Big News in Small Samples *

Peter Schotman
University of Maastricht and L.I.F.E.

Stefan Straetmans
University of Maastricht and L.I.F.E

Casper G. de Vries
Erasmus University

Aug. 2001

Abstract

Univariate time series regressions of the forex return on the forward premium generate mostly negative slope coefficients. Simple and refined panel estimation techniques yield slope estimates that are much closer to unity. We explain the two apparently opposing results by allowing for both additive and multiplicative news. No arbitrage arguments imply that the multiplicative news component must be identical across all exchange rates at a given point in time. Cross section estimates reveal that the movements in the multiplicative news component are so large that a negative slope coefficient for the post Bretton Woods time series regressions is not improbable.

*We benefitted from presentations in Groningen, Aarhus, Rotterdam, London, Toulouse and Brussels. We also like to thank G. Jiang, N. Mark, L.P. Hansen and T. Nijman for their stimulating comments. *Address:* Tinbergen Institute, Erasmus University, P.O. Box 1738, 3000 DR Rotterdam, Netherlands, *e-mail:* CdEVRIES@FEW.EUR.NL

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 $E_t\left[s_{t+1} - s_t\right]$. 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 Fisher hypothesis (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 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 2.1 replicates this stylized fact for sixteen 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 December 1999[1]. One can see that the slope estimate is more often negative than positive, and on average it is -0.492. When testing the null hypothesis that the slope should equal 1, against the (two-sided) alternative (H_1: $\beta \neq 1$) the Fisher hypothesis can be rejected for 9 out of 16 currencies at the 5 percent significance level.

The evidence for a less than unitary slope coefficient in Table 2.1 accords well with the abundant literature on the topic, see the surveys by Hodrick (1987),

[1]Because Datastream did not provide data on Japanese 1-month forwards before July 1978 we extracted these data from the Harris Bank database. (Wie /waar hiervoor iemand bedanken?).

Table 1: OLS results (1976.02-1999.12)

$s_{t+1} - s_t = \alpha + \beta (f_t - s_t) + v_{t+1}$				
Currency	$\widehat{\beta}$	$s.e.(\widehat{\beta})$	R^2	$D.W.$
German Mark	-0.707	0.720	0.003	2.025
French Franc	0.017	0.571	0.000	2.013
UK Pound	-1.648	0.732	0.017	1.889
Spanish Peseta	0.774	0.304	0.022	1.768
Danish Krona	-0.810	0.499	0.009	2.024
Norwegian Krona	-0.560	0.491	0.005	2.033
Swedish Krona	1.075	0.454	0.019	1.831
Italian Lira	0.092	0.400	0.000	1.825
Japanese Yen	-2.235	0.798	0.027	1.982
Canadian Dollar	-1.086	0.480	0.018	2.152
Belgian Franc	-0.635	0.720	0.003	2.001
Dutch Guilder	-1.598	0.707	0.018	2.030
Portuguese Escudo	0.478	0.195	0.021	1.910
Swiss Franc	-1.364	0.684	0.014	1.934
Irish Punt	0.509	0.647	0.003	2.018
Austrian Shilling	-1.096	0.717	0.008	2.075
$\sum \widehat{\beta}_i/16$	-0.550			

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.

In fact, the empirical literature contains quite some evidence which puts the seemingly exact magnitude of the downward bias into doubt. Bilson (1981) already notices that the downward bias disappears if the most sizeable forward premia are dummied out. Under stationarity of the process the role of outliers is diminished in larger samples. Bilson's result also indicates that the relation between expected changes and the forward premium does not work at every point in time. Time variation in the slope coefficient β will be the main topic in this paper. Several regression techniques 'artificially' lengthen the sample size. For example, the Seemingly Unrelated Regression technique (SUR) is a particular weighing scheme that uses the information contained in all variables of the system. Fama (1984) reports the following results for a set of 9 dollar rates over the years 1973-1982: the average of the OLS slope estimates is -0.99, this increases to -0.68 upon application of unrestricted SUR, and restricted SUR (with regard to the slope) produces an estimate of -0.58. More recently panel methods have been employed. For a panel of 3 dollar rates with three and six month forward maturities Mayfield and Murphy (1992) report slopes of respectively 0.79 and

Table 2: Slope estimates by pooled regression analysis

Method	$\widehat{\beta}$	$s.e.(\widehat{\beta})$
Stacked GLS	0.622	0.048
Unrestricted SUR		
average estimate	0.064	
slope restricted		
SUR	0.259	0.053
slope and intercept		
restricted SUR	0.366	0.045

0.54. Flood and Rose (1994) on a stacked data set of 6 EMS currencies against the Deutsche Mark find a restricted OLS slope estimate of 0.58. Flood and Rose report that this result is robust against adding fixed time (year) effects. Very similar results are reported in Table 2.2 for our data set. It follows that the downward bias may be less severe than univariate OLS regressions do suggest.

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 (1990), 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.

In this chapter we take a third route which does not fall within the categories of omitted variable bias or systematic forecast errors. Note that under rational expectations eq.(2) is not the only admissable stochastic specification consistent with (1). Specifically, consider the following more general specification:

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

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. In a time series context the term $\varepsilon_{t+1}(f_t - s_t)$ induces heteroskedasticity in the errors. 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. But as we argue in the next section, there are strong economic arguments in favor of the hypothesis that the ex post returns do lie on straight lines with identical slopes, i.e., $\varepsilon_t^{ij} = \varepsilon_t$ for all currency pairs ij.

The purpose of this paper is to identify this slope variation. This is accomplished in a series of cross sectional regressions in a system of N different exchange rates. We then show that the time variation in this multiplicative news factor is almost dramatic. We find that the slope varies between -8.91 and 16.52, with a variance of 7.85 around the mean value of 0.44. It is argued that this high variability is a plausible explanation for the standard univariate time series result. In the estimation procedure we also try to take care of the fact that the intercepts may be time varying due to the presence of a time varying risk premium, or that this is due to time varying additive news. The time variation in the additive parts is also found to be considerable. But this variation is not the cause of the premium puzzle. It appears to be entirely caused by the sizeable variation in the multiplicative news factor.

2 Economic Theory and Econometric Techniques

One way to interpret the test results of the unbiasedness of the forward premium is that the parameters in the regression model

$$s_{t+1} - s_t = \alpha + \beta(f_t - s_t) + v_{t+1}, \tag{4}$$

are time varying. The suggestion is not new. Bekaert and Hodrick (1993), for example, conclude that "formal tests of the stability of the coefficients indicate that the parameters have changed over time". General time variation in the intercept and slope is tested in Barnhart and Szakmary (1991) and De Koning and Straetmans (1996). These studies use a rolling regression procedure to investigate the variability of the slope coefficient. They find evidence of the time variation of the slope, but, except for the latter half of the 1980s, the hypothesis $\beta = 1$ can often not be rejected. As noted in the introduction, time variation in the slope is not necessarily inconsistent with the Fisher hypothesis provided the slope variation is pure white noise.

The difficulty with this approach, and probably any univariate time series approach towards time varying parameters, is the lack of instruments. Which variables determine the time variation of the parameters? Rolling regressions and other nonparametric methods implicitly assume that the parameters vary slowly over time. In this section we make a different identifying assumption. Under the maintained hypothesis that the time varying slope coefficient is uncorrelated with the forward premium, the slope coefficients have to be the same for each bilateral exchange rate by a no arbitrage argument. A cross sectional regression of, say, multiple dollar spot returns against the respective forward premia will then provide an estimate of the slope for each time period.

