

The Forex Forward Puzzle: the Career Risk Hypothesis

Fang Liu* and Piet Sercu[†]

JEL-codes: F31, G15.

Keywords: forward puzzle, uncovered interest parity, risk premium

First draft: July 2006

this version formatted at 20:55 on Thursday 25th December, 2008

*Katholieke Universiteit Leuven, Naamsestraat 69, B- 3000 Leuven; Tel: +32 16 326632; Fax: +32 16 326732; email: fang.liu@econ.kuleuven.be.

[†]Katholieke Universiteit Leuven, Naamsestraat 69, B- 3000 Leuven; Tel: +32 16 326756; Fax: +32 16 326732; email: piet.sercu@econ.kuleuven.be. We thank Stan Beckers, Tom Copeland, Lieven De Moor, Marco Lyrio and Kristien Smedts, and other participants in presentations at K.U. Leuven, at the ACE Conferences, Hong Kong, 2006 and 2007, and the Asia-Pacific Conference at Seoul, 2006, for useful suggestions. All remaining errors are ours, though.

Abstract

We conjecture that the forward puzzle may reflect career risks: when professional investors observe public danger signals about a currency, they require a premium for holding it. We find evidence of this in ERM rates. As deep discounts do signal danger, we next specify nonlinear variants of the Fama regression to capture this risk. We also decompose the forward premium into a long-memory trend and short-term component. We find empirical evidence for a career-risk premium; risk is in fact dominant in the trend component while the short-term component loads more on expectations. All confidence intervals are calculated via Monte Carlo.

1 Introduction

One empirical puzzle in international finance is the size of the bias in the forward premium as a predictor of future exchange rate changes. The Unbiased Expectations Hypothesis (UEH) posits that, in the Fama regression of exchange rate changes on beginning-of-period forward premia, the slope should equal unity and the intercept zero. However, as shown by Cumby and Obstfeld (1984), Fama (1984), and many others after them, the empirical coefficients are not only systematically below unity, but often even negative. The empirical results are all the more unexpected as, in unconditional tests over long periods, the cross section of time-series-average interest differentials does match the cross-section of time-series-average rates of appreciation quite well (Backus and Smith, 1993). One interpretation of the downward bias is that there is a missing variable that correlates with the forward premium, and a prime candidate is a risk premium. Another view is that, because of the near-unit-root characteristics of the regressor, the usual confidence intervals are vastly understated, see Roll and Yan (2000). In light of this, any new tests should be careful about the confidence intervals. While the Roll-Yan observation might help one to accept occasional estimates that are far from unity, it still does not explain why the coefficients are so consistently below unity across periods and currency pairs.

Some of the above issues become even more puzzling if one considers exchange rates within the European Exchange Rate Mechanism (ERM). First, in these rates there is strong mean-reversion, that is, there is predictability; so the usual remark that interest rates do not predict because there is nothing to predict does not apply here (Sercu, Vandebroek and Wu, 2008). Second, in the ERM context also the long-memory property of forward premia is a real puzzle. Member countries did coordinate their monetary policies, which should have led to co-movement in the interest rates, not randomly diverging rates. Also, it is hard to imagine that expectations of percentage changes in ERM exchange rates would be non-stationary. In light of this, the most likely cause of near-nonstationarity in forward premia would then be the missing variable (like a risk premium), a conjecture we try to substantiate in this paper. A third reason for focusing on the ERM is that it had clear anchors: the official bilateral parities and the admissible band; the associated multilateral measure of a currency's relative strength, called the divergence indicator; and the interest differential relative to the DEM, which was even canonized as an EMU accession criterion later on. All this made it very clear whether a currency was strong or looked threatened. These elements are useful to test a new hypothesis about the forward bias, inspired by the fallen-angel effect in stock markets: bearing in mind their career prospects, portfolio managers shun assets that fell badly, in the recent past, or

emit other very visible danger signals.

In the remainder of this intro we outline the paper's contributions: we study competing non-linear relations, and we filter out the long-memory component of the forward premium. In all of this, we take into account the near-unit-root problem in our significance tests.

Non-linearities have long antecedents as a potential explanation of the forward puzzle. Such a non-linearity can arise because the risk premium is approximately quadratic in the forward premium, as Bansal (1997) points out. The non-linearity could also be due to transaction costs (Huisman, R., Kee, Kool, C. and Nissen, F., 1998; Obstfeld and Rogoff, 2000) or "limits to arbitrage" (Lyons, 2001, Villanueva; 2005, Sarno, Valente and Leon; 2006, Baillie and Kiliç, 2006), notably when traders ignore gains that are of insufficient size relative to risk or transaction costs. Lastly, the risk premium could come from a career-risk-premium effect. In this view any public danger signal, like a pronounced forward discount, adds to the portfolio manager's reluctance to invest in the "fallen angel" (as the stock-market phrase goes). Like the Bansal and transaction-cost models, the career-risk hypothesis proposes a particular nonlinear model that can be written as a Fama regression whose slope, beta, varies depending on the forward premium. In this paper we specify the Fama beta as a quadratic function of the forward premium. We estimate these nonlinear models using overlapping one-month observations for EMU-member exchange and forward rates for the DEM. Wald test confirms the presence of the nonlinearities, and the models outperform the Fama version in terms of all standard goodness-of-fit measures. Also, more flexible functions than the quadratic add little explanatory power. The cubic model produces an inverse U-shaped beta plot, supporting the fallen-angel or career-risk hypothesis.

Our second contribution, after non-linearity, is filtering. Filtering is inspired by the fact that any nonstationarity in forward premia must come from either the expectations, or the risk premium (or, more generally, the missing variable), or both. Expected changes in ERM exchange rates are even less likely to have unit-root characteristics than those of floating rates, so the risk premium is the more likely source of the long memory in forward premium.

We decompose the time series of forward premia into a Hodrick-Presscott "trend" (which turns out to be non-stationary) and a stationary filtered component, and we re-run our generalized Fama regressions with as the regressor either this long-run component or its filtered part. Consistent with the idea that the long-memory part is more closely related to the risk premium while the filtered component loads more heavily on the expectations, we see that betas for the filtered premia are much higher, while those of the "trend" component in the

forward premium are clearly negative. In both betas, there is an inverted U shape, suggesting that the fallen-angel or career-risk factor has both long- and short-memory components; but the effect stronger in the long-memory component.

All our significance tests are based on Monte Carlo simulations where forward premia have strong memory and where the observations periods overlap, like in our data.

2 Non-linear variants of the Fama regression

2.1 *Earlier non-linear models*

Huisman et al. (1998) condition the Fama regression coefficients on the day-by-day cross-sectional variation of forward premia. They start from a particular view on where the missing variable comes from, namely, friction in the market. Since markets are subject to transaction costs, they argue, uncovered interest arbitrage cannot perfectly align expected exchange rate changes and forward premia. Most of the time, expectations of exchange rate changes are so small and diffuse that this friction-induced noise between expectations and premia largely obscures the theoretical parity between the two. However, there may be occasions where the market does expect unusually large changes; and if the impact of friction is essentially unaffected by the size of the expected change, then in these instances the signal-to-noise ratio must be relatively favorable. Highly positive or negative forward premia should therefore be better predictors than small premia. Cast in familiar statistical terms, the Fama regression suffers from an errors-in-the-regressor type bias towards zero, and for a given variance of the noise term this bias can be reduced by constructing a subsample where the cross-sectional variances of the regressor are larger. Huisman et al. test this model using panel techniques with a cross-currency constraint that ensures numeraire-invariance of the estimates. They report that large-variance observations generate Fama regression coefficients close to unity, and even substantially above unity if the definition of “large variance” is very strict.

Huisman et al.’s approach is similar, in spirit, to an earlier regression by Bilson (1981), who works with an equation of the type

$$\tilde{s}_{t,\Delta} = I_t^{extr} \cdot [\alpha_1 + \beta_1 f_{t,\Delta}] + (1 - I_t^{extr}) \cdot [\alpha_0 + \beta_0 f_{t,\Delta}] + \eta_{t,\Delta}, \quad (2.1)$$

where $\tilde{s}_{t,\Delta}$ is the percentage change, or log change, in the spot rate in the period $[t, t + \Delta]$; I_t^{extr} is an indicator that the forward premium observed at t is among the n percent most extreme observations; and $f_{t,\Delta}$ is the percentage or log forward premium at t for delivery at

$t + \Delta$. Thus, in both Huisman et al. and Bilson the Fama regression abruptly switches between parameters (α_1, β_1) and (α_0, β_0) depending on whether the forward observation is classified as extreme. One can experiment with the classification criterion n — *e.g.* the five, ten or twenty percent biggest premia—and see which one works best.

The more recent Limits-to-Arbitrage literature comes up with transition functions that are, basically, smoother and more flexible version of the Bilson equation. The Sarno et al. (2006) variant, for instance, would read like¹

$$\tilde{s}_{t,\Delta} = [\alpha_1 + \beta_1 f_{t,\Delta}] + \Phi(f_{t,\Delta}, \gamma) \cdot [\alpha_2 + \beta_2 f_{t,\Delta}] + \eta_{t,\Delta}, \quad (2.2)$$

with $\Phi(f_{t,\Delta}) := 1 - \exp[-\gamma(f_{t,\Delta})^2]$. Φ is an inverse bell-shaped function that assigns zero weight to the (α_2, β_2) version of the regression when $|f|$ is zero, and almost unit weight when $|f|$ is very large. Accordingly, Sarno et al. hypothesize that while α_1 and β_1 may be close to zero, we still have $\alpha_1 + \alpha_2 = 0$ and $\beta_1 + \beta_2 = 1$. Baillie and Bollerslev (2000) similarly use the lagged forward premium as the transition variable of the Logistic Smooth Transition Regression and also find evidence supporting the transaction-cost or limit-to-arbitrage theories. Mark and Moh (2002) describe similar effect concerning the presence of limit to speculation resulting from central bank intervention.

Bansal (1997) takes a very different perspective, focusing on the risk premium instead of friction. He starts from the CCAPM and establishes that its currency risk premium is approximately quadratic in the forward premium. Thus, the entire relation between expected change and forward premium becomes quadratic—inverse U-shaped, to be more precise. In his tests, Bansal approximates this by a piecewise-linear, inverse V-shaped relation,

$$\tilde{s}_{t,\Delta} = I_t^+ \cdot [\alpha_+ + \beta_+ f_{t,\Delta}] + (1 - I_t^+) \cdot [\alpha_- + \beta_- f_{t,\Delta}] + \eta_{t,\Delta}, \quad (2.3)$$

where I_t^+ is an indicator that the forward premium for period t is positive. Thus, in this model the Fama β changes discretely around $f = 0$; the hypothesis is that positive f s have a negative β and vice versa. If the approximation of the inverse U-shape by an inverse V is omitted, the Bansal equation can be written as involving a beta that, itself, is a negative linear function of f , thus producing the overall quadratic relation between the expectation and the forward premium. Table 1 sums up the models.

¹The argument in their Φ is the expectation, not the premium. Baillie and Kiliç (2006) do use the premium (in a logistic transition function that is similar to Sarno et al.'s).

Table 1: Overview of the nonlinear models

Key: "rp" denotes the risk premium, " β " the Fama (1983) beta and " f " the forward premium.

Model	rp in terms of f	β in terms of f
Huisman et al. (1998) and limit-to-arbitrage	—	U-shaped or inverse bell pattern
Bansal (1997)	– inverse U-shaped pattern	β negatively linear in f
	– inverse V-shaped approximation	$\beta > 0$ for $f < 0$; $\beta < 0$ for $f > 0$
Fallen-angel hypothesis	Cotangent shape, possibly asymmetric	inverse U or inverse V, possibly asymmetric

There are some other econometric studies on the nonlinearity in Data Generating Process (DGP). Domowitz and Hakkio (1985) propose a GARCH-in-mean model to capture the heteroscedastic conditional variance and any related time-varying risk premium. Although they find evidence of a varying risk premium, there is little support for the conditional variance as an important determinant. Tai (2001), in contrast, concludes that the multivariate GARCH-in-mean is picking up a clear time-varying risk premium.

