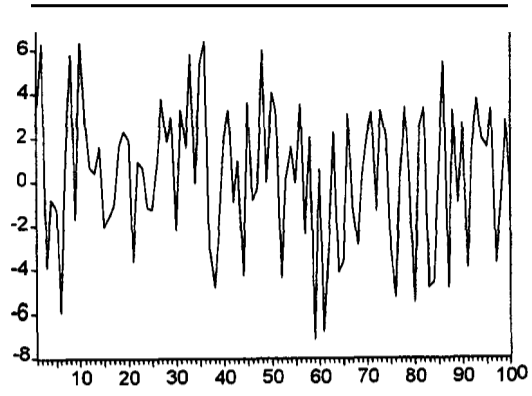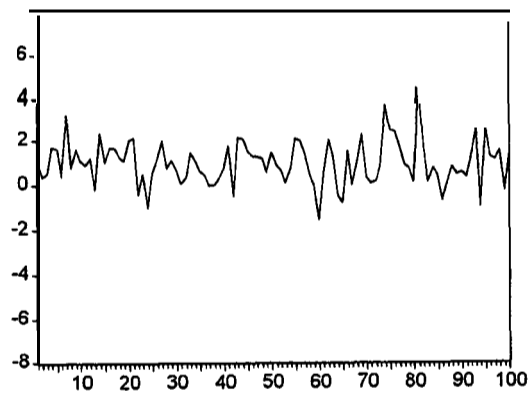
Figure 3: Simulated slopes for T=200

Figure 4: Simulated slopes for T=2000

slope coefficient. Note that we equalized the denominator of ε to the numerator of $(f - s)$ so that the product $\varepsilon (f - s)$ is again Student-(3) distributed. By a proper choice of a, c and ρ the additive noise term dominates $\varepsilon (f - s)$ as is also evident from the real data.

We perform the experiment for two different sample sizes in order to get an idea of the degree of convergence. Figures 3 and 4 report 100 simulations of the forward premium slope for $T=200$ and $T=2000$ respectively. For the smaller samples of size 200 the order of magnitude of the OLS slopes seems to correspond to the -1 or -2 values commonly reported in the literature on the forward discount bias. Also the OLS slopes converge in probability to their 'true' value but much slower than in the case where we would not have included a multiplicative news factor in the simulation equation. Indeed, we also performed simulations for the case where the multiplicative news factor is identically equal to zero. For comparable sample sizes, we find very small sampling variability in the slope estimates. In addition, the standard error of the OLS series seems to decrease more quickly under the 'no time variation' scenario indicating that OLS convergence is hampered if the underlying parameter is time varying.

4 Summary and conclusions

An 'academic industry' developed upon trying to rationalize the apparent rejection of the Fisher hypothesis for forward exchange, even though the markets never paid any attention to the apparent downward bias. Most of this literature has tried to explain the rejection by economic arguments such as irrational expectations, risk premia, peso problems or learning by speculators. Neither of these theories has been particularly succesful in dealing with the forward discount bias, specifically because it is difficult to nest different economic explanations for the bias in the same model and to test for these jointly.

In this article we took another route and allowed for both a multiplicative and an additive news component. We documented coefficient time variation in the forward premium regression for testing Uncovered Interest Parity. More specifically, we tried to assess the time variation result's sensitivity to changing the techniques for modelling and detecting time variation. Although commonly applied coefficient stability tests and rolling regressions have multiple weaknesses, the coefficient time variation patterns did not dramatically change when using more advanced time varying coefficient techniques like the return to normality model. We estimated this model by using the Kalman Filter in order to compare the model's likelihood for different time series processes of the Fama slope. It was found that β_t is most likely to be indistinguishable from white noise around some negative long run mean that is still significantly smaller than 1 in small samples.

We argued that the time varying slope varies so much that it is not improbable

that univariate time series estimates of β are well below 1 for the typical post Bretton Woods sample size. We are currently investigating the same issue for the term structure of interest rates which can be analyzed by the same methodology.

References

[1] Ansley, C.F., Spivey, W.A. and W. Wrobleski (1977), 'On the Structure of Moving Average Processes', Journal *of Econometrics,* 6, 121-134.

[2] Barnhart, S.W. and A.C. Szakmary (1991), 'Testing the unbiased forward rate hypothesis: Evidence on unit roots, co-integration, and stochastic coefficients', *Journal of Financial and Quantitative Analysis, 26, 245-267.*

[3] Bekaert, G. and R. J. Hodrick (1993), 'On Biases in the Measurement of Foreign Exchange risk premiums', *Journal of International Money and Finance,* 12, 115-138.

[4] Bilson, J.F.O. (1981), 'The speculative efficiency hypothesis', *Journal of Business, 54, 435-45 1.*

[5] Cavaglia, S.F.M.G., Verschoor, W.F.C. and C.C.P. Wolff (1994), 'On the biasedness of Forward Foreign Exchange rates: Irrationality or Risk Premia?', *Journal* of Business, 3, 321-343.