Consider a multicurrency world of $N+1$ different currencies, numbered $j = 0, \ldots, N$. Let s_t^{ij} refer to the time t log currency price of currency j in terms of numeraire currency i. For brevity we introduce the shorthand notation $y_t^{ij} \equiv s_t^{ij} - s_{t-1}^{ij}$ and $x_t^{ij} \equiv f_{t-1}^{ij} - s_{t-1}^{ij}$. In its most general form the multicurrency version of model (4) reads

$$y_t^0 = \mathbf{B}_t^0 x_t^0 + \psi_t^0, \tag{5}$$

where $y_t^0 = (y_t^{01}, \ldots, y_t^{0N})'$ is a vector of length N, and where the vectors x_t^0 and ψ_t^0 are defined analogously. The coefficient matrix $\mathbf{B}_t^0$ is of order $(N \times N)$ and is fixed at each instant of time, but can be time varying. The vector of intercepts ψ_t^0 comprises both the stochastic and the nonstochastic additive terms of the model. Model (5) is written with currency 0 as the numeraire. When $\mathbf{B}_t^0$ is a full matrix, the equation for y_t^{0i} contains all N forward premia x_t^{0j} $(j = 1, \ldots, N)$ as explanatory variables, and not just its own forward premium x_t^{0i}. Although the model is written with currency 0 as the numeraire, the choice of numeraire is arbitrary.

Definition 1 *Let $\mathcal{M}_i$ be the system of equations of the form (5) describing all exchange rates against the common numeraire i with coefficient matrices ψ_t^i and $\mathbf{B}_t^i$. Further define $\tilde{\mathcal{M}}_{ij}$ as the system of equations for exchange rates against numeraire i that is implied by $\mathcal{M}_j$. A system of exchange rate equations is numeraire invariant if $\mathcal{M}_i = \tilde{\mathcal{M}}_{ij}$ for all i and j.*

Proposition 1 states the effects of a change of numeraire. Numeraire invariance requires for example that both the induced systems $\tilde{\mathcal{M}}_{ij}$ and $\mathcal{M}_i$ exhibit identical restrictions on the parameters in the coefficient matrix.

Proposition 1 *Model (5) is numeraire invariant with parameters for numeraire currency 1 if*

$$\psi_t^1 = \mathbf{P}\psi_t^0 \tag{6}$$
$$\mathbf{B}_t^1 = \mathbf{P}\mathbf{B}_t^0\mathbf{P}, \tag{7}$$

with

$$\mathbf{P} = \begin{pmatrix} -1 & \mathbf{0}' \\ -\iota & \mathbf{I} \end{pmatrix},$$

where ι is a vector of ones.

Proof. The matrix $\mathbf{P}$ specifies the transformation from currency 0 to currency 1. By direct multiplication it can be verified that $\mathbf{P}^{-1} = \mathbf{P}$, so that $\mathbf{P}$ is unipotent.[2] Applying $\mathbf{P}$ to the system (5) yields

$$\begin{aligned} y_t^1 &= \mathbf{P}y_t^0 = \left(\mathbf{P}\mathbf{B}_t^0\mathbf{P}^{-1}\right)\mathbf{P}x_t^0 + \mathbf{P}\psi_t^0 \\ &\equiv \mathbf{B}_t^1 x_t^1 + \psi_t^1. \end{aligned} \tag{8}$$

Since by reordering the variables in y_t^0 we can have any currency in position 1, the numeraire invariance holds with respect to any currency i. ■

The coefficients in (5) are $(N \times N)$ matrices that depend on the time subscript t but are fixed at each instant in time. Without restrictions the parameters are not identified. One set of identifying assumptions is that the $\mathbf{B}_t^0$ are constant over time. Then (5) is a system of N linear time series regressions. Moreover, since each regression equation contains the same set of regressors x_t^0, the parameters can be estimated by Ordinary Least Squares. A serious drawback of this specification is the large number of parameters. The number of parameters increases with the cross sectional sample size. For instance, it would not be possible to estimate (5) on a broad panel of countries, like in the PPP analysis of Frankel and Rose (1996). Another drawback is of course the assumption of constant parameters itself. Above we noted that there is quite some evidence in the literature suggesting that the parameters are time varying. Therefore we see as the main purpose of our analysis to investigate the time variation in the slope coefficient β in the univariate equations like (4). Moreover, preliminary univariate regression analysis whereby spot returns were regressed on the set of N premia did not appear to warrant further investigation.

To reduce the number of parameters, we like to go back to the univariate model in which y_t^{ij} only depends on x_t^{ij}. The reason for selecting this specific way of reducing the parameters is of course the Fisher hypothesis and the subsequent empirical literature that imposes the same restriction; it is also consistent with the theoretical economic literature on the subject (see below).

Assumption 1 *The coefficient matrix $\mathbf{B}_t^i$ is diagonal, with elements β_t^{ij}, $j = 0, ..., N$,*

$j \neq i$.

Under Assumption 1 the system (5) implies Proposition 2 which first appeared in Koedijk and Schotman (1990):

[2] Economically this means that applying the same transformation twice we go from currency 0 to currency 1 and back to currency 0 again. See Koedijk and Schotman (1989) for further discussion.

Proposition 2 (Koedijk and Schotman) *Given assumption 1, model (5) is numeraire invariant if and only if at any time* $\beta_t^{ij} = \beta_t$ *for all currency pairs* ij*, and* $\psi_t^{ij} = \psi_t^{kj} + \psi_t^{ik}$.

Proof. We can write the equation for an exchange rate return y_t^{ik} in two alternative ways. The first way is

$$\begin{aligned} y^{ik} &= y^{ij} - y^{kj} \\ &= \beta^{ij}x^{ij} - \beta^{kj}x^{kj} + \psi^{ij} - \psi^{kj} \\ &= \beta^{ij}x^{ik} + (\beta^{ij} - \beta^{kj})x^{kj} + \psi^{ij} - \psi^{kj}. \end{aligned}$$

And the second way is

$$y^{ik} = \beta^{ik}x^{ik} + \psi^{ik}.$$

Subtract the latter expression from the former to obtain

$$(\beta^{ij} - \beta^{ik})x^{ik} + (\beta^{ij} - \beta^{kj})x^{kj} + \psi^{ij} - \psi^{kj} - \psi^{ik} = 0. \tag{9}$$

Since (9) must hold for all $\left(x^{ik}, x^{kj}\right) \in \mathbf{R}^2$ we must have that the (time t fixed) coefficients satisfy $\beta^{ij} = \beta^{ik} = \beta^{kj}$ and the additive terms satisfy $\psi^{ij} = \psi^{kj} + \psi^{ik}$. Since, if not all coefficients β_t^{ij} would be equal, then $\psi^{ij} - \psi^{kj} - \psi^{ik}$ would be a function of some of the premia x^{ij} and thereby violating Assumption 1. The system would thus not be numeraire invariant. ■