2.2 The fallen-angel hypothesis

Our own tentative explanation is inspired by Sercu and Vinaimont's (2006) work on Peso risk, which seemed a promising but ultimately unsuccessful candidate explanation for the forward bias in the private ECU. Like the original Peso hypothesis, our hypothesis invokes "dark matter", that is, risks not observed by the econometrician. In the original Peso version, the dark matter is a low-probability but huge change. The potential change being huge, it does affect the expectation and therefore the forward premium; but in a finite sample the low-probability change may never be observed, so that the statistician would conclude that the forward premium systematically mispredicts the future spot rate.² One problem with this view is that most of the empirical evidence comes from floating rates, and one wonders what the

²For this to affect the regression coefficient rather than the intercept, the Peso risk must be time-varying and correlated with the forward premium. A plausible mechanism is as follows. When bad news about the foreign currency hits the market, the spot rate drops. But the concomitant selling of short-term paper (or borrowing against deposits) also pushes up the foreign interest rate, thus seemingly foretelling a further drop—or, if you wish, slowing down the immediate drop. If the Peso event then fails to materialize, the downward-pointing forward premium tends to be followed by a recovery in the spot rate, producing the negative regression coefficients.

huge Peso event might be if there is no system of interventions or exchange restrictions that keep the accumulating tensions bottled up for longish times. If one accordingly rejects the Peso view as implausible for floating rates, then it may seem that we are only inches away from the overreaction hypothesis. In this view, the huge change fails to materialize not because its probability is low, but because it exists only in the minds of the traders. People are subject to bouts of panic or overoptimism, causing soon-corrected movements in spot rates accompanied by changes in interest rates in the opposite direction. This fads & fashions view is what Sercu and Vinaimont ultimately come to for the private ECU.

Our own dark-matter variant, in contrast, invokes no such irrationality. The starting point is that the market is dominated by professional investors (traders or portfolio managers), not individuals playing with their own stakes. For a professional, the ultimate decision criterion is the personal career and remuneration prospects. This is not the same as the return on the portfolio to be managed because reputation and PV-ed remunerations are not linear in the portfolio return, and depend also on how and when any losses have occurred. Suppose there is bad news about a foreign currency, immediately showing up in a falling spot rate and a rising foreign money-market rates, *i.e.* a falling forward premium. The manager may play it safe and liquidate the foreign positions, thus risking to miss a recovery; or she may act contrarian and stay long, risking a further drop in the spot rate. In making the choice she will note that a cash loss looks worse than an opportunity loss, in general. But a cash loss from being contrarian (when there has been a clear and publicly observable bad initial signal) looks much worse than an opportunity loss from missing a rally which, judging by the initial forward premium, was deemed to be the less likely outcome anyway. Any cash loss from going against the flow will be met with the comment that the trader “should have seen it coming”, but the opportunity loss from following the consensus signal will not. In short, when bad news hits the market, professional investors head for the exit even if there is an expected gain from the subsequent recovery, because the expected gain from the recovery is counterbalanced by a dark matter, the potential damage to the professional investor’s career if expectations turn out to be wrong. This career-risk hypothesis is consistent with the stop-loss behavior (selling after a loss) that is so pervasive among currency traders. In the stock market this is known as the “fallen angel” effect: stocks that did badly are shunned by portfolio managers and, therefore, generate high returns (Ikenberry, D., Lakonishok, J. and Vermaelen.T., 1995).

This particular dark-matter theory is testable, unlike the strict Peso view or its (statistically indistinguishable) overreaction counterpart, because it does not involve invisible variables.

Instead, it predicts a risk premium for holding currencies with public danger signals. An unusually negative forward premium would certainly be one such warning sign, triggering a positive extra required return. The amount of expected return the manager is willing to give up is the career risk premium, which is expected to be positive and large when the forward premium is negative and to be unimportant when the forward premium is around zero. Similarly, the manager would be willing to give up some expected return for the safety promised by a markedly positive forward premium. In short, the missing variable exhibits a cotangent-shaped relation to f : sloping down in quadrant IV, leveling off around the origin, and then dipping down again in quadrant II. Note that the relation between the forward premium and the fallen-angel premium needs not be symmetric. While it takes a large expected reward to go against a warning signal and risk a cash loss, a smaller return shortfall may already be enough to make the trader ignore a positive signal, because the risk of ignoring the signal is just an opportunity cost.³

Thus, to model the private risk premium as a function of f we chose a negative-sloping and possibly asymmetric function that is probably quite flat when f is close to zero but is steeper for values of the forward premium farther away from zero. Let us use the notation $x_+ := \text{Max}(x, 0)$ and $x_- := \text{Min}(x, 0)$ to denote observations selected by sign. Within the class of low-order polynomials we could then chose a piecewise quadratic, like $\zeta f_-^2 - \eta f_+^2$ with $\zeta > 0$, $\eta > 0$, or a piecewise cubic like $-\zeta f_-^3 - \eta f_+^3$. The corresponding test equations are,

$$\begin{aligned} & \text{fallen-angel premium} \\ \text{(quadratic:)} \quad E_t(\tilde{s}_{t,\Delta}) - f_{t,\Delta} & \approx \zeta (f_{t,\Delta})_-^2 - \eta (f_{t,\Delta})_+^2; \\ & \Rightarrow E_t(\tilde{s}_{t,\Delta}) \approx [1 + \zeta (f_{t,\Delta})_- - \eta (f_{t,\Delta})_+] f_{t,\Delta}, \end{aligned} \quad (2.4)$$

$$\begin{aligned} \text{(cubic:)} \quad E_t(\tilde{s}_{t,\Delta}) - f_{t,\Delta} & \approx -\zeta (f_{t,\Delta})_-^3 - \eta (f_{t,\Delta})_+^3; \\ & \Rightarrow E_t(\tilde{s}_{t,\Delta}) \approx [1 - \zeta (f_{t,\Delta})_-^2 - \eta (f_{t,\Delta})_+^2] f_{t,\Delta}. \end{aligned} \quad (2.5)$$

Thus, here the betas (shown in the square brackets) are predicted to be inverted U- or V-functions of f .

While it is not difficult to test whether the predicted patterns are present or not, finding that it does not necessarily mean that they reflect a fallen-angel effect. Thus, before starting

³This would be even more so if the foreign currency is exotic and the home currency a major one. The default view in academia is that the choice of the base currency does not matter, fundamentally, but it is also a fact that more managers report in, say, GBP or USD than in DKK or BEF; thus, if there is an asymmetry in the private risk premia, the bigger currency's point of view is likely to dominate.

the main tests we want to verify whether it is generally true that danger signals other than the forward premium seem to generate a fallen-angel risk premium. If so, this finding would lend extra credibility to our interpretation that also a big forward premium is a fallen-angel signal.

3 Preliminary tests

3.1 Data

The currencies we work with are the Belgian Franc (BEF), German Mark (DEM), Danish Krone (DKK), French Franc (FRF), Dutch Guilder (NLG), Spanish Peseta (ESP), Irish Punt (IEP), Italian Lira (ITL) and Austrian Schilling (ATS). All data are acquired from DataStream. We use weekly observations on one-month forward contracts, and the sample period is usually from January 1st, 1976 to December 31st, 1998 (1200 observations). The one exception is the IEP, whose Datastream coverage starts on April 2nd, 1979 (1030 observations). The future spot rate is the spot rate on the delivery day, *i.e.* two working days plus one month after the date when the transaction is agreed. The DEM acts as the base currency; that is, exchange rates show the value of one DEM in units of the other currency.

3.2 Do danger signals lead to excess returns?

Let us define the expected excess return, $R_{t,\Delta}^e$, as the exchange-rate change in excess of the forward premium. The expected excess return should be zero in absence of the risk premium and irrational expectations. In the modified Fama regression,

$$\begin{aligned} R_{t,\Delta}^e &:= \tilde{s}_{t,\Delta} - f_{t,\Delta}, \\ &= \alpha + (\beta - 1)f_{t,\Delta} + \sum_j \gamma_j X_{j,t} + \eta_{t,\Delta}, \end{aligned} \tag{3.6}$$

the intercept α and slope $(\beta - 1)$ should both equal zero under the null hypothesis. In addition, extra regressors X_j should have no explanatory power. In this section we verify whether excess returns have non-zero γ s for danger signals other than the forward premium, as our fallen-angel hypothesis predicts. As possible proxies for danger signals, we use the first three principal components of seven possible danger signals; the underlying variables and the eigenvectors are discussed in the Appendix. The three synthetic risk signals refer to (i) the multilateral position (“divergence”) in the ERM band; (ii) the change of the bilateral value of the DEM; and (iii) the change of the divergence. The signs of the weights are such that a higher value of the

Table 2: **The predictive power of the principal components of seven danger signals**

Key: Data are daily observations with the DEM as base currency, for overlapping monthly returns in excess of the forward premium. The sample is within the narrow-band period of the ERM, from April 2nd, 1979 till August 3rd, 1993.

$$R_t^e = \alpha + \gamma_1 \xi_{1,t} + \gamma_2 \xi_{2,t} + \gamma_3 \xi_{3,t} + v_t$$

	γ_1	γ_2	γ_3	R^2
BEF	***0.0008	***0.0003	***0.0003	0.024
DKK	***0.0005	0.0001	*0.0002	0.006
FRF	***0.0003	***0.0005	-0.0001	0.008
NLG	***0.0006	***0.0004	***-0.0002	0.078
ITL	***-0.0008	** -0.0003	-0.0002	0.009
IEP	***-0.0003	*0.0002	0.0003	0.002
ESP	***0.0007	0.0003	0.0004	0.002
number of right signs	5	6	4	

In this paper, ***, ** and * indicate statistical significance at the 0.001, 0.01 and 0.05 levels, respectively.

principal component should be associated with a higher expected return on the DEM; that is, the regression coefficient should be positive if danger signals lead to risk premia.

We now test if the conjecture is right or not by regressing the excess return on the principal components in Equation (3.7),

$$R_t^e = \alpha + \gamma_1 \xi_{1,t} + \gamma_2 \xi_{2,t} + \gamma_3 \xi_{3,t} + v_t. \quad (3.7)$$

The estimated slopes and R^2 s are reported in Table 2. While under the null no coefficient should be significant, bar perhaps one or two on a pure-chance basis, we see fifteen starred estimates out of twenty-one coefficients. The intervention factor ξ_1 has five times a significant positive coefficient (the two exceptions being the currencies with systematically large forward premia, ITL and IEP). ξ_2 comes up with a positive value in six out of seven equations, albeit with unsatisfactory significance in two cases. The third component, ξ_3 , has a positive coefficient in four equations, although only two of them are significant. Most of the wrong signs are insignificantly different from zero. For each currency, the R^2 is low but is nevertheless much higher than for the regression where the forward premium is the regressor. So, the fallen-angel hypothesis acquires some credibility: the empirical evidence implies that general danger signals other than the interest differential trigger higher expected returns, consistent with the idea that career- or image-risk considerations drive the manager to ask for extra return.

3.3 Does more volatility lead to excess returns?

The standard Fama regressions are consistent and efficient under the assumptions of constant variance, but it is well proven that the variance of exchange rate changes is heteroscedastic. In

Table 3: Fama regressions and the GARCH-in-mean Versions

Key: Data are Monday-weekly observations, with the DEM as base currency, from April 2nd, 1979 to September 14th, 1992.

Panel A: GARCH-in-Mean $\tilde{s}_{t,\Delta} = \alpha + \beta f_{t,\Delta} + \delta \sigma_t + \epsilon_t$;
 $\sigma_t^2 = \omega + \alpha_\epsilon \epsilon_{t-1}^2 + \alpha_\sigma \sigma_{t-1}^2 + \sum_{k=1}^m d_k I_{t,k}$.