[6] Cheung, Y. (1993), 'Exchange rate risk premiums', *Journal of International Money and Finance, 12,* 182-194.

[7] Chiang, T.C. (1988), 'The forward rate as a predictor of the future spot rate: a stochastic coefficient approach', Journal of *Money,* Credit *and Banking, 20,* 212-232.

[8] Cooley, T.F. and E.C. Prescott (1976), 'Estimation in the presence of stochastic parameter variation', *Econometrica, 44,* 167-184.

[9] Domowitz, I. and C. Hakkio (1986), 'Conditional Variance and the Risk Premium in the Foreign exchange Market', *Journal of International Economics,* 19, 47-66.

[10] Engel, C. (1996), 'The For-ward Discount Anomaly and the risk premium: A Survey of Recent Evidence', *Journal of Empirical Finance, 3,* 123-192.

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

[12] Fisher, I. (1930), *The theory of interest, New* York (MacMillan).

[13] Frankel, J.A. and K.A. Froot (1989), 'Forward discount bias: Is it an exchange risk premium?', *Quarterly Journal of Economics, 54,* 139-161.

[14] Fraser, P. and M. Taylor (1990), 'Modelling isk in the interwar foreign exchange market', *Scottich Journal of Political Economy , 37,* 241-258.

[15] Frenkel, J.A. (1981), 'Flexible exchange rates, prices and the role of 'news': lessons from the 1970's', *Journal of Political Economy, 89, 665-705.*

[16] Hamilton, J.D. (1994), *Time Series Analysis,* Princeton University Press.

[17] Harvey, A.C. (1989), *Forecasting, structural* time series *models and the Kalman Filter,* Cambridge University Press.

[18] Harvey, A.C. and G.D.A. Phillips (1982), 'The estimation of regression models with time-varying parameters', in Deistler M., Fürst E., and G. Schwödiauer eds., *Games, economic dynamics, and time series analysis,* Wien-Würzburg (Physica-Verlag), 306- 321.

[19] Hodrick, R.J. (1987), *The empirical evidence on the efficiency of forward and futures foreign exchange markets,* Chur (Harwood publishers).

[20] Hodrick, R.J. and S. Srivastava (1986), 'The covariation of risk premiums and expected future spot exchange rates', *Journal of International Money and Finance,* 5, (Supplement), S5-22.

[21] Judge, G.G., Hill, R.C., Griffiths, W.E., Lutkepohl, H. and T.C. Lee (1985), *The theory and practice of econometrics,* New York (John Wiley).

[22] Kalman, R.E. (1960), 'A new approach to linear filtering and prediction problems', *Tmnsactions ASME Journal of Basic Engineering, 82, 35-45.*

[23] Kramer W. and P.C. Schotman (1992), 'Range vs. maximum in the OLS-based version of the CUSUM test', *Economics Letters, 40,* 379-381.

[24] Lewis, K.K. (1995), 'Puzzles in international financial markets', in G.M. Grossman and K. Rogoff eds., *Handbook of International Economics Vol. III,* Amsterdam (North-Holland), 1913-1971.

[25] Lucas, A. (1996), 'Outlier Robust Unit Root Analysis', Ph. D. thesis, Tinbergen Institute Research Series, Erasmus University Rotterdam.

[26] McCallum, B.T. (1994), 'A reconsideration of the Uncovered interest Parity Relationship', *Journal of Monetary Economics, 33,* 105-132.

[27] Nijman, T.E., F.C. Palm and C.P. Wolff (1993), 'Premia in forward foreign exchange as unobserved components: A note', Journal *of* Business *& Economic Statistics, 11, 361-365.*

22

[28] Rosenberg, B. (1973), 'Random Coefficient Models: The analysis of a Cross-Section of Time Series by Stochastically Convergent Parameter Regression', *Annals of Economic and Social Measurement, 2, 399-428.*

[29] Schaefer, S. (1975), 'Alternative Models of systematic Risk. International Capital Markets: an Inter and Intra country analysis, ed. by E. Elton and M. Gruber, Amsterdam, 150-161.

[30] Schotman, P., Straetmans S. and C.G. de Vries (1996), 'Big news in Small Samples', Tinbergen Institute Discussion Paper.

[31] Taylor, M. (1988), 'A DYMIMIC model of forward foreign exchange risk, with estimates for three major exchange rates', Manchester School of Economics and Social Studies, 56, 55-68

[32] Watson, M.W. and R.F. Engle (1983), 'Alternative algorithms for the estimation of dynamic factor, mimic and varying coefficient regression models', *Journal of Econometrics, 23, 385-400.*

[33] Wolff, C.C.P. (1987), 'Forward Foreign Exchange Rates, Expected Spot Eates, and Premia: a Signal- Extraction Approach', *Journal of Finance, 42, 395-406.*

[34] Vries, C.G. de (1994), 'Stylized facts of nominal exchange rate returns', in F. van der Ploeg, *The Handbook of International Macroeconomics,* Basil Blackwell, 348-389.