An alternative short proof to the above proposition starts from a diagonal matrix $\mathbf{B}_t^i$ but with non-identical diagonal elements. The transformation $\mathbf{PB}_t^i x_t^i$ then yields equations containing diagonal and non-diagonal elements. Which implies that $\mathcal{M}_i$ is not numeraire invariant because $\mathbf{PB}_t^i\mathbf{P}$ is not diagonal. Below we focus on the time varying nature of the coefficients. Therefore it is useful to discuss an alternative approach that emphasizes the stochastic nature of the β (and ψ) as in the econometric varying coefficient models literature. Let $\mathbf{B}^i$ be a stochastic diagonal matrix. Suppose the conditional distributions of the β_t^{ij} and the ψ_t^{ij} do not depend on the forward premia x_{t-1}^{ij}. This assumption is the analogue of the assumption that $\mathbf{B}_t^i$ is a fixed coefficient matrix. The news interpretation of β_t^{ij} and ψ_t^{ij} may be motivated by considering a first order Taylor approximation of $y_t^{ij}(x_{t-1}^{ij}, .)$ at x_{t-1}^{ij} in the direction of x_{t-1}^{ij}. Then take conditional (time $t-1$) expectations of the variables in eq. (9). Because the moments are independent of the premia, and the two premia (out of the three) can be freely varied, the coefficients β_t^{ij} are equal in expectation $E_{t-1}\beta_t^{ij} = \beta$ and $E_{t-1}(\psi_t^{ij} - \psi_t^{kj} - \psi_t^{ik}) = 0$. Then square eq. (9) and again take expectations. This quadratic equation in x^{ik} and x^{kj} can only hold if all second moments are zero as well. This implies that $\beta_t^{ij} = \beta_t$ for all i, j. Note that this setup is more general than the Fisher model in (3) which imposes $E_{t-1}\beta_t^{ij} = 1$. In that sense the above model offers a well-defined alternative to the Fisher hypothesis and hence is amenable to statistical testing. To summarize, the foregoing literature on the Fisher hypothesis implicitly imposes the restriction that $\mathbf{B}_t^0$ is diagonal. What this literature did not recognize is that the no arbitrage relations then imply the equality of these coefficients. This restriction can be exploited in a panel study. To complete the specification we now split the additive news terms ψ_t^{ij} into a news component v_t^{ij} and a (time t fixed) intercept term α_t^{ij} :

$$\psi_t^{ij} = \alpha_t^{ij} + v_t^{ij}.$$

We make two further assumptions for identification purposes.

Assumption 2 (Orthogonality) *The news components have the conditional expectations* $E(v_t^{ij}|\ x_{t-1}^{ij}) = 0$.

In a time series context this is the standard orthogonality assumption. In the cross section it is made explicit. The other assumption distinguishes the intercept α_t^{ij} from the error term v_t^{ij}.

Assumption 3 (Factor structure) *The intercept has the factor structure*

$$\alpha_t^{ij} = \alpha_t^j - \alpha_t^i\ , \tag{10}$$

$$\alpha_t^i = \sum_{l=1}^{K} \delta_{il}\gamma_{lt}. \tag{11}$$

The second assumption is analogous to Mahieu and Schotman (1994). The first part of the assumption guarantees that the intercepts satisfy the triangular structure of exchange rates. The second part is the actual factor assumption.

A special case is the one factor assumption ($K = 1$) with all δ_j ($j = 0, \ldots, N$) equal to zero, except for the numeraire $\delta_i = 1$. Also setting $\beta_t = \beta$ we obtain the panel model with time effects

$$y_t^{ij} = \gamma_t + \beta x_{t-1}^{ij} + v_t^{ij} \qquad\qquad j = 0, \ldots, N,\ j \neq i \tag{12}$$

The linear specification (12) has been used in panel studies like Mayfield and Murphy (1992) and Flood and Rose (1994). The idea is that there are fixed effects due to time. In these studies the fixed effects are studied by considering a panel of exchange rates against a specific numeraire currency, say currency i. But a time effect γ_t in an exchange rate panel is only related to the numeraire currency. Changing to a different numeraire means subtracting two equations, and the time effect will drop out. For this reason we favor our more general and numeraire invariant specification.

In general a time effect only appears for those currency pairs for which the corresponding factor loadings δ_i and δ_j are different. In the empirical work we will therefore use the single factor model

$$\alpha_t^{ij} = (\delta_j - \delta_i)\gamma_t \tag{13}$$

In the special case that γ_t is constant over time the model reduces to a panel with individual effects. We will refer to (13) as a panel with multiplicative time and individual effects. The δ_j's in (13) are not identified in a single cross section, but only in a panel data framework under the identifying restriction that all δ_j are constant over time. Since the δ_j only appear as differentials, a normalization rule is necessary to fix their absolute value, for example $\delta_0 = 0$.

The multiplicative specification (13) of the intercepts is also economically motivated. The intercepts can be given an interpretation as risk premia. A straightforward exposition of risk premia in foreign exchange is available from Robichek and Eaker (1978) and Hodrick (1987, p.87). Let R_t^i and R_t^j denote the nominal returns in local currency on internationally traded comparable one-period bonds denominated in currency i and j respectively. The return on the foreign bond expressed in the currency i equals $R_t^j + s_{t+1}^{ij} - s_t^{ij}$. Suppose that the riskless rate of return in country i equals r_t. Finally let $E_t R_t^M$ be the world market risk premium in a one factor version of the International CAPM, see e.g. Dumas (1994) or Lewis (1995). The ICAPM then implies the two relations

$$\begin{aligned} E_t(R_t^j + s_{t+1}^{ij} - s_t^{ij}) &= r_t + \delta_j E_t R_t^M, & (14) \\ R_t^i &= r_t + \delta_i E_t R_t^M, & (15) \end{aligned}$$

where δ_ℓ ($\ell = i, j$) is the sensitivity of the return with respect to the return on the world market portfolio.[3]. Subtract eq.(15) from (14) to get

$$E_t(s_{t+1}^{ij} - s_t^{ij}) = (\delta_j - \delta_i) E_t R_t^M + R_t^i - R_t^j. \qquad (16)$$

From (16) we see that the expected rate of depreciation equals the interest differential plus a risk premium. The risk premium is the product of a common factor, which is the price of risk, and the relative amount of risk measured by the difference in the 'beta's'. The time varying part of the intercept in (16) could thus be interpreted as a risk premium, i.e., $\gamma_t = E_t R_t^M$. Ferson and Harvey (1991) provide evidence that this is a reasonable modelling assumption. Furthermore, for the purpose of identification some restriction on the intercept is needed anyhow. From a cross sectional perspective, the δ_i, $(i = 0, \ldots, N)$ are given.

The latent variable interpretation of the risk premium term $(\delta_j - \delta_i)\gamma_t$ would require a set of instruments when dealt within a time series model, see the eloquent exposition in Hansen and Hodrick (1983). The panel approach allows one to identify this term from the time 'dummies'.

Without imposing further structure on the error terms v_t^{ij} the model is still not identified. First, we make the standard assumptions that v_t^{ij} has mean zero and is orthogonal to all information variables dated t or earlier. From the time series

[3] That means δ_i is the covariance of the return R_t^i with the return on the world market portfolio divided by the variance of the world market return. We assume that the convexity term that arises from the covariance between the rates of inflation and depreciation is negligible.

regressions it is known that the error terms are conditionally heteroskedastic. We would therefore like to allow for time varying variances. Analogous to β_t, we would also like to refrain from making assumptions on the time series properties of the error variance. This can be done in a panel data framework at the price of making additional cross sectional assumptions. In particular, we assume the cross covariance structure proposed by Koedijk and Schotman (1990) and Mahieu and Schotman (1994). In these papers the error term is decomposed as

$$v_t^{ij} = v_t^i - v_t^j. \tag{17}$$

This decomposition expresses the additive news in exchange rate s^{ij} as the difference between news about currencies i and j. It is consistent with the triangular identity, and assures that the whole system is completely numeraire invariant, see Koedijk and Schotman (1990). We assume that the components are mutually uncorrelated and have a common variance $\frac{1}{2}\sigma_t^2$. These assumptions are formalized below.