	ATS	BEF	DKK	FRF	NLG	ITL	ESP	IEP
α_ϵ	*0.15	***0.12	*0.10	*0.21	***0.22	***0.16	***0.13	***0.15
α_σ	***0.59	***0.57	***0.54	***0.50	***0.52	**0.47	**0.51	***0.60
$\alpha_\epsilon + \alpha_\sigma$	0.74	0.69	0.64	0.71	0.74	0.63	0.64	0.75
β	-0.28	0.08	0.04	***0.48	***-0.62	0.28	*0.55	-0.08
δ	-0.89	**1.32	1.22	*1.33	***0.42	—	***0.56	***1.79

Panel B: Standard Fama $\tilde{s}_{t,\Delta} = \alpha + \beta f_{t,\Delta} + \epsilon_t$;

	ATS	BEF	DKK	FRF	NLG	ITL	ESP	IEP
β	-0.23	**0.49	0.37	**1.11	*-0.58	0.32	*1.03	0.28

addition, a changing variance may be associated with a changing risk premium. In standard asset pricing, this could be true only if variance is related to some more fundamental covariance-risk measure, but under the career-risk view a higher volatility would be a danger signal like the others we just considered. We look for a variance effect via the GARCH-in-mean model, as proposed by Engel, Lilien and Robins (1987), in which the first equation links the expectation to high volatility:

$$\tilde{s}_{t,\Delta} = \alpha + \beta f_{t,\Delta} + \delta \sigma_t + \epsilon_t, \quad (3.8)$$

$$\sigma_t^2 = \omega + \alpha_\epsilon \epsilon_{t-1}^2 + \alpha_\sigma \sigma_{t-1}^2 + \sum_{k=1}^m d_k I_{t,k}. \quad (3.9)$$

In the mean Equation (3.8) the standard deviation σ_t is added in the traditional Fama regression in the hope of capturing part of the missing variable. In the volatility Equation (3.9), the conditional variance σ_t^2 depends on its own lag σ_{t-1}^2 and the previous realized volatility ϵ_{t-1}^2 . We add dummy variables $I_{t,k}$ that indicate whether or not a realignment took place during the period t because the realignments do not fit the GARCH model. This lack of fit holds not only at conceptual level, but also empirically in the sense that without the dummies the value of $(\alpha_\epsilon + \alpha_\sigma)$ is larger than unity.⁴ We find that the F-statistics of the joint test on the dummy coefficients d_k are significant,⁵ and the values of $(\alpha_\epsilon + \alpha_\sigma)$ do become smaller than unity after filtering out the realignments effect.

The estimates in Table 3 are based on the Monday-weekly data from April 2nd, 1979

⁴The constraint of $(\alpha_\epsilon + \alpha_\sigma) \leq 1$ makes sure the volatility is converging.

⁵This empirical result is available on request.

to September 14th, 1992. These changes reflect the considerations: first, most realignments happened on Sundays, and the variances on the following Mondays should be peaking relative to those of the rest working days once the news was released; second, the ITL and the GBP dropped out of the EMS after September 1992, so, we study the sub-sample to avoid a potential bias due to the regime switching in the volatilities. The estimations of the variance Equation (3.9) are presented in the first two rows of Panel A. All the α_ϵ and α_σ are positive and significant, and the values of $(\alpha_\epsilon + \alpha_\sigma)$ are below unity but large, indicating that the conditional variance is heteroscedastic, or the shocks of the volatilities are highly persistent. So, larger jumps are associated with higher expectations. In the main test, we proxy the jump effect with cubic model in terms of forward premium. When we come to the mean equations, we find that: (i) the β becomes smaller compared with the β from the traditional Fama regressions in Panel B, and (ii) the slope for the standard deviation δ is significant and positive except for the ATS. Therefore, the decreased β s and the positive δ s may suggest that the traditional Fama regressions miss a correlation between f and σ and there exists a time-varying risk premium or another missing variable related to σ . Since increased σ is a danger signal, the findings are consistent with the career-risk hypothesis.

3.4 *Can VECM regressions solve the puzzle?*

Nonstationarity in a variable is one of the pieces of the forward puzzle. Regression makes sense only with stationary variables or if the dependent variable co-integrates with the independent variable. Many studies⁶ find that the forward rate F_t co-integrates with the spot rate S_t , and the cointegrating vector is around (1,-1); furthermore, the forward premium often rejects the unit root but is still highly persistent; and in the Vector Error Correction Model (VECM), the Fama β is much lower than unity, so the UEH is rejected.

Here, we analyze the intra-ERM data with a VECM, which is designed to capture both the long-run relation and short-term dynamics of the co-integrated variables. Judging by the Johansen cointegration test (available on request), F is co-integrated with S for all the observed

⁶See the summary of cointegrating vectors estimated of the studies on currency markets in Table 2 of Brenner and Kroner (1995).

Table 4: The estimations of the VECM

Key: Here, the data frequency is monthly, to make the left-hand side variable ΔS_{t+1} match the rates on one-month forward contract. So, the standard deviations do not suffer the overlapping problem anymore and are calculated under the OLS. We study the whole sample period, from January 1976 to December 1998.

	$\Delta S_{t+1} = \alpha_s + \beta_s(F_t - S_t) + \sum_i \rho_{f,i}\Delta F_{t-i} + \sum_i \rho_{s,i}\Delta S_{t-i} + \epsilon_{st}$							\bar{R}^2
	β_s	$\rho_{f,1}$	$\rho_{s,1}$	$\rho_{f,2}$	$\rho_{s,2}$	$\rho_{f,3}$	$\rho_{s,3}$	
ATS	**0.51	***-0.99	**0.56	*-0.41	0.24			0.18
BEF	***0.63	-0.21	0.22					0.04
DKK	0.18	**0.64	**0.73					0.01
FRF	***0.82	**0.83	**0.91	***-1.25	***1.17	*0.53	*-0.55	0.14
NLG	**0.51	**0.41	**0.45					0.07
ITL	*-1.10	-0.83	0.91	-0.07	0.09	1.31	-1.15	0.07
ESP	0.43	**1.01	**0.99	**1.33	***-1.34			0.05
IEP	*-0.60	0.25	-0.19					0.01

currencies, thus the UEH can be tested with the VECM,

$$\Delta F_{t+1} = \alpha_f + \beta_f(F_t - S_t) + \sum_i \rho_{f,i}\Delta F_{t-i} + \sum_i \rho_{s,i}\Delta S_{t-i} + \epsilon_{ft}, \quad (3.10)$$

$$\Delta S_{t+1} = \alpha_s + \beta_s(F_t - S_t) + \sum_i \rho_{f,i}\Delta F_{t-i} + \sum_i \rho_{s,i}\Delta S_{t-i} + \epsilon_{st}. \quad (3.11)$$

Equation (3.11) is the Fama regression, extended by adding the lagged differentials of F_t and S_t , ΔF_{t-i} and ΔS_{t-i} . The co-integrating vector between F_t and S_t , whose general form would be $(a F_t - b S_t)$, has been restricted to equal $(1,-1)$,⁷ so, the cointegration equation $(F_t - S_t)$ is the forward premium and the β_s is the Fama beta. Under the UEH, the additional regressors should have zero coefficients. So if the slope coefficients $\rho_{f,i}$ and $\rho_{s,i}$ are significantly different from zero, the dynamic items must pick up either some inefficiency or part of the missing variable(s).

Table 4 reports the estimates of Equation (3.11), where lag orders of dynamic items have been selected in terms of the AIC. In this test the data frequency is monthly nonoverlapping returns (not weekly observations on overlapping one-month returns), to make the left-hand side variable ΔS_{t+1} match the rates on a one-month forward contract. The β_s in the second column are still smaller than unity, indicating the UEH is rejected. However, the adjusted \bar{R}^2 s improve by allowing for the short-term dynamics. In the columns of slope coefficients of the differentials, most of the $\rho_{f,i}$ and $\rho_{s,i}$ are significantly different from zero. We also notice

⁷The estimated co-integrating vector is insignificantly different from $(1,-1)$. This empirical result is available on request.

that, in all cases, the $\rho_{f,i}$ is similar in size but opposite in sign to the $\rho_{s,i}$. Wald tests, not shown here, establish that in all cases except the ATS there is no significant difference between $|\rho_{f,i}|$ and $|\rho_{s,i}|$. Therefore, the dynamic part can be interpreted as referring to changes in the forward premia:

$$\begin{aligned} \sum \rho_i \Delta F_{t-i} - \sum \rho_i \Delta S_{t-i} &= \sum \rho_i \Delta (F_{t-i} - S_{t-i}), \\ &= \sum \rho_i \Delta f_{t-i}. \end{aligned} \quad (3.12)$$

The significant coefficients imply that the changes in the forward premia are related to the missing variable(s). All significant coefficients except the one of the ESP are negative, in fact. Thus, a rising forward premium for the DEM is followed by a below-average returns on the DEM or, in other words, a falling premium for the foreign currency is followed by a above-average excess return on that currency. The message from the VECM, then, is that changes in forward premia seem to act as additional danger signals, exactly like the other ones we considered before.

All this lends credibility to the notion that, within the ERM, danger signals were followed by excess returns on the foreign currency. The standard Fama test ignores all these risk effects because they should all be reflected in the forward premium anyway; but, as we know, this hypothesis is systematically rejected in the literature. The question we now address is to what extent the forward premium may have picked up the missing risks. So we again omit the danger signals from the model and return to the original Fama version, except that we allow for non-linearities.⁸

4 Main test

4.1 *Description on the nonlinear models*

Our main test models are part of the literature that explains the puzzle as a failure to allow for a risk premium in the Fama regression. Generalizing Bansal's approach, we present two models that approximately fit all of the non-constant-beta hypotheses discussed in Section 2.

⁸A separate decision was to drop not just the mean equation of GARCH, but also the variance equation. The reasons relate to fact that the main tests use monthly holding periods. In monthly returns, GARCH effects are much weaker than in the weekly returns we studied in this section. In addition, we use overlapping returns, but GARCH requires non-overlapping returns, which would mean we lose a lot of observations. Note that the confidence intervals are based on Monte Carlo, not on the standard errors derived from homoscedasticity-based theory.

The first one is a cubic model where the risk premium takes the form of higher orders of the forward premia,

$$\begin{aligned} E_t(\tilde{s}_{t,\Delta}) &= \alpha + \beta_1 f_{t,\Delta} + \beta_2 f_{t,\Delta}^2 + \beta_3 f_{t,\Delta}^3, \\ &= \alpha + f_{t,\Delta} + \underbrace{(\beta_1 - 1)f_{t,\Delta} + \beta_2 f_{t,\Delta}^2 + \beta_3 f_{t,\Delta}^3}_{\text{risk premium}}. \end{aligned} \quad (4.13)$$

Thus, the Fama beta becomes quadratic in the forward premia,

$$\beta(f_{t,\Delta}) = \beta_1 + \beta_2 f_{t,\Delta} + \beta_3 f_{t,\Delta}^2. \quad (4.14)$$

In empirical section we will examine the nonlinearity by testing the joint hypothesis $\beta_2 = \beta_3 = 0$. If and when the null of a linear model is rejected, we can check the observed pattern against the alternatives set forth in Table 1. But the quadratic approximation for the Fama β may lack flexibility; the tails of an inverse bell shape, for instance, can not be captured. As a more flexible alternative to the cubic model in Equation (4.13), we therefore let the Fama beta be a quadratic spline function of the forward premia:

$$\beta(f_{t,\Delta}) = \beta_1 + \beta_2 f_{t,\Delta} + \beta_3 f_{t,\Delta}^2 + d_1(f_{t,\Delta} - k_1)_+^2 + \dots + d_p(f_{t,\Delta} - k_p)_+^2. \quad (4.15)$$

In this equation the Fama regression is implicitly revised into a nonlinear form,

$$E_t(\tilde{s}_{t,\Delta}) = \alpha + [\beta_1 + \beta_2 f_{t,\Delta} + \beta_3 f_{t,\Delta}^2 + d_1(f_{t,\Delta} - k_1)_+^2 + \dots + d_p(f_{t,\Delta} - k_p)_+^2] f_{t,\Delta}. \quad (4.16)$$

Familiarly, the quadratic spline function in equation (4.15) consists of two parts: a quadratic function of f and the plus functions, $(f - k_i)_+^2$ for $i = 1, \dots, p$. The pre-set parameters k_i are referred to as the knot points; so the squared positive differences between f and the knot points are included into the plus functions. Like a $(p + 3)$ -th degree polynomial, the spline function is continuous in its level and first-order derivative, but it allows the second-order derivative to change at each knot point without any repercussions on the function for lower values of f . Here, we set the knot points as follows. The percentage forward premia are ranked by size, and for each currency six knot points separate the whole set of observations into seven bands. The values of the knots for a given currency are the 5th, 10th, 20th, 80th, 90th and 95th top percentile values of the sample of the forward premia of the currency. Note that the first band corresponds to the lowest premia, and the seventh band to the highest premia. While the central zone is wide in terms of frequencies, in algebraical terms it is not wide because the density is much higher in the middle. For example, in Table 5 we show the knot points for the raw forward premium of the FRF, as well as for its two decompositions to be introduced in Section 5.