Assumption 4 *The covariance matrix of the errors v_t^{ij} against the common numeraire currency i is given by*

$$\mathrm{E}(v_t v_t^{'}) = \sigma_t^2 \mathbf{S}^{-1}, \tag{18}$$

where v_t is the $(N \times 1)$ vector containing the error terms $v_t^j - v_t^i, (j = 1, ..., N, j \neq i)$ and where $\mathbf{S} = \frac{1}{2}\left(\mathbf{I} + \iota\iota^{'}\right)$ is an $(N \times N)$ matrix.

The decomposition (18) explicitly recognizes that the error terms contain the 'numeraire news' as a common factor v_t^i, introducing strong cross sectional correlations. In other words, it states that all exchange rates against the dollar will be correlated, because they share the dollar component. The decomposition also provides a justification of a time effect in a panel framework. From a cross sectional perspective there is no difference between the fixed effects model

$$y_t^{ij} = \gamma_t + \beta_t x_{t-1}^{ij} - v_t^{ij}, \ j = 1, \ldots, N, \ j \neq i \tag{19}$$

with cross sectionally uncorrelated errors v_t^{ij}, and the random effects model

$$y_t^{ij} = \beta_t x_{t-1}^{ij} + v_t^i - v_t^j, \ j = \ 1, \ldots, N. \tag{20}$$

The intercept γ_t in (19) is indistinguishable from the numeraire 'error term' ν_t^i in (20). Interpretation of our time varying intercept $(\delta_j - \delta_i)\gamma_t$ in (16) hinges on the assumption that the news components ν_t^j have equal variance, not depending on j. The estimates of γ_t can therefore not be unambiguously interpreted as a risk premium. In the empirical application we do not stress any risk premium interpretation, but focus on time variation of the slope coefficient β_t. It is the slope that constitutes the forward premium puzzle.

If the panel data model described above is a good description of the data, it reduces the information content of multiple univariate time series analyses. We emphasize that the parameters γ_t, β_t and σ_t^2 are common to all exchange rates, i.e., using several time series of exchange rates yields no more information concerning the Fisher hypothesis than using a single series. Hence, for the typical post 1973 samples used in the premium puzzle studies essentially only one single history is available concerning the multiplicative factor β_t.

Estimation of model (5) with covariance structure (18) proceeds through Quasi Maximum Likelihood. Assume that currency 0 is the numeraire, and write the model in vector notation as

$$y_t = X_t\theta_t + v_t \qquad\qquad t = 1, \dots, T \tag{21}$$

where

$$\begin{aligned} y_t &= (y_t^{01}, \dots, y_t^{0N})' \\ x_t &= (x_t^{01}, \dots, x_t^{0N})' \\ X_t &= (\delta \; x_t) \\ \delta &= (\delta_1 - \delta_0, \dots, \delta_N - \delta_0)' \\ \theta_t &= (\gamma_t, \beta_t)' \\ v_t &= (v_t^1 - v_t^0, \dots, v_t^N - v_t^0)' \end{aligned}$$

The log-likelihood function takes the form

$$\ln L = -\frac{1}{2}\sum_{t=1}^{T}\left(N \ln \sigma_t^2 + \frac{v_t'\mathbf{S}^{-1}v_t}{\sigma_t^2}\right), \tag{22}$$

The likelihood function can be concentrated with respect to the time effects β_t and γ_t, and the time dependent cross sectional variances σ_t^2. From the first order conditions with respect to θ_t and σ_t^2 we find the conditional estimators

$$\hat{\theta}_t = \left(X_t'\mathbf{S}^{-1}X_t\right)^{-1} X_t'\mathbf{S}^{-1}y_t \tag{23}$$

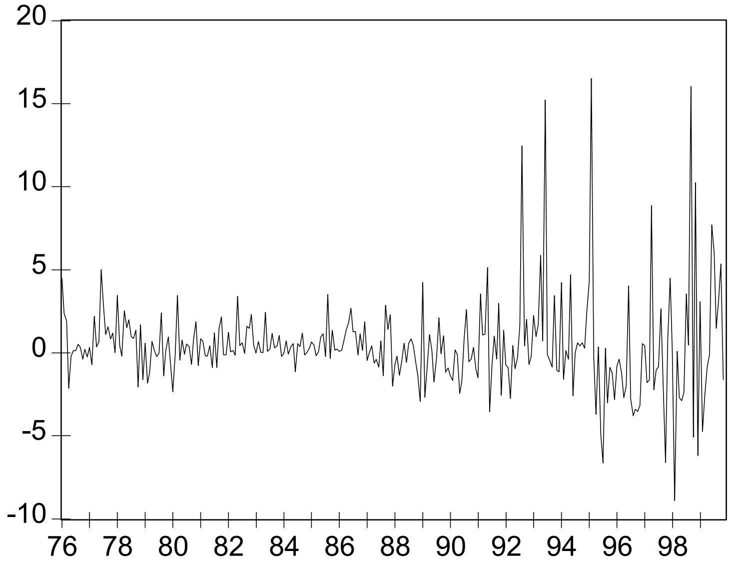
$$\hat{\sigma}_t^2 = \frac{1}{N}y_t'\left(\mathbf{S}^{-1} - \mathbf{S}^{-1}X_t(X_t'\mathbf{S}^{-1}X_t)^{-1}X_t'\mathbf{S}^{-1}\right)y_t \tag{24}$$

Substituting (24) in (22), the concentrated likelihood function becomes

$$\ln L^* = -\frac{N}{2}\sum_{t=1}^{T}\ln \hat{\sigma}_t^2 \tag{25}$$

This is still a function of δ, and must be maximized numerically. Robust standard errors for δ are obtained from the Hessian and Information matrix of the concentrated likelihood function. The information matrix can be computed from the outer product of the scores of the concentrated likelihood function (25).

Figure 1: Iterative GLS estimate of multiplicative news factor



The variance-covariance matrix for the time effects β_t and γ_t can be computed conditional on the estimated value of δ using the standard least squares formula

$$\mathbf{V}\left(\widehat{\theta}_t\right) = \hat{\sigma}_t^2 \left(X_t'^{-1}\mathbf{S}^{-1}X_t\right)^{-1}. \tag{26}$$

In this section we report and discuss estimates of the model

$$s_{t+1}^{i0} - s_t^{i0} = (\delta_i - \delta_0)\gamma_t + (f_t^{i0} - s_t^{i0})\beta_{t+1} + v_{t+1}^i - v_{t+1}^0, \tag{27}$$

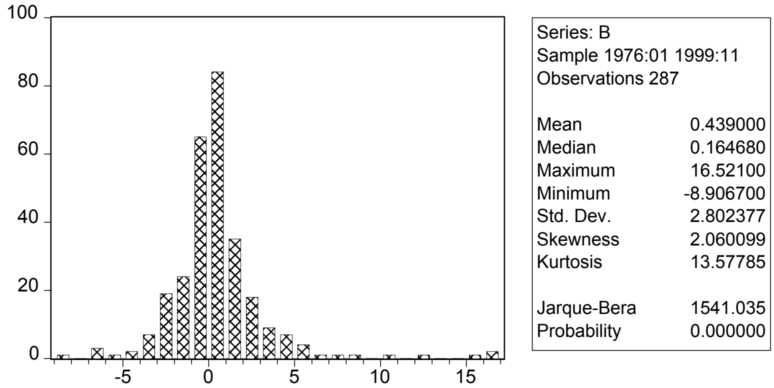
and variance-covariance matrix $\Omega_t = \sigma_t^2\mathbf{S}^{-1}$, for our dataset with 16 currencies and 210 months. The estimation procedure is the QML estimator outlined in (22) and (26). We first discuss the slope estimates.

The striking properties of the $\hat{\beta}_t$ estimates are easily visualized by the time plot in Figure 2.1.

First of all, the average of $\widehat{\beta}_t$ is 0.44, and is thus positive in contrast with much of the literature. Second, the sheer magnitude and variation in the estimates is remarkable in view of the presumed smoothness that is inherent to the traditional univariate time series approach. Although some of the volatility of the $\widehat{\beta}_t$ can be attributed to estimation error, the cross sectional standard errors $\mathbf{V}(\widehat{\beta}_t)$ are small enough to conclude that β_t is far from constant. To compute the true variance of β_t , as distinct from the estimated $\widehat{\beta}_t$, we use the decomposition

$$\widehat{\beta}_t = \beta_t + (\widehat{\beta}_t - \beta_t) \tag{28}$$

Figure 2: Distributional characteristics of multiplicative news factor



Since $\widehat{\beta}_t$ is the best linear estimator of β, the estimation error $(\widehat{\beta}_t - \beta_t)$ is uncorrelated with β_t. The time series variance follows as

$$var(\widehat{\beta}_t) = var(\beta_t) + E[var(\widehat{\beta}_t - \beta_t)], \tag{29}$$

from which we can estimate the variance of β_t as

$$var(\beta_t) = \frac{1}{T}\sum_{t=1}^{T}(\widehat{\beta}_t - \overline{\beta})^2 - \frac{1}{T}\sum_{t=1}^{T}\mathbf{V}\left(\widehat{\beta}_t\right) \tag{30}$$

The average cross sectional estimation variance of $\widehat{\beta}_t$ filters out the noise from the variance of $\widehat{\beta}_t$. For our dataset this results in

$$var(\beta_t) = 7.85 - 1.26 = 6.59 \tag{31}$$

Therefore most of the variation in β_t in Figure 2.1 is real, and not due to estimation noise. The unconditional distributional characteristics of the $\hat{\beta}_t$ are displayed in Figure 2.2 and normality is strongly rejected.

Plots in Appendix 2.B of the Hill statistic, see e.g. de Vries (1994), for the lower and upper tail of the distribution of $\hat{\beta}_t$ indicate that the number of bounded moments is in the neighborhood of 3. Hence the $\widehat{\beta}_t$ distribution is heavy tailed and may just have a finite variance. In the simulation study in the fourth section below this fact is investigated more thoroughly.

Before we continue we also investigate visually whether the $\hat{\beta}_t$ estimates are reliable. For example, one might worry that the sizeable movements in the $\widehat{\beta}_t$ are caused by a few outliers. Proposition 2 holds that the ex post slope coefficients β_t^{ij} should be identical for all i and j at a given time instant. Therefore one would like to see that the points $x_t^{ij} = (f_{t-1}^{ij} - s_{t-1}^{ij})$ and $y_t^{ij} = (s_t^{ij} - s_{t-1}^{ij})$ lie approximately on the same straight line, irrespective of the currency pair ij,

Table 3: Correlation matrix between multiplicative news factors for different cross sectional dimensions

	$\widehat{\beta}_{1,t}$	$\widehat{\beta}_{3,t}$	$\widehat{\beta}_{5,t}$
$\widehat{\beta}_{1,t}$	1		
$\widehat{\beta}_{3,t}$	0.96691	1	
$\widehat{\beta}_{5,t}$	0.95162	0.99330	1

when adjusted for the intercept effect $(\delta_j - \delta_i)\gamma_t$. For each month in the sample we therefore constructed scatter plots for all $\frac{1}{2}n(n-1) = 240$ currency pairs. For a given numeraire i, each of the spot returns y_t^{ij} and the forward premia x_t^{ij} $(j = 0, ..., N, j \neq i)$ were adjusted by the estimated intercept. All the intercept adjusted points were plotted, which in principle should locate a single straight line through the origin. Within the same figure 16 of such clouds are recorded by varying the numeraire i.

Some scatter plots are reported in Appendix 2.C. We chose those plots that correspond with the most salient episodes of recent international monetary history. The year 1982 (Figure 2.6) stands out because of the unexpected Belgian devaluation in February and the widely anticipated EMS realignment of June. The year 1992 (Figure 2.7) also produced some extreme events on the monetary scene when both the lira and the pound were forced to leave the EMS. The year after (Figure 2.8), these currency crashes apparently also contagioned the French Franc which left the band in July-August. Subsequently, European monetary authorities decided to widen the band for most EMS currencies. It is remarkable to see that whenever the slope is sizeable the scatter points are always very well aligned along the estimated line. The rotation in the line does not seem to be caused by the presence of outliers.

We also investigated whether the time variation result persists when omitting important currencies from the cross section. Table 2.3 represents correlations between 3 cross sectional $\widehat{\beta}_t$ estimates.

While $\widehat{\beta}_{1,t}$ represents the original slope series, $\widehat{\beta}_{3,t}$ is calculated over a smaller cross section excluding the United Kingdom, Germany and France. If one additionally excludes the Netherlands and Austria and implements the iterative GLS estimator one obtains $\widehat{\beta}_{5,t}$. From Table 2.3 it is clear that the specific time series pattern of the slope estimates is hardly altered by excluding the considered countries.

Returning to the $\hat{\beta}_t$ estimates as depicted in Figure 2.1, we also investigated the time series behavior of the slope estimates. We checked for the presence of autocorrelation in the $\hat{\beta}_{t+1}$ and the $\hat{\beta}_{t+1}^2$ (ARCH-behavior), but no evidence for such intertemporal dependencies was found. However, looking back at Figure 2.1 the magnitude of the slope estimates is increasing towards the end of the sample, suggesting increased turbulence in the forex market. This is, in a way, more apparent than real. Consider the current situation of the hard core EMS currencies. In anticipation of the monetary unification in 1999, and partly due to convergence in monetary policy, the (short-term) interest spreads have narrowed considerably (interest futures rates have neared each other up to a few basis points). Consider again as shorthand notations for the spot return and the forward premium $y_{t+1}^{ij} \equiv s_{t+1}^{ij} - s_t^{ij}$ and $x_t^{ij} = f_t^{ij} - s_t^{ij}$. In essence the panel estimate for the slope reflects the difference between the ratio y_{t+1}^{ij}/x_t^{ij} and the ratio $(\widehat{\gamma}_t \hat{\delta}_{ij} + \widehat{v}_{t+1}^{ij})/x_t^{ij}$. This interpretation of β has no meaning when $x_t^{ij} = 0$. But it may explain the increased magnitude of the $\hat{\beta}_{t+1}$'s toward the end of the sample. Because the magnitude of the denominator x has gone down over time. (Apart from the European Monetary Union effect, x has also diminished because of the decline in world inflation , i.e., nominal interest rates cannot be negative, and because world capital markets have become more open). In contrast the variation in the spot returns y has not changed much. The composite Figure 2.9 in Appendix 2.D shows cross section variance estimates $\hat{\sigma}_c^2(y_{t+1}^{ij})$ and $\hat{\sigma}_c^2(x_t^{ij})$. Moreover, it also plots the respective components of the following approximate identity

$$\widehat{\beta}_{t+1}^2 \approx \frac{\hat{\sigma}_c^2 \left(y_{t+1}^{ij} \right)}{\hat{\sigma}_c^2 \left(x_t^{ij} \right)} - \frac{\hat{\sigma}_c^2 \left(\widehat{\gamma}_t \hat{\delta}_{ij} + \widehat{v}_{t+1}^{ij} \right)}{\hat{\sigma}_c^2 \left(x_t^{ij} \right)}. \tag{32}$$

We omit the plot for $\widehat{Cov}_c(\hat{\beta}_{t+1} x_t^{ij}, \widehat{\gamma}_t \hat{\delta}_{ij} + \widehat{v}_{t+1}^{ij} \hat{\sigma}_c^2)(x_t^{ij})$ because it is essentially 0. It appears that the divergence in $\left| \hat{\beta}_{t+1} \right|$ over time is caused by convergence in the interest differentials x, while the return innovations have not levelled off.