Table 5: Knot-point values of the FRF in the spline

Key: "f" is the forward premium, " \hat{f} " is the (common) long-memory component, "ff" is the filtered forward premium, "Cum prob" is the cumulative probability and "S.D." is the standard deviation. All the numbers are expressed in percent.

variable (%)	Mini- mum	knot1	knot2	knot3	knot4	knot5	knot6	Maxi- mum	Mean	S.D.
Cum prob (%)	0	5	10	20	80	90	95	100		
f	-0.15	0.00	0.01	0.04	0.53	0.69	0.88	2.61	0.33	0.33
\hat{f}	-0.05	0.06	0.08	0.10	0.53	0.63	0.69	0.98	0.33	0.21
ff	-0.52	-0.28	-0.22	-0.12	0.09	0.21	0.39	2.08	0.00	0.25

4.2 Empirical tests and model evaluation

Table 6 shows the summary results for the linear, cubic and spline regressions. The values for the standard Fama β in the second column are significant and positive, a rare result in this literature. However, we still find the Fama model is misspecified: there is a clear non-linear relation between the spot changes and the forward premia. Here, nonlinearity is demonstrated if the Wald test rejects the null hypothesis of a linear relation, the original Fama model; in other words, we jointly test whether the coefficients β_2 and β_3 equal zero for the cubic model or similarly for $\beta_2, \beta_3, d_1, d_2, \dots, d_p$ equal zero for the spline model. From Table 6 both the cubic and the spline have five regressions rejecting the joint tests, indicating the presence of nonlinearity. The beta plots in Figure 1 also illustrate the nonlinearity.

Next, we evaluate the goodness-of-fit via three measures: \bar{R}^2 , the Akaike Information Criterion (AIC) and Schwarz Information Criterion (SIC). We compare the nonlinear test models to the Fama linear model in Table 7. The cubic regression outperforms the Fama

Figure 1: Beta plot with the cubic models

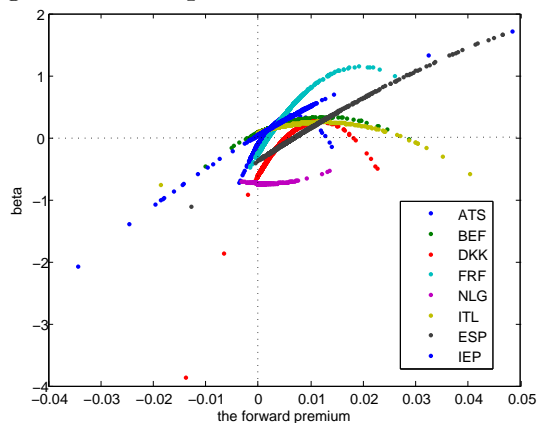


Table 6: Fama, cubic and spline regressions

Key:The Fama panel reports constant Fama β s; the Cubic panel reports all the coefficients of the cubic models and Wald tests for the joint hypothesis that β_2 and β_3 equal zero; the spline panel shows values of $\beta(f_{t,\Delta})$ computed at each band's midpoint value, not the coefficients themselves. The significance tests for each $\beta(f_{t,\Delta})$ and the Wald tests are based on the Monte-Carlo simulations described in the Appendix. The Wald tests show the presence of nonlinearity if the joint hypothesis that all the coefficients of the plus functions and of higher orders of $f_{t,\Delta}$ equal zero. The main tests are based on the weekly observations on the whole sample period, from January 1st, 1976 to December 31th, 1998, by default in the following tables without special descriptions on the data.

	Fama		Cubic				Spline							Wald
	β		β_1	β_2	β_3	Wald	band 1	band 2	band 3	band 4	band 5	band 6	band 7	
ATS	*0.11		-0.11	***138.02	-1.00e+4	***4.75	-0.36	0.21	0.01	0.27	-0.12	0.07	***0.31	1.55
BEF	***0.25		0.08	**38.60	-1.47e+3	0.45	***-1.70	***-2.19	*-1.41	-0.08	*0.26	**0.28	***0.27	***31.86
DKK	***0.15		***-0.61	***148.65	-6.35e+3	***22.93	-0.14	0.55	0.49	0.22	***0.47	***0.31	0.11	***142.85
FRF	***0.78		-0.26	***142.65	-3.63e+3	*2.62	***-5.51	-0.13	1.27	*0.73	***0.81	**0.42	***0.95	1.01
NLG	***-0.66		***-0.74	-5.16	1.50e+3	0.08	***-1.97	***-1.54	-0.44	0	-0.31	***-0.46	***-0.65	0.64
ITL	***0.16		0.08	26.30	-1.06e+3	0.75	-0.68	-0.49	-0.03	-0.01	0.23	**0.56	0.04	***4.08
ESP	*0.91		-0.37	55.55	-251.97	*2.38	**3.68	2.73	**1.71	**0.79	**0.56	*0.34	***1.20	***7.89
IEP	*0.15		0.04	***50.40	-329.33	***23.84	-0.96	** -1.22	***-1.79	** -0.36	-0.19	0.23	0.53	***34.68

$$\text{Spline: } \tilde{s}_{t,\Delta} = \alpha + \beta(f_{t,\Delta})f_{t,\Delta} = \alpha + [\beta_1 + \beta_2 f_{t,\Delta} + \beta_3 f_{t,\Delta}^2 + d_1(f_{t,\Delta} - k_1)_+^2 + \dots + d_p(f_{t,\Delta} - k_p)_+^2]f_{t,\Delta} + \tilde{\epsilon}_{t,\Delta}$$

$$\text{Cubic: } \tilde{s}_{t,\Delta} = \alpha + \beta f_{t,\Delta} + \tilde{\epsilon}_{t,\Delta}$$

$$\text{Fama: } \tilde{s}_{t,\Delta} = \alpha + \beta f_{t,\Delta} + \tilde{\epsilon}_{t,\Delta}$$

Table 7: **Goodness of fit: Fama, cubic and spline regressions**

Key: “#(Regressions that outperform Fama)” is the number of equations where the nonlinear models outperform their linear counterparts in terms of the measures of Goodness of fit; “#(Regressions that outperform Cubic)” is the number of equations where the cubic model outperforms spline counterpart.

	Fama			Cubic			Spline		
	\bar{R}^2	AIC	SIC	\bar{R}^2	AIC	SIC	\bar{R}^2	AIC	SIC
ATS	0.28%	-8.598	-8.590	1.11%	-8.605	-8.588	1.30%	-8.602	-8.559
BEF	0.96%	-6.899	-6.891	1.06%	-6.898	-6.881	2.37%	-6.907	-6.864
DKK	0.19%	-6.671	-6.663	4.96%	-6.720	-6.701	7.16%	-6.737	-6.694
FRF	6.84%	-6.250	-6.242	7.92%	-6.262	-6.243	8.88%	-6.266	-6.223
NLG	5.13%	-8.131	-8.123	5.03%	-8.129	-8.112	5.62%	-8.130	-8.088
ITL	-0.04%	-4.913	-4.905	0.17%	-4.914	-4.897	0.61%	-4.913	-4.871
ESP	8.46%	-4.799	-4.790	10.90%	-4.842	-4.807	12.19%	-4.834	-4.791
IEP	0.18%	-5.826	-5.187	9.83%	-5.933	-5.912	10.90%	-5.932	-5.884
#(Regressions that outperform Fama)				7	6	4	8	6	3
#(Regressions that outperform Cubic)				—	—	—	8	4	0

version in terms of adjusted \bar{R}^2 and AIC: seven \bar{R}^2 s are better and so are six AICs. The SIC, the most stringent measure, still reports four cubic models that outperform the Fama regressions. Also, the spline regressions have an advantage in \bar{R}^2 and AIC compared to Fama regressions: all \bar{R}^2 s are better, and so are six AICs. Different from the cubic model, however, the spline model does a bad job in terms of SIC: only three regressions get improved betas relative to the constant-beta model. The third pairwise comparison is between these two nonlinear models, presented in the last row of Table 7. The spline is superior to the cubic only in terms of \bar{R}^2 , but turns out to do worse in terms of SIC for all the regressions. To summarize, the nonlinear models have clearly better goodness-of-fit than the linear Fama model for \bar{R}^2 and AIC, but relative to the cubic the spline adds little extra and does even worse in terms of SIC.

As mentioned before, the spline regression is recommended when we have no strong priors on the shape of the beta plot. Table 6 reports the estimated mid-point betas of each band. The estimates turn out to be quite imprecise, though; this may be the reason why we do not find any clear pattern. The cubic models, in contrast, produce similar patterns: low or negative values for β_1 (much lower than the standard Fama betas), positive coefficients for the quadratic term in seven cases out of eight, and negative ones for the cubic in seven cases out of eight again. The results are investigated more closely in the following section.

At this point we conclude that the Fama model misses a nonlinearity, and is inferior to the nonlinear models in terms of goodness-of-fit. There is, however, no evidence that we need to go beyond a cubic model. So we now abandon splines and restrict the generalized Fama beta to a simpler form, a simple quadratic function of the forward premium f .

4.3 *The cubic model: further discussion*

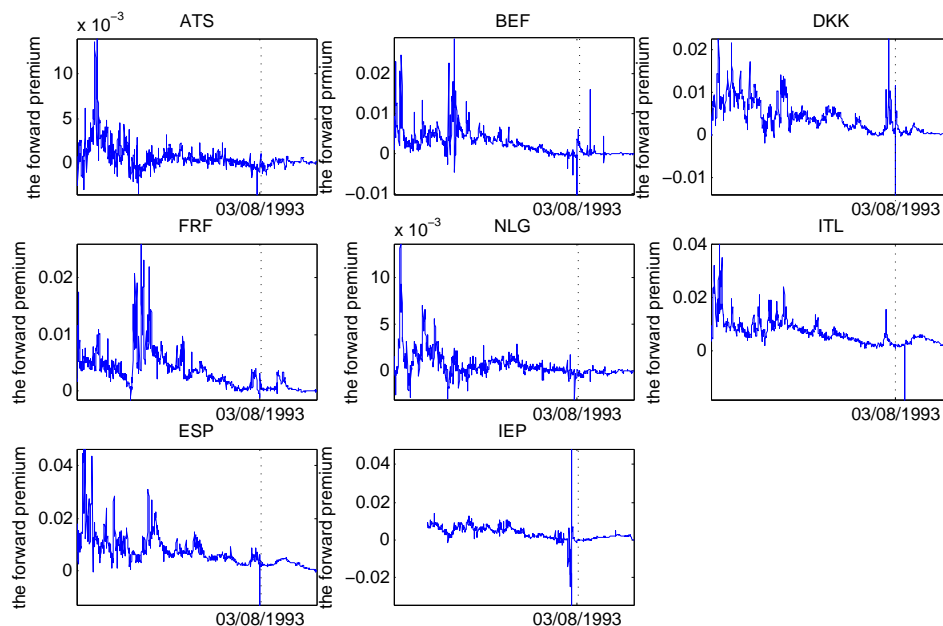
From Table 6, we see that there is a nonlinear relation: the coefficients β_2 are significant in five cases out of eight and the Wald tests reject the constant beta in five regressions. Even though the cubic coefficient β_3 is never significant individually (t-test), its sign is always negative except for the NLG, suggesting that the parabola of the beta plots opens downward. Consistent with the negative signs of β_3 , Figure 1 shows that the Fama betas are inverse U-shaped functions of the forward premia. True, the inverse U-shape is not uniform across currencies: we see only a weak convexity for the ESP and the IEP, and the NLG actually has a slightly concave line; but there certainly is no pattern of betas that are negative functions of the forward premium, the pattern that would have supported the Bansal risk premium; nor is there any U or V shaped pattern as proposed by the transaction cost or limit-to-arbitrage hypothesis. The only story that seems to be consistent with the data, in short, is the fallen-angel hypothesis. A possible explanation for the concave line for the NLG is that this currency was closest to being a DEM clone, so the career-risk logic might not have applied at all. In most cases, the max of the inverse U tends to occur at positive forward premium, 1% to 2%, indicating that the bad vibes seem to start not when forward premia fall below zero but already when they fall below a critical positive level.