Concerning the weak Fisher hypothesis that $E_t \beta_{t+1} = 1$, the descriptive statistics in Figure 2.2 tell about the same story as the other procedures for pooling the data that were discussed in the introduction. The mean and median are much closer to 1 than the univariate time series estimates. On basis of normality of the $\hat{\beta}_{t+1}$, however, a simple t-test would still reject the null. In the next section we try to argue, however, that the current sample size, given the nature of the data, is still too small to be able to confidently reject the Fisher hypothesis. Albeit the fact the panel estimation procedure already goes some way in 'extending the sample size'.

We now turn our attention to the intercept terms $(\delta_j - \delta_i)\gamma_t$. Figure 2.3 summarizes the distributional aspects of $\widehat{\gamma}_t$, whereas Table 2.4 reports the estimates of $(\delta_j - \delta_i)$ with respect to two numeraire choices j. Note that the scale of the intercept is not numeraire invariant.

Depending on the choice of the numeraire, more or less weight is attributed to $\hat{\sigma}_t^2$. From the descriptive statistics in Figure 2.4 we see that $\widehat{\gamma}_t$ behaves quite differently from $\widehat{\beta}_{t+1}$. Normality of $\widehat{\gamma}_t$ cannot be rejected, and the mean is indistinguishable from 0. Time series tests for autocorrelation or conditional heteroskedasticity show that $\widehat{\gamma}_t$ is indistinguishable from white noise.

Figure 3: Iterative GLS estimate of common intercept factor

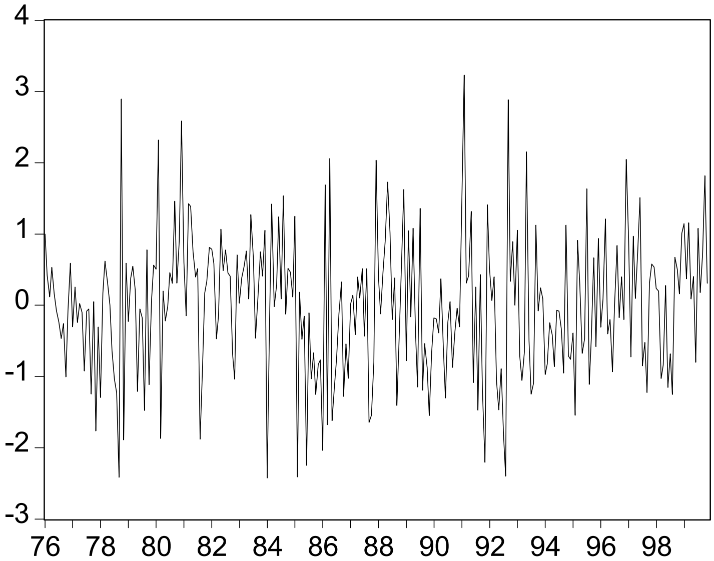


Figure 4: Time series properties of common intercept factor

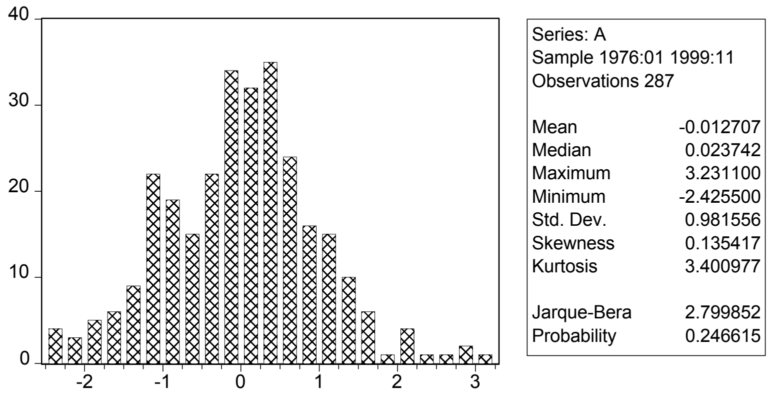


Table 4: Iterative GLS estimate of currency-specific intercept

Currency	$\widehat{\delta}_i$	$\widehat{\delta}_i - \widehat{\delta}_k$
German Mark	0.03134	-0.03134
French Franc	0.03068	-0.00066
UK Pound	0.02413	-0.00721
Spanish Peseta	0.02610	-0.00524
Danish Krona	0.03186	0.00052
Norwegian Krona	0.02596	-0.00539
Swedish Krona	0.02475	-0.00660
Italian Lira	0.02865	-0.00269
Japanese Yen	0.01577	-0.01558
Canadian Dollar	-0.00094	-0.03229
Belgian Franc	0.03123	-0.00011
Dutch Guilder	0.03144	0.00009
Portuguese Escudo	0.02804	-0.00330
Swiss Franc	0.03198	0.00063
Irish Punt	0.03018	-0.00117
Austrian Shilling	0.03149	0.00014

As Table 2.4 shows the country specific effects relative to the US$ are all significant except for the Canadian Dollar. This can be rationalized by the high degree of monetary integration between the two countries. If monetary integration is indeed reflected in the significance of $\hat{\delta}_{ij}$, it is also of interest to consider the significance of the country specific effect in a target zone such as the EMS. As expected, we could not detect a significant price of risk in any of the EMS cross rate returns (second column).

At this point it is of some interest to start discussing what causes the 'downward bias' in the univariate slope coefficients. One hypothesis which has been advanced by e.g. Mayfield and Murphy (1992) is that this is due to the omission of the time fixed effect. Certainly, if one pools the data and allows for such an effect than the slope estimate becomes 0.54. This constitutes a considerable increase over the average of the univariate time series estimates. But a similar increase occurs when one uses other methods for pooling data, cf. Table 2.2 above. Vice versa, if one restricts the time effect $\gamma_t = \gamma$, then the $\hat{\beta}_t$ estimates are not very much influenced by this (the correlation between the 'γ_t- restricted' and unrestricted $\hat{\beta}_t$ is 0.9). The reason is that the premia are small in magnitude compared to the returns, i.e., the so-called news dominance phenomenon (see Table 2.5 in Appendix 2.A). Consequently the intercept restriction is relatively harmless for the purpose of finding β_t.

Another celebrated cause for the downward bias in the univariate analyses is the omitted variable bias due to correlation between the forward premium and, say, a (time varying) risk premium. But it is questionable whether this would affect the univariate estimates relative to the pooled estimates very much. For this to be the case the effect would have to go into different directions for different exchange rates, because simple pooling of the data 'apparently' averages this effect out. But then more of the univariate estimates would have to be positive. At any rate, this type of omitted variable bias problem can still play a role in the discrepancy between the average of our panel estimates of $\hat{\beta}_t$ and the Fisher hypothesis. But it seems unlikely that a risk premium can explain the wide variation in the $\hat{\beta}_t$, due to the stability of the x_t over time. This can be seen as follows. The x_t exhibit high first order autocorrelation of about 0.7 (due to Central Bank money market management). Hence the x_t are relatively slowly moving when compared to $s_{t+1} - s_t$. Take a first order Taylor approximation of the risk premium with respect to x_t. In the previous section we argued that the presence of a risk premium which depends on x_t does not affect the conclusion that cross sectionally all slopes have to be identical, i.e., $\beta_t^{ij} = \beta_t$. Hence, a cross section regression picks up β_t plus the partial derivative of the risk premium with respect to x_t at the point of linearization. This causes a bias in the measurement of β_t one way or another. But the $\hat{\beta}_t$ display an enormous variation over time. This is unlikely due to wide variations in the partial derivative, since the x_t time series is so smooth.[4] The next section tries to argue that it is the large movements in the multiplicative factor which causes the difference between the univariate and pooled estimates.

3 Slow convergence

Consider again the univariate benchmark model

$$y_{t+1} = \beta_{t+1} x_t + \psi_{t+1}, \;\; t = 1, ..., T. \tag{33}$$

For simplicity we have again combined the intercept and the additive news term into ψ_{t+1}. This section focuses on the likelihood that in a single time series sample the slope estimate is negative when β_{t+1} is time varying with the distributional properties that were detected in the previous section. Note that the inquiry into $\Pr\left\{\widehat{\beta} < 0\right\}$ is somewhat different from the more usual question concerning the rate of convergence of $\widehat{\beta}$ to the unconditional mean of β_{t+1}, i.e., the standard $\sqrt{T}$ result. We are instead interested in a large deviation probability. Basically we like to ask how likely it is that a univariate time series sample of the standard post Bretton Woods size turns up a sizeable negative $\widehat{\beta}$. By the arbitrage arguments from the second section this implies that if the forex time series generate

[4] It could be caused by another omitted variable that both changes the risk premium all the time and is uncorrelated with x_t. But then, as the scatter plots show, this factor has to be common to all currencies (there is not much deviation from the slopes).

a negative $\widehat{\beta}$, it is then likely to occur for all forex series. Standard OLS implies $\widehat{\beta} = \sum\limits_{t}^{n} x_t y_{t+1} / \sum\limits_{t}^{n} x_t^2$. We investigate $\Pr\left\{\widehat{\beta} \leq -q\right\}$, when $q \to +\infty$. To study this probability, we make one distributional assumption. In line with the observed heavy tails of the (spot) return distribution, it is assumed that the innovation ψ is fat tailed distributed. Evidently, this assumption then applies to both tails. The concept of a fat tailed distribution is well defined in the literature. One says that $F(\psi)$ has fat tails if there exists an index $\alpha > 0$, and functions $L_1(y)$ and $L_2(y)$ such that

$$\lim_{t \to -\infty} \frac{F(t)}{(-t)^{-\alpha} L_1(t)} = 1 \text{ and } \lim_{t \to +\infty} \frac{1 - F(t)}{t^{-\alpha} L_2(t)} = 1, \tag{34}$$

whereby

$$\lim_{t \to -\infty} \frac{L_1(ts)}{L_1(t)} = 1 \text{ and } \lim_{t \to +\infty} \frac{L_2(ts)}{L_2(t)} = 1. \tag{35}$$

One says that $F(t)$ is a regularly varying function with tail index α, while $L_1(t)$ and $L_2(t)$ are slowly varying functions. When $F(t)$ is regularly varying, the tails of its density $f(t)$ decay by a power and hence only the moments less than α are bounded. This excludes e.g. the normal distribution as it displays exponential decay. A necessary and sufficient condition for the fat tail property is that the following limits hold, $b > 0$,

$$\lim_{t \to -\infty} \frac{F(bt)}{F(t)} = b^{-\alpha} \text{ and } \lim_{t \to +\infty} \frac{1 - F(bt)}{1 - F(t)} = b^{-\alpha}, \tag{36}$$

see e.g. Leadbetter et al. (1983). These conditions are easily verified for the Student-t distribution, but also apply to the ARCH(1) process with conditionally normally distributed innovations. The evidence is that these distributions/processes better describe the spot return process than the normal distribution. We first need the following result.

Lemma 1 *Assume that the X_t and Ψ_t are each i.i.d. random variables and that X_t is independent of Ψ_t. Suppose $F(\psi)$ satisfies condition (36), and that X with density $g(x)$ has more bounded moments then Ψ. Then the distribution of the convolution $\sum\limits_{t=1}^{T} X_t \Psi_t$ asymptotically equals*

$$\Pr\left\{\sum_{t=1}^{T} X_t \Psi_t \leq -q\right\} \approx T q^{-\alpha} L_1(-q)$$

as $q \to +\infty$.

Proof. Let $H(q) = \Pr\{X\Psi \leq -q\}$, so that

$$\begin{aligned} H(q) &= \int_{-\infty}^{0} g(x) \int_{-q/x}^{+\infty} f(\psi)d\psi dx + \int_{0}^{+\infty} g(x) \int_{-\infty}^{-q/x} f(\psi)d\psi dx \\ &= \int_{-\infty}^{0} g(x) \left[1 - F\left(-\frac{q}{x}\right)\right] dx + \int_{0}^{+\infty} g(x)F\left(-\frac{q}{x}\right) dx, \end{aligned}$$

We investigate $H(q)$ as $q \to +\infty$ and want to verify the left tail version of condition (36). To this end we investigate $A(s) = \lim_{t\to+\infty} H(st)/H(t)$. Because $F(.)$ is bounded between 0 and 1, it follows that the limit can be taken inside the integrals by the Lebesgue Convergence Theorem. Hence

$$A = \frac{\int_{-\infty}^{0} g(x) \lim_{t\to+\infty} (1 - F(-\frac{st}{x}))dx + \int_{0}^{+\infty} g(x) \lim_{t\to+\infty} F(-\frac{st}{x})dx}{\int_{-\infty}^{0} g(x) \lim_{t\to+\infty} (1 - F(-\frac{t}{x}))dx + \int_{0}^{+\infty} g(x) \lim_{t\to+\infty} F(-\frac{t}{x})dx}.$$

By the regular variation properties (34), (35), and further manipulation we get

$$\begin{aligned} A &= \frac{\int_{-\infty}^{0} g(x) \lim_{t\to+\infty} (-\frac{st}{x})^{-\alpha} L_2(-\frac{st}{x})dx + \int_{0}^{+\infty} g(x) \lim_{t\to-\infty} (-\frac{st}{x})^{-\alpha} L_1(\frac{st}{x})dx}{\int_{-\infty}^{0} g(x) \lim_{t\to+\infty} (-\frac{t}{x})^{-\alpha} L_2(-\frac{t}{x})dx + \int_{0}^{+\infty} g(x) \lim_{t\to-\infty} (-\frac{t}{x})^{-\alpha} L_1(\frac{t}{x})dx} \\ &= s^{-\alpha}. \end{aligned}$$

Note that in the last step we replace $L_2\left(-\frac{st}{x}\right)$ by $L_2\left(-\frac{t}{x}\right)$ and $L_1\left(\frac{st}{x}\right)$ by $L_1\left(\frac{t}{x}\right)$. Also note that by the moment assumption

$$\int_{-\infty}^{0} g(x)(-x)^{\alpha} dx + \int_{0}^{+\infty} g(x)x^{\alpha}dx$$

is bounded, and hence the above ratio of integrals is well defined. It follows that (36) is satisfied and $H(q)$ is a regularly varying function. Next, we turn to the probability of the sum $\Pr\left\{\sum_{t=1}^{T} X_t\Psi_t \leq -q\right\}$. By the independence assumptions each of the product random variables $X_t\Psi_t$ is independent. It then follows directly from Feller's theorem [1971, VIII.8] on the convolution of independent fat tailed random variables that asymptotically

$$\Pr\left\{\sum_{t=1}^{T} X_t\Psi_t \leq -q\right\} \approx Tq^{-\alpha}L_1(-q).$$