To control for a possible regime shift, we now separate the whole sample into two parts, subsample 1, before August 3rd, 1993, and subsample 2 after this date. On August 3rd, 1993 the ERM exchange rate fluctuation bands were widened from $\pm 2.25\%$ to $\pm 15\%$ around the central parities to relieve heavy speculation pressure.⁹

Figure 2 depicts the movements of the forward premia. Compared with subsample 1, the forward premia in subsample 2 move within a much narrower zone. This might mean that with the wide fluctuation band the investors expected less speculation, but towards the end of the sample period they surely also anticipated more financial integration among ERM

⁹During 1992-93 the ERM experienced a crisis due to persistent inflation differentials and diverging competitiveness across the members. After the German reunification, capital flows from West to East Germany caused a large fiscal deficit, followed by an inflationary tendency, so the Bundesbank raised its discount rate. Meanwhile, the other ERM countries wanted to lower interest rates to get out of recession. Inflation differentials also aggravated the diverging competitiveness across members. Italy's competitiveness, for example, suffered from persistent inflation and rising labor costs, while Germany, France and Denmark had no such competitive difficulties. The conflicting monetary policy objectives within the ERM stimulated market speculation on a realignment. (Soros, notably, gambled massively, and lost the better part of his 1992 speculative gains). Eventually, the fluctuation bands were widened—not just so as to avoid a realignment but also to make life more difficult for speculators later on.

Figure 2: The movement of the forward premium



countries. Whatever the reason, danger signals occur much more rarely in subsample 2 and the fallen-angel effect should be hard to document.

Panel A of Table 8 summarizes the empirical results from the cubic models for the whole sample and the two subsamples. The slopes in the first subsample are quite close to the slopes in the whole sample. But in the second subsample, there are more egregious numbers, whether in the coefficients β_1 or the higher-order coefficients β_2 and β_3 . We find similar results later, when the regressor is the filtered forward premium or the long-memory component: in the post-93 period, the regressors show very little variation, and as a result, the estimated coefficients are erratic.

4.4 Robustness test on the career-risk effect

The inverse U-shaped betas generated by the cubic model are consistent with the fallen-angel hypothesis that the traders tend to shun long positions when facing danger signals. Alternatively, however, such selling behavior can also be explained by a common behavioral bias, extrapolative expectations: following a realized depreciation of the foreign currency the traders predict a further depreciation and step out, unless there is a hefty risk premium. The realized depreciation may very well be followed by a change in the forward premium, so we might have

Table 8: The cubic models with different sample periods and various regressors

Key:The data are weekly observations, and the whole sample period is from January 1st, 1976 to December 31th, 1998. Subsample 1 contains the observation before August 3rd, 1993, and subsample 2 after this date.

Panel A: regressor is f											
$\tilde{s}_{t,\Delta} = \alpha + \beta_1 f_{t,\Delta} + \beta_2 f_{t,\Delta}^2 + \beta_3 f_{t,\Delta}^3 + \tilde{\epsilon}_{t,\Delta}$											
whole sample				subsample 1				subsample 2			
	β_1	β_2	β_3	β_1	β_2	β_3	$Wald$	β_1	β_2	β_3	$Wald$
ATS	-0.11	***138.02	-1.00e+4	-0.07	135.19	-1.00e+4	***4.74	-0.48	636.92	3.16e+5	0.82
BEF	0.08	**38.60	-1.47e+3	-0.02	40.73	-1.42e+3	0.46	-2.73	678.63	-3.19e+4	1.35
DKK	***-0.61	***148.65	-6.35e+3	-0.84	144.10	-5.67e+3	***21.35	-4.36	23.42	4.89e+4	***93.40
FRF	-0.26	***142.65	-3.63e+3	-0.89	201.81	-5.17e+3	**3.82	4.24	-4.19e+3	7.74e+5	1.82
NLG	***-0.74	-5.16	1.50e+3	***-0.92	8.13	1.51e+3	0.26	-0.67	-276.67	2.20e+5	0.56
ITL	0.08	26.30	-1.06e+3	1.26	-75.59	1.01e+3	1.98	12.88	-1.78e+3	-1.41e+5	***29.98
ESP	-0.37	55.55	-251.97	-0.54	*59.96	-279.16	2.00	12.00	-4.00e+3	2.97e+5	***15.08
IEP	0.04	***50.40	-329.33	-0.06	**47.52	-236.02	***21.91	4.37	-4.19e+3	8.51e+5	***6.69
Panel B: regressor is ff											
$\tilde{s}_{t,\Delta} = \alpha + \beta_1 ff_{t,\Delta} + \beta_2 ff_{t,\Delta}^2 + \beta_3 ff_{t,\Delta}^3 + \tilde{\epsilon}_{t,\Delta}$											
whole sample				subsample 1				subsample 2			
	β_1	β_2	β_3	β_1	β_2	β_3	$Wald$	β_1	β_2	β_3	$Wald$
ATS	0.25	117.42	-1.26e+4	0.26	122.86	-1.31e+4	***8.19	-0.60	260.80	3.36e+5	0.42
BEF	-0.15	68.16	-1.99e+3	-0.03	59.66	-1.86e+3	1.50	*-1.98	469.89	-1.28e+4	2.09
DKK	-0.49	***124.63	-6.24e+3	-0.38	111.82	-6.23e+3	***21.06	***-3.69	-88.36	6.42e+4	***88.00
FRF	-0.64	**441.00	-1.87e+4	-0.57	*441.92	-1.73e+4	***15.84	-4.02	-2.01e+3	1.59e+6	***4.70
NLG	***-1.27	-68.98	-349.88	***-1.27	62.11	253.75	1.07	-0.88	-661.89	3.07e+5	1.81
ITL	**-1.20	110.14	-1.55e+3	-1.18	372.77	-1.51e+4	***6.33	***-1.91	-3.89e+3	-1.85e+5	***37.13
ESP	*0.36	**85.81	422.63	**0.45	**76.67	562.51	**3.54	-5.65	-3.33e+3	9.29e+5	***18.07
IEP	0.05	**49.37	-302.77	0.19	49.84	-338.40	***25.36	-1.91	-3.89e+3	-1.85e+5	***37.13
Panel C: regressor is \hat{f}											
$\tilde{s}_{t,\Delta} = \alpha + \beta_1 \hat{f}_{t,\Delta} + \beta_2 \hat{f}_{t,\Delta}^2 + \beta_3 \hat{f}_{t,\Delta}^3 + \tilde{\epsilon}_{t,\Delta}$											
whole sample				subsample 1				subsample 2			
	β_1	β_2	β_3	β_1	β_2	β_3	$Wald$	β_1	β_2	β_3	$Wald$
ATS	-1.27	1.82e+3	-5.23e+5	-1.13	1.68e+3	-4.85e+5	1.27	-0.30	-1.98e+4	-1.03e+8	1.48
BEF	-1.20	854.09	-1.00e+5	** -3.49	1.56e+3	-1.60e+5	***14.39	***36.50	***-9.32e+4	***-6.87e+7	**3.89
DKK	-2.25	**684.55	** -4.38e+4	***-5.32	1.26e+3	-7.56e+4	***6.49	***112.51	***-9.45e+4	***2.51e+7	***8.79
FRF	-2.39	*909.82	-6.20e+4	** -4.31	***1.34e+3	** -9.14e+4	**4.49	36.28	1.55e+4	-3.07e+6	***6.08
NLG	-1.16	***4.26e+3	***-1.84e+6	-1.58	**5.08e+3	***2.15e+6	***6.91	** -33.71	** -2.14e+5	-3.78e+8	***5.43
ITL	***-11.56	***1.56e+3	-5.71e+4	***-15.22	***1.91e+3	***-6.75e+4	***9.00	-389.22	1.66e+5	-2.31e+7	***7.57
ESP	***-10.75	***1.37e+3	-4.61e+4	***-14.63	***1.76e+3	***-5.83e+4	***12.91	** -232.38	**9.86e+4	-1.35e+7	***8.25
IEP	***-4.98	***1.48e+3	***-1.10e+5	***-8.93	***2.30e+3	***-1.63e+5	***9.09	***85.78	***-1.13e+5	***4.20e+7	***5.04

Table 9: **Robustness test of the Fallen-angel effect**

Key: “ $s_{t-\Delta,\Delta}$ ” and “ $s_{t-2\Delta,\Delta}$ ” are the exchange rate changes of the one-period and two-period lags respectively.

$$\tilde{s}_{t,\Delta} = c + \beta_1 f_{t,\Delta} + \beta_2 f_{t,\Delta}^2 + \beta_3 f_{t,\Delta}^3 + \rho_1 s_{t-\Delta,\Delta} + \rho_2 s_{t-2\Delta,\Delta} + \tilde{\varepsilon}_{t,\Delta}$$

	c	β_1	β_2	β_3	ρ_1	ρ_2
ATS	** 0.00	0.11	71.68	** -6206.56	*0.07	0.00
BEF	0.00	0.06	43.89	-1731.57	**0.12	-0.06
DKK	*0.00	** -0.58	***147.17	*** -6356.78	0.02	-0.01
FRF	0.00	-0.46	154.49	-3730.25	-0.03	0.09
NLG	***0.00	* -0.83	13.21	299.40	-0.04	-0.04
ITL	0.00	0.04	32.04	-1233.17	0.07	0.00
ESP	0.00	-0.64	66.31	-351.67	-0.06	0.06
IEP	0.00	0.03	***50.49	* -336.47	0.08	* -0.09

ascribed the effect of the past depreciation to the forward premium. To test this we add two lags of the dependent variable into the cubic model,

$$\tilde{s}_{t,\Delta} = c + \beta_1 f_{t,\Delta} + \beta_2 f_{t,\Delta}^2 + \beta_3 f_{t,\Delta}^3 + \rho_1 s_{t-\Delta,\Delta} + \rho_2 s_{t-2\Delta,\Delta} + \tilde{\varepsilon}_{t,\Delta}, \quad (4.17)$$

where $s_{t-\Delta,\Delta}$ and $s_{t-2\Delta,\Delta}$ are the exchange rate changes of the one-period and two-period lags respectively. If the traders react solely on the basis of extrapolative expectations, ρ_1 or ρ_2 might be significantly positive, or zero, or negative, depending on whether the expectations are correct, or irrelevant, or flatly wrong. More importantly, if the risk premium comes from extrapolation then the presumed fallen-angel effect would be weakened or perhaps even subsumed entirely.