■

Remark 1 *In the above it is straightforward to allow for asymmetries in the tail shape. But this may be not very relevant for the log-transformed exchange rates. If the moment condition in the statement of the lemma is not satisfied, to the extent that $G(x)$ has a tail index $\widetilde{\alpha}$ such that $0 < \widetilde{\alpha} < \alpha$, then α should be replaced by $\widetilde{\alpha}$ in the claim.*

Remark 2 *The Hill estimates for the tail shape of the distribution of β_{t+1} should be identical to the tail index estimates for the y_{t+1} as reported in the literature.*

We can now address how $\Pr\left\{\widehat{\beta} \leq -q\right\}$ behaves for large q. Assume for simplicity the case that the time varying coefficient model (33) only exhibits multiplicative news, i.e., $\beta_{t+1} = 1 + \varepsilon_{t+1}$ and $\psi_{t+1} = 0$. Then it easily follows that standard OLS boils down to:

$$\begin{aligned}\widehat{\beta} &= 1 + \frac{\sum x_t^2 \varepsilon_{t+1}}{\sum x_t^2} \\ &= \sum w_t \varepsilon_{t+1},\end{aligned}$$

with $0 < w_t \leq 1$, $\forall t$. Because w_t is a bounded random variable the conditions of Lemma 1 are satisfied and $\widehat{\beta}$ is hampered by slow convergence. A proof for the general case of nonzero additive news is quite tedious but available upon request.

4 Monte Carlo evidence

The benchmark model we consider for simulation is the following:

$$s_{t+1} - s_t = (1 + \varepsilon_{t+1})(f_t - s_t) + v_{t+1}. \tag{37}$$

The RHS variables are sampled from the following distributions:

$$\begin{aligned} v &\sim cN_1, \\ (f - s) &\sim \frac{aM_t}{\sqrt{(N_2^2 + N_3^2 + N_4^2)/3}}, \\ M_t &= \rho M_{t-1} + N_5, \\ \varepsilon &\sim bN_6/M_t, \end{aligned} \tag{38}$$

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 Table 2.5 in Appendix 2.A. It can be easily shown that the choice $(a, b, c, \rho) = (1/1500; 10; 1/10; 0.7)$ is a good calibration for (37).

Besides their relative orders of magnitude, we also want to mimic the distributional characteristics of the variables in (37). The additive noise term is sampled from a standard normal distribution because in general we do not find that this variable is fat tailed. In contrast we find that the forward premium is highly fat tailed, see Table 2.5. Therefore we sample this term from a Student (3) distribution constructed as a ratio of a standard normal aM_t and the square root of a $\chi^2(3)$ variable divided by 3. Forward premia are highly dependent: one typically finds 1st order autocorrelations between 0.6 and 0.9 [5]. To reproduce this feature in the data M_t is drawn out of an AR(1) process with normally distributed innovations and where $\rho = 0.7$. Multiplicative news shocks are modelled as Cauchy-distributed innovations in order to create some large outliers ('big news') in the time varying 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. Appendix 2.E reports 100 simulations for the Fisher slope for T=200 and T=2000 respectively, see the composite Figure 2.10. Note that in the smaller samples of size 200 the order of magnitude of the OLS slopes corresponds to the -1 or -2 values commonly reported in the (univariate time series-oriented) 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 smaller variability in the OLS estimates. In addition, the standard error of the OLS series seems to decrease more quickly under the 'no time variation' scenario than under the 'time variation' scenario which supports the convergence result derived in the preceding theoretical section.

5 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,

[5] See also McCallum (1994) for this feature.

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. Simple no arbitrage arguments than showed that the multiplicative news component must be the same in all exchange rates. Hence, this news component can be identified from a cross section of different currencies. A panel estimator was used to estimate the time varying slope and to take account of the fixed time and currency effects in an appropriate manner. It turned out 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] Barnhart, S.W. and A.C. Szakmary, " Testing the unbiased forward rate hypothesis : Evidence on unit roots, co-integration, and stochastic coefficients," Journal of Finance and Quantitative Analysis 26, 1991, 245-267.

[2] Bekaert, G. and R.J. Hodrick, "On biases in the Measurement of Foreign Exchange risk premiums, " Journal of International Money and Finance 12, 1993, 115-138.

[3] Bilson, J.F.O., "The speculative efficiency hypothesis," Journal of Business, 54, 1981, 435-451.

[4] Campbell, J.Y. and R.H. Clarida, "The dollar and real interest rates," Carnegie-Rochester conference Series on Public Policy 27, 1987, 103-140.

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

[6] Dumas, B., "Partial equilibrium versus general equilibrium models of international capital markets," in F. van der Ploeg ed., The Handbook of International Economics, (Oxford, Blackwell), 1994, 301-347.

[7] Engel, C., "The Forward Discount Anomaly and the risk premium: A Survey of Recent Evidence, " NBER Working Paper 5312, 1995.

[8] Fama, E., "Forward and spot exchange rates," Journal of Monetary Economics 14, 1984, 319-338.

[9] Feller, W., An introduction to Probability Theory and its Applications, Vol.2, (New York, Wiley), 1971.

[10] Ferson, W.E. and C.R. Harvey, " The variation of economic risk premiums, " Journal of Political Economy 99, 1991, 385-415.

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

[12] Flood, R.P. and A.K. Rose, "Fixes: on the forward discount puzzle, " NBER Working Paper 4928, 1994.

[13] Flood, R.P. and A.K. Rose, "Fixes: on the forward discount puzzle, " Review of Economics and Statistics, 78, 1996, 748-752.

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

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

[16] Hansen, L.P. and R.J. Hodrick, "Forward exchange rates as optimal predictors of future spotrates: an econometric analysis," Journal of Political Economy 88, 1980, 829-853.

[17] Hansen, L.P. and R.J. Hodrick, " Risk averse speculation in the forward foreign exchange market: An econometric analysis of linear models, " in J.A. Frenkel (ed.), Exchange Rates and International Macroeconomics, (Chicago, Univ. Of Chicago Press), 1983, 113-152.

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

[19] Koedijk, K. and P. Schotman, "Dominant real exchange rate movements," Journal of International Money and Finance, 8, 1989, 517-531.

[20] Koedijk, K. and P. Schotman, "How to beat the random walk, an empirical model of real exchange rates," Journal of International Economics, 1990, 311-332.

[21] Koning, C. de, and S. Straetmans, "Variation in the slope coefficient of the Fama regression for testing uncovered interest rate parity: Evidence from fixed and time-varying coefficient approaches, mimeo, Tinbergen Institute, 1996.

[22] Lewis, K.K., "Puzzles in international financial markets", in G.M. Grosman and K. Rogoff eds., Handbook of International Economics Vol. III (Amsterdam, North-Holland), 1995, p. 1913-1971.

[23] Mayfield, E.S. and R.G. Murphy, "Interest rate parity and the exchange risk premium," Economics Letters 40,1992, 319-324.

[24] Mahieu, R.J. and P. Schotman, A factor approximation model for the international arbitrage pricing theory, mimeo, 1994.

[25] Nijman, T.E., F.C. Palm and C.P. Wolff, "Premia in forward foreign exchange as unobserved components: A note," Journal of Business & Economic Statistics 11, 1993, 361-365.

[26] Schotman, P.C.,"Empirical studies on the behavior of interest rates and exchange rates," Ph.D. thesis, Erasmus University Rotterdam, 1989.

[27] Robichek, A.A. and M.R. Eaker, " Foreign Exchange Hedging and the Capital Asset Pricing Model, " Journal of finance 33, 1978, 1011-1018.

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