Table 9 reports the estimates for Equation (4.17). The autoregressive coefficients ρ_1 and ρ_2 are insignificant and small in most cases, so the exchange rate change is not persistent, and any extrapolative behavior would be incorrect. More importantly, the quadratic coefficient β_2 remains positive for all the currencies, indicating the beta initially goes up in the forward premium; and the cubic coefficient β_3 stays negative except for NLG, showing that the inverse U-shaped pattern does not change. Therefore, the presumed fallen-angel explanation remains an acceptable explanation of the forward puzzle.

5 Filter out long-memory component in the forward premia

5.1 Long- and short-term components of forward premia

In this section we consider the near-nonstationarity of the forward premium. As mentioned before, non-stationarity in a variable invalidates the standard statistical model and makes

Table 10: **Fractional integration test**

Key: s denotes the spot return, f_Δ the forward premium for horizon Δ , \tilde{f} its H-P-filtered version, and ω^* is the estimated common trend in f . If d is larger than 0.5, the diagnosis is that the time series looks non-stationary, otherwise, it is viewed as stationary.

d	$\tilde{s}_{t,\Delta}$	$f_{t,\Delta}$	\tilde{f}	ω^*
ATS	0.109	0.544	-0.098	—
BEF	0.299	0.507	-0.312	—
DKK	0.257	0.608	-0.317	—
FRF	0.235	0.572	-0.201	—
NLG	0.109	0.561	-0.043	—
ESP	0.239	0.564	-0.294	—
IEP	0.184	0.581	-0.363	—
ITL	0.205	0.444	-0.310	—
ω^*	—	—	—	***-0.84

the usual t-statistics unreliable except when the dependent variable co-integrates with the independent variable. So we check the nonstationarity of the dependent variable and regressor with the Fractional Integration test (FI)¹⁰ in Table 10. Unsurprisingly, the exchange rate changes $\tilde{s}_{t,\Delta}$ are classified as stationary with $d < 0.5$ in all currencies. The forward premia $f_{t,\Delta}$, in contrast, seem non-stationary except those for the ITL. So, the traditional regression may be imbalanced: a stationary $\tilde{s}_{t,\Delta}$ can not cointegrate with a nonstationary $f_{t,\Delta}$.

Roll and Yan (2000) speculate on the likely source of the non-stationarity of the forward premium. They conceptually decompose the nominal interest rate differential into the real interest rate differential and the expected inflation differential. The real interest rate differential might be a unit-root process when there is an economically vast gap between countries, but that is not the case for the mainstream economies studied in most published tests. So, they conjecture that the persistence in the forward premium is due to a non-stationary inflation risk premium or a non-stationary expected inflation differential.

An alternative approach is to conceptually dissect the forward premium into the expected spot-rate change and the missing variable, possibly a currency-risk premium. In principle, any non-stationary expectations for exchange-rate changes should show up in the Fractional Integration tests on the realized changes. True, in finite samples there is an issue of power

¹⁰In the analysis of persistency of the time series, we usually consider the order of integration, d . When d is smaller than 0.5, the process is (weakly) stationary with finite variance. If d equals 0.5 or greater, the process is non-stationary with infinite variance. Infinite variance is more general than the unit-root case ($d = 1$). Here, we estimate d with Wavelet Ordinary Least Square (WOLS), see Tkacz (2001).

here: the variability of even a non-stationary expectation may still be small and hard to detect relative to the white-noise component in $\tilde{s}_{t,\Delta}$, and, therefore, be hard to establish beyond the usual doubt. In our case, there is a good *a priori* argument, though, that the apparent stationarity of \tilde{s} is not the result of low power: it is hard to imagine non-stationary expectations for percentage changes in exchange rates that all belong to the ERM.

If observed forward premia are non-stationary or close to it and if we reject unit roots for the expectations part, then the non-stationarity must come from the residual, the missing variable. Being likely to be a long-run process in the time domain or a low-frequency noise in the frequency domain, we estimate the long-memory component on the basis of the Hodrick-Prescott trends in each forward premia series. Appendix B provides the details of how the Hodrick-Prescott (HP) filtering technique produces the trend f_j^{tr} for every currency j . We find that the trends are quite similar across currencies, so much so that we actually treat them as a common trend, ω_t , computed as the average of the original trend series. The filtered premia, ff_j , are the residuals in the regression of the raw premia on the trend ω , and the long-term trends per currency, \hat{f}_j , are their fitted values.

If our inference that the trend component \hat{f}_j catches the nonstationarity of f_j is right, then its complement, the filtered premium $ff_{j,t}$, may be stationary. Table 10 compares the Fractional Integration parameter d of the filtered forward premia with the raw forward premia. In each currency $ff_{j,t}$ has a d smaller than 0.5, indicating that we seem to have a stationary series after filtering out the trend component from f . At the same time, the filtered premia should also be more independent from each other. From Panel D of Table 13, the correlations between the ff s are much lower than the correlations between the f s.

The fact that the (common) trend component is strongly long-memory leads to the conjecture that it contains little or no forecasts of exchange-rate changes; that is, the presence of this common trend may have been responsible for the poor predictive power of the forward premium—if at least the other component, the filtered part, does load more heavily on the expectations and not just on the transient part of the risk premium. This is what the tests of Section 5.2 are about.

5.2 *The predictive power of components of the forward premia*

To evaluate our conjecture on the forward-premium decomposition, we relate the exchange-rate change in turn to each of the two components. The regression with ff should generate higher β s with less of the inverse U-shaped pattern than those with \hat{f} . Panel B and Panel C in Table

Table 11: **Comparing the filtered premium or the common trend to the original forward premium in terms of goodness of fit.**

Key: Number of times the alternative right-hand-side variable beats the original one in terms of fit.

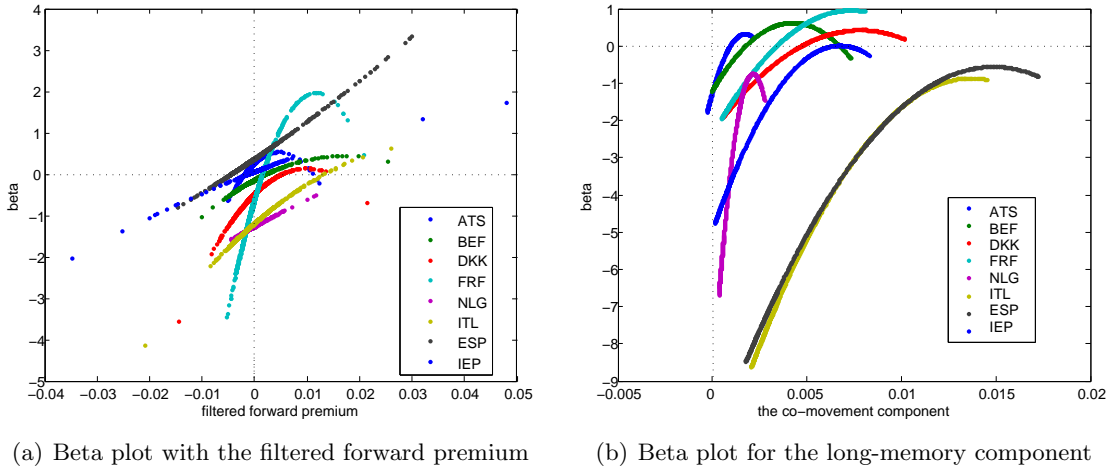
	$R_{ff}^2 \succ R_f^2$	$AIC_{ff} \succ AIC_f$	$SIC_{ff} \succ SIC_f$	$R_{\hat{f}}^2 \succ R_f^2$	$AIC_{\hat{f}} \succ AIC_f$	$SIC_{\hat{f}} \succ SIC_f$
whole sample	7	7	6	2	2	2
subsample 1	7	7	7	2	2	2
subsample 2	5	6	6	4	5	5

8 present the results.

Compared with the raw f in Panel A, the filtered forward premium ff produces higher coefficients β_1 and more significant Wald tests in the first subsample. In addition, ff produces less egregious slopes in the second subsample. For the whole sample, four regressions have higher coefficients β_1 with the ff , and more regressions (seven out of eight) reject the linear model specification via the significant Wald test. The long-memory component, in contrast, brings out much lower and negative coefficients in the whole sample and subsample 1, and many erratic coefficients β_1 in subsample 2. (In the second subsample, the lack of variability in the long-memory part mentioned above may also contribute to make the betas even worse.) The more negative β_1 s are consistent with the idea that the common component is related to the long-memory part of the risk premium; meanwhile, the more positive β_1 s with ff are consistent with the conjecture that the filtered forward premium ff loads more heavily on the expectation than \hat{f} . The cubic slope β_3 stays negative in most cases, implying an inverse U-shaped pattern of the beta as shown in Figure 3(a) and Figure 3(b). The inverse U-shaped pattern that we see in plots of the beta against ff again supports the fallen-angel hypothesis. But the pattern is far less pronounced than for the raw f . With \hat{f} as the regressor, in contrast, the pattern is very clear, whether statistically or graphically.

Table 11 assesses the goodness-of-fit — \bar{R}^2 , AIC and SIC — when the cubic model is run on different specifications of the forward premia. All three measures indicate a better fit for the filtered forward premia than for the raw forward premia: seven out of eight regressions show an improvement in the whole sample and the first subsample; six regressions gain in terms of AIC and SIC in the second subsample. In contrast, the long-memory component does worse than the original regressor with six regressions worsening in the whole sample and subsample 1, and with mixed results in subsample 2. So, the model evaluation is consistent with our conjecture

Figure 3: Beta plots in terms of the components



that the short-term filtered forward premium loads on the expectation (but probably also on the transient part of the risk premium), while the long-memory component does a bad job as a predictor, has a strong inverse-U pattern expected from the risk premium, and is, therefore, closer to the missing variable.

6 Conclusions

This paper is related to two issues in the forward-puzzle literature: non-linearities in the relation between \tilde{s} and f , and non-stationarity in the forward premium.

One idea we bring into this literature is career risk considerations among money managers: to go against publicly observed danger signals, one needs either guts or extra expected returns. In an exploratory test we find that danger signals other than the forward premium do seem to be followed by extra excess returns. The next question is whether the forward premium may also be doubling as a danger signal, in which case the Fama beta should be allowed to be a non-linear function or a spline function of the forward premium. We do find that a nonlinear relationship is more proper to describe the relation between the percentage change in the spot rate and the forward premium. Both the cubic and spline regressions outperform the linear model in terms of goodness-of-fit. The cubic model produces an inverse U-shaped beta that rises from very negative to a higher level as the forward premium approaches zero and dips again as the forward premium increases further. This pattern again fits in with the fallen-angel hypothesis that traders or portfolio managers shun long positions in assets with danger signals.

Our second contribution is the decomposition of the highly persistent forward premium

into a long-memory component and a filtered forward premium that is definitely not unit-root. The former component associates with a strong common trend. It seems to load heavily on the missing regressor which, in Fama's conjecture, is responsible for the forward puzzle: when we run the cubic models with the long-memory variable on the right-hand side, betas are much lower than for the total- f regressions and the filtered- ff regressions. The shape of the betas even provides a clue as to what the long-memory part stands for: it could again be a fallen-angel risk premium that makes traders skeptical about assets with danger signals. The filtered forward premia, in contrast, seems to load on both the expectations and the risk premium: the betas are generally positive and do a better job indicating the future change, but also exhibit a (weaker) inverse U-shaped pattern that one expects from career-risk effects.

References

- Backus, D. and Smith G., 1993. Consumption and Real Exchange Rates in Dynamic Economies with Non-traded Goods, *Journal of International Economics* **35**, 297-316.
- Baillie, R. T., and Kiliç, 2006. Do Asymmetric and Nonlinear Adjustments Explain the Forward Premium Anomaly?, *Journal of International Money and Finance* **25**, 22-47.
- Baillie, R. T., and Bollerslev, T., 2000. The Forward Premium Anomaly is Not as Bad as You Think, *Journal of International Money and Finance* **19(4)**, 471-488.
- Bansal, R., 1997. An Exploration of the Forward Premium Puzzle in Currency Market, *the Review of Financial Studies* **10**, 369-403.
- Bilson, J. F. O., 1981. The Speculative Efficiency Hypothesis, *Journal of Business* **54**, 435-451.
- Brenner R. J. and Kroner K. F., 1995. Arbitrage, Conintegration, and Testing the Unbiasedness Hypothesis in Financial Markets, *The journal of Financial and Quantitative Analysis* **30(1)**, 23-42.
- Cumby, R. and Obstfeld, M., 1984. International Interest Rate and Price Level Linkages under Flexible Exchange Rates: A Review of Recent Evidence, in J. Bilson and R. Marston eds. *Exchange Rate Theory and Practice*, University of Chicago Press, Chicago, IL.
- Domowitz, I. and Hakkio, C.S., 1985. Conditional Variance and the Risk Premium in the Foreign Exchange Market, *Journal of International Economics* **19**, 47-66.
- Engel, R. F., Lilien, D. M., and Robins, R.P., 1987. Estimating time-varying risk premia in the term structure: the ARCH-M model, *Econometrica* **55**, 319-407.
- Fama, E., 1984. Forward and Spot Exchange Rates, *Journal of Monetary Economics* **14**, 319-338.
- Huisman, R., Koedijk, Kee., Kool, C. and Nissen, F., 1998. Extreme support for uncovered interest parity, *Journal of International Money and Finance* **17**, 211-228.
- Ikenberry, D., Lakonishok, J., and Vermaelen, T., 1995. Market Underreaction to Open Market Share Repurchases, *Journal of Financial Economics* **39**, 181-208.
- Lothian, J. R., Wu, L., 2005. Uncovered Interest Rate Parity over the Past Two Centuries, *working paper*.
- Lyons, R. K., 2001. *The Microstructure Approach to Exchange Rates*, (MIT Press).
- Mark, N. C. and Moh, Y. K., 2002. Continuous-Time Market Dynamics, ARCH Effect, and the Forward Premium Anomaly, *working paper SSRN*.

- Mark, N.C. and Y. Wu, 1997. Risk, Policy Rules, and Noise: Rethinking Deviations from Uncovered Interest Parity, Tinbergen Institute Discussion Papers 97-041/2, Tinbergen Institute.
- Ramberg, R., Dudewicz, E.J., Tadikamalla, P.R., Mykytka, E.F., 1979. A probability Distribution and its Uses in Fitting Data, *Technometrics*, **21(2)**, 201-214.
- Obstfeld, M. and Rogoff, K., 2000. New Directions for Stochastic Open Economy Models, *Journal of International Economics* **50(1)**, 117-153.
- Roll, R. and Yan, Sh., 2000. An Explanation of the Forward Premium Puzzle, *European Financial Management* **6(2)**, 121-148.
- Sarno, L. Valente, G. and Leon, H., 2006. Nonlinearity in Deviations from Uncovered Interest Parity: An Explanation of the Forward Bias Puzzle, *CEPR Discussion Papers*, 5527.
- Sercu, P., Vandebroek, M., and Wu, X., 2008. Is the Forward Bias Economically Small? Evidence from European Rates, *Journal of International Money and Finance*, forthcoming.
- Sercu, P. and Vinaimont, T., 2006. The Forward Bias in the ECU: Peso Risks *v.* Fads & Fashions, *Journal of Banking and Finance*, **30**, 2409-2432.
- Tai, Chu-Sheng, 2001. A multivariate GARCH in mean approach to testing uncovered interest parity: Evidence from Asia-Pacific foreign exchange markets, *The Quarterly Review of Economics and Finance*, **41**, 441-460.
- Tkacz, G., 2001. Estimating the Fractional Order of Integration of Interest Rates Using a Wavelet OLS Estimator, *Studies in Nonlinear Dynamics and Econometrics*, **5 (1)**, Article 2.
- Villanueva, O. M., 2005. FX Dynamics, Limited Participation, and the Forward Bias Anomaly, *The Financial Review* **40**, 67-93.

A Principal Component Analysis of the Danger Signals

We select the following seven variables related to the divergence indicator and to recent trends in the spot rate over one day:

1. Position in the ERM band

The European Exchange Rate Mechanism (ERM) was built around the ECU, a basket of all EU currencies. For each currency there was a target value in the basket, called central parity. Whenever the actual value of the ECU moved too far from the central parity against the ECU, the member state had to take “policy measures” to bring back its currency into line. The divergence indicator (D) provides the official signal for such measures. It was published every day in all major newspapers, and was calculated as the divergence between the actual value and central parity of the ECU in units of home currency, as a percentage of the allowed maximum divergence,¹¹

$$D := \frac{[\text{actual value} - \text{central parity}]/\text{central parity}}{\text{maximum divergence}}. \quad (1.18)$$

¹¹The maximum divergence depends on the currency’s weight in the ECU. Against the DEM, for example, the ECU cannot drop as far as against the IEP, since 30 percent of the the ECU was DEM and only 1 percent was IEP.

A positive D means a strong ECU, that is, a weak home currency and therefore, under the career-risk hypothesis, a risk premium for holding it. Since the exchange-rate data in the tests are units of home currency per DEM, the positive career-risk risk premium on home currency translates into a negative risk premium for holding DEM. That is, the relation between our risk premium and D should be negative.

The longer the time the divergence indicators has been positive, the worse is the signal. Two versions of the D are considered that measure the persistence of the D , namely, the one-week and one-month moving averages of D , denoted as $\bar{D}_{w,t}$ and $\bar{D}_{m,t}$:

$$\begin{aligned}\bar{D}_{w,t} &:= \sum_{l=1}^7 D_{t-l}/7, \\ \bar{D}_{m,t} &:= \sum_{l=1}^{30} D_{t-l}/30.\end{aligned}\tag{1.19}$$

2. Change of the position within the band

A weakening of home currency against the ECU could be another danger signal, over and above the level of the D . We look at the change over the last 24 hours:

$$D_{t,-1} = D_t - D_{t-1}.\tag{1.20}$$

A positive change means a strengthening ECU, that is, a weakening home currency. Like a positive level of the ECU, a positive change should therefore get a negative sign in the risk premium on the DEM.

3. The change of the bilateral exchange rate for the DEM.

The variables in the first two groups are in terms of the basket, the ECU. The divergence indicator is, however, a relatively weak constraint in the sense that it looks at the average deviation of the member currencies from their central parities, not the highest pairwise deviation; and it may trigger “policy measures” like interest-rate changes but not intervention in the exchange market. Any such intervention was instead based on the bilateral rates, which should stay within the $\pm 2.25\%$ band ($\pm 6\%$ for the ITL and ESP). We look at the bilateral rate that is most likely to be troublesome, for weak currencies: the value of the DEM. Changes in that rate are measured in three time spans: daily, $(s_{t,-1})$, weekly $(s_{t,-7})$, and monthly $(s_{t,-30})$:

$$\begin{aligned}s_{t,-1} &= \ln S_t - \ln S_{t-1}, \\ s_{t,-7} &= \ln S_t - \ln S_{t-7}, \\ s_{t,-30} &= \ln S_t - \ln S_{t-30}.\end{aligned}\tag{1.21}$$

Table 12: **Weights of the risk variables in the principal components ξ_1 , ξ_2 and ξ_3 , and explanatory power**

Key: Seven potential danger signals are compressed into three principal components: divergence (current, and one-week and one-month moving average, denoted D , \bar{D}_w and \bar{D}_m), change in the bilateral log rate of the DEM (over one day, one week, one month, denoted s_{-1} , s_{-7} and s_{-30}), and change in the divergence (D_{-1}). Data are daily observations with the DEM as base currency, for overlapping monthly returns in excess of the forward premium. The sample period is from April 2nd, 1979 till August 3rd, 1993.

$$\begin{aligned}\xi_1 &= \kappa_{11}D + \kappa_{12}\bar{D}_{w,t} + \kappa_{13}\bar{D}_{m,t} + \kappa_{14}s_{t,-30} + \kappa_{15}s_{t,-1} + \kappa_{16}s_{t,-7} + \kappa_{17}D_{t,-1} \\ \xi_2 &= \kappa_{21}D + \kappa_{22}\bar{D}_{w,t} + \kappa_{23}\bar{D}_{m,t} + \kappa_{24}s_{t,-30} + \kappa_{25}s_{t,-1} + \kappa_{26}s_{t,-7} + \kappa_{27}D_{t,-1} \\ \xi_3 &= \kappa_{31}D + \kappa_{32}\bar{D}_{w,t} + \kappa_{33}\bar{D}_{m,t} + \kappa_{34}s_{t,-30} + \kappa_{35}s_{t,-1} + \kappa_{36}s_{t,-7} + \kappa_{37}D_{t,-1}\end{aligned}$$

		weights κ							variance explained
		D	$\bar{D}_{w,t}$	$\bar{D}_{m,t}$	$s_{t,-1}$	$s_{t,-7}$	$s_{t,-30}$	$D_{t,-1}$	
BEF	ξ_1	-0.58	-0.58	-0.58	-0.02	-0.02	-0.03	-0.03	0.43
	ξ_2	0.02	0.03	0.03	-0.53	-0.62	-0.56	-0.14	0.26
	ξ_3	-0.05	0.04	0.03	-0.09	0.15	0.18	-0.97	0.14
DKK	ξ_1	-0.58	-0.58	-0.58	0.02	0.03	0.01	-0.02	0.42
	ξ_2	-0.04	-0.00	-0.01	-0.53	-0.61	-0.53	-0.24	0.26
	ξ_3	-0.04	0.05	0.02	-0.11	0.17	0.33	-0.92	0.14
FRF	ξ_1	-0.57	-0.57	-0.57	0.05	0.09	0.14	-0.02	0.43
	ξ_2	-0.09	-0.10	-0.09	-0.52	-0.64	-0.54	-0.03	0.25
	ξ_3	-0.04	0.06	0.04	-0.11	0.06	0.08	-0.99	0.14
NLG	ξ_1	-0.58	-0.58	-0.57	-0.01	-0.02	-0.06	-0.03	0.42
	ξ_2	0.02	0.04	0.04	-0.54	-0.61	-0.55	-0.17	0.26
	ξ_3	-0.06	0.06	0.03	0.07	0.11	0.12	-0.98	0.14
ITL	ξ_1	-0.57	-0.57	-0.57	-0.05	-0.08	-0.10	-0.02	0.43
	ξ_2	0.08	0.08	0.08	-0.53	-0.63	-0.55	-0.04	0.26
	ξ_3	-0.04	0.05	0.03	0.15	-0.05	0.00	-0.99	0.14
IEP	ξ_1	-0.58	-0.58	-0.57	0.02	0.04	0.05	-0.03	0.42
	ξ_2	-0.03	-0.04	-0.04	-0.52	-0.64	-0.57	-0.01	0.26
	ξ_3	-0.06	0.07	0.04	0.03	0.00	-0.01	-0.99	0.14
ESP	ξ_1	-0.50	-0.49	-0.49	0.17	0.28	0.39	-0.09	0.51
	ξ_2	-0.14	-0.26	-0.24	-0.63	-0.38	-0.14	0.54	0.23
	ξ_3	-0.21	-0.15	-0.18	0.10	-0.58	-0.43	-0.61	0.13

A rise in the DEM's value means a danger signal and, therefore, a lower risk premium for the DEM.

In short, under the career-risk view all seven potential danger signals should be negatively correlated with the risk premium on the DEM.

The Principal Component Analysis reduces the seven variables into three principal components, ξ_1 , ξ_2 and ξ_3 . From Table 12, three components account for more than 80% of the total variance; the first principal component ξ_1 explains around 40% of the variance and can be summarized as the position factor, for in this linear combination the variables D , $\bar{D}_{w,t}$ and

$\bar{D}_{m,t}$ have dominant weights relative to the other variables; the second component ξ_2 can be viewed as the bilateral change factor because of the prominent weights for $s_{t,-1}$, $s_{t,-7}$ and $s_{t,-30}$. The third component ξ_3 relates to the change in the position, primarily weighted by $D_{t,-1}$. As can be seen from the weights of seven variables in the table, all the principal components negatively correlate with their dominant variables, so if danger signals lead to risk premia we expect the slope coefficients in the regression of returns on the signals.

B Hodrick-Prescott Decomposition of the Premia

The Hodrick-Prescott (HP) filter is a standard instrument to capture the trend in time series. The HP-filter works as follows. A series f_j for currency j is to be decomposed into “trend” and “cycle” components,

$$f_{j,t} = f_{j,t}^{tr} + f_{j,t}^c. \quad (2.22)$$

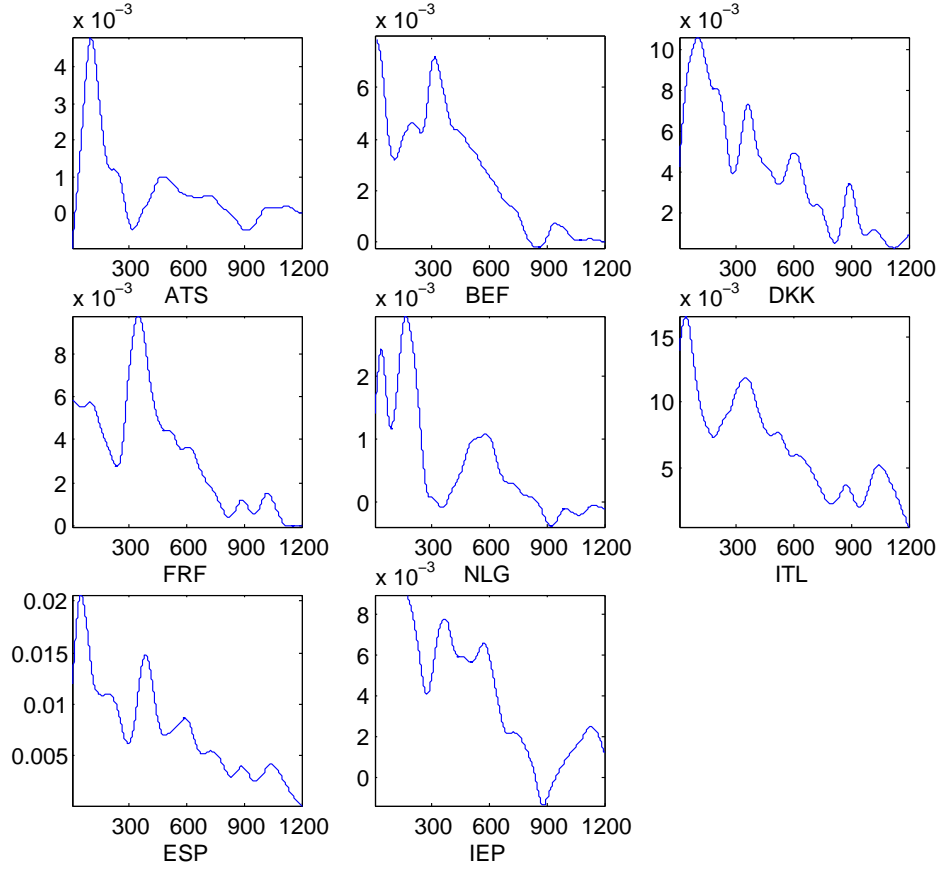
The HP filter estimates the trend from the solution to the following minimization problem with a pre-set smoothing parameter λ :

$$\min_{f_{j,t}^{tr}, \forall t} \sum_{t=1}^T (f_{j,t} - f_{j,t}^{tr})^2 + \lambda \sum_{t=1}^T \left[\underbrace{(f_{j,t+1}^{tr} - f_{j,t}^{tr}) - (f_{j,t}^{tr} - f_{j,t-1}^{tr})}_{\text{second difference in the trend}} \right]^2. \quad (2.23)$$

This objective function consists of two parts: the squared sum of the deviations of the series from its trend, summing up the badness-of-fit of the trend series vis-a-vis the raw data; and the sum of the squared trend reversals (the difference between two consecutive first differences), measuring the smoothness of the trend series. λ , the smoothing parameter, weighs the components in the objective. For quarterly data, for instance, it is conventionally set at 1600; we use the default value in Eviews 5.0 for weekly data, $\lambda = 270400$.

Figure 4 shows the eight trends. A strong similarity of the patterns is immediately obvious: forward premia against DEM are generally falling over time, in line with the general level of interest rates and, towards the end, converging to zero when the single monetary policy under the authority of the ECB approaches. The existence of important links between the eight trend series is confirmed when we compute correlation coefficients. Panel B of Table 13 shows the correlations between the eight HP trend components. Most of the values are above 0.60, much higher than the correlation coefficients between the raw forward premia in Panel A. Thus, series-by-series trends that consist of the sluggish (or, tentatively, long-memory) components

Figure 4: The HP trends in individual currencies



of the premia are behaving quite similarly, suggesting that they load more heavily on one underlying common trend.

Because a strong similarity is found between the individual trends f_j^{tr} , we assume there is a common trend, ω_t^* , underlying the individual trend variables. The motivation for using one common trend is that the individual trends may be over-fitting their series: by retaining only the common part, an additional filter is administered. Our proxy for the common component ω_t^* is calculated as a weighted average of f_j^{tr} across currencies. We first scale each trend series f_j^{tr} by its mean, \bar{f}_j , and we then average across the currencies j :

$$\omega_t := \frac{\sum_{j=1}^J \frac{f_{j,t}^{tr}}{\bar{f}_j}}{J}. \quad (2.24)$$

Secondly, because our data set has a small cross-section (only eight currencies), we truncate the input data: only the values within the two-standard-deviation interval around the medians

Table 13: Correlations of f or components (eight currencies against dem)

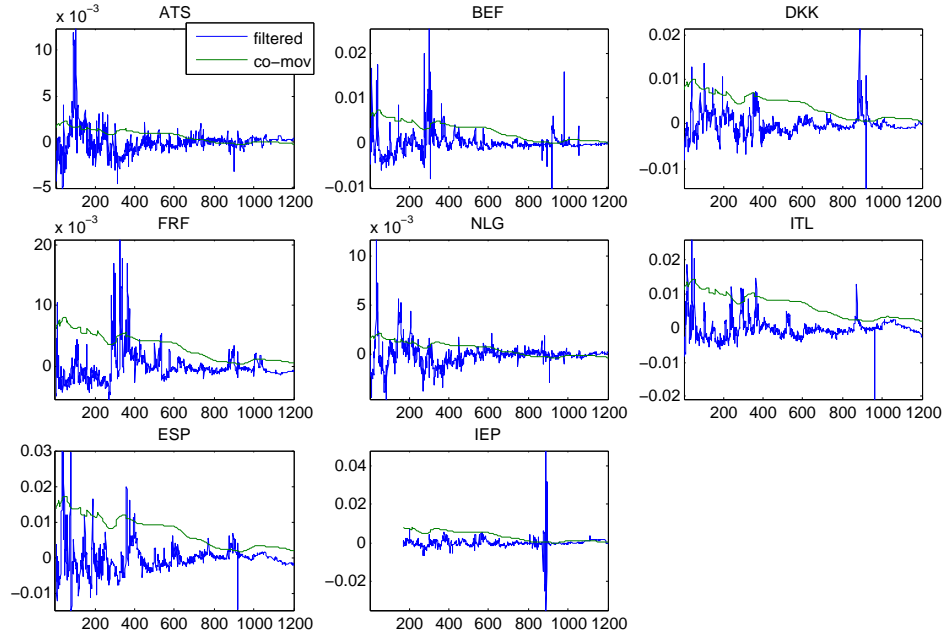
	ATS	BEF	DKK	FRF	NLG	ESP	IEP	ITL
Panel A: Raw forward premia, $\text{corr}(f_j, f_k)$								
ATS	1							
BEF	0.139	1						
DKK	0.328	0.561	1					
FRF	0.111	0.681	0.611	1				
NLG	0.551	0.313	0.420	0.179	1			
ESP	0.304	0.455	0.657	0.557	0.362	1		
IEP	0.277	0.469	0.393	0.453	0.387	0.449	1	
ITL	0.269	0.684	0.602	0.792	0.266	0.601	0.450	1
Panel B: HP trends, $\text{corr}(f_j^{tr}, f_k^{tr})$								
ATS	1							
BEF	0.264	1						
DKK	0.736	0.742	1					
FRF	0.293	0.897	0.739	1				
NLG	0.666	0.548	0.767	0.354	1			
ESP	0.582	0.790	0.887	0.767	0.682	1		
IEP	0.626	0.860	0.826	0.789	0.672	0.858	1	
ITL	0.420	0.913	0.795	0.860	0.547	0.908	0.827	1
Panel C: $\text{corr}(\omega^*, f_j)$								
correlation coefficient	ATS	BEF	DKK	FRF	NLG	ESP	IEP	ITL
	0.61	0.89	0.91	0.82	0.78	0.93	0.96	0.92
Panel D: HP-filtered premia, $\text{corr}(ff_j, ff_k)$								
ATS	1							
BEF	0.118	1						
DKK	0.257	0.395	1					
FRF	0.131	0.302	0.346	1				
NLG	0.150	0.357	0.298	0.039	1			
ESP	0.065	0.220	0.269	0.044	0.380	1		
IEP	0.065	0.071	-0.065	0.058	0.067	-0.027	1	
ITL	0.227	0.340	0.351	0.367	0.189	0.128	0.003	1

are included in the average calculations.

$$\omega_t^* := \frac{\sum_{j=1}^J I_{j,t}^{2\sigma} \frac{f_{j,t}^{tr}}{f_j}}{\sum_{j=1}^J I_{j,t}^{2\sigma}}, \quad (2.25)$$

where $I_{j,t}^{2\sigma}$ indicates whether the (j, t) -th trend observation is within two standard deviations from its median. So, when the value is outside the two- σ interval, it has no weight in the average.

The better the individual forward premia f_j correlate with the common trend ω^* , the more co-movement it captures. Panel C of Table 13 shows high correlation coefficients between these two variables, from 0.61 to 0.96, indicating that there are lots of commonalities and that ω_t^*

Figure 5: **The filtered forward premia and their trends**

works well as a proxy for the eight original HP trends. ω_t^* also has long memory: the unit root can not be rejected by Dicky Fuller test shown in Table 10. Having obtained a single long-memory variable for all currencies under the HP filter method, we lastly decompose the eight series of f_j into a part that is linearly related to the common trend, and an orthogonal component:

$$\begin{bmatrix} f_1 \\ f_2 \\ \dots \\ f_J \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \dots \\ \mu_J \end{bmatrix} + \begin{bmatrix} \phi_1 \\ \phi_2 \\ \dots \\ \phi_J \end{bmatrix} \times \omega_t^* + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \dots \\ \epsilon_J \end{bmatrix}. \quad (2.26)$$

In the main text the fitted values $\hat{f}_{j,t} := \mu_j + \phi_j \omega_t$ are referred to as the long-memory component; and the retrieved residuals ϵ_j are henceforth referred to as the filtered forward premia, denoted as ff_j , and expressed as $ff_{j,t} := f_{j,t} - \hat{f}_{j,t}$. Figure 5 shows graphs of the filtered premia, for comparison, along with the time series plots of the long-memory components \hat{f}_j underlying ω in Figure 4. By design, while for these trend components—and, by implication, the common trend—there is a lot of smoothness and inertia, in Figure 5 the random component for the ff_j is quite strong. A second difference between \hat{f}_j and ff_j is that most of time the plots of the filtered premia are below the plots of the long-memory components. Third, the long-memory component \hat{f}_j has a narrower standard deviation than the filtered forward premium ff_j . For the FRF, this can also be seen from the standard deviations of f , \hat{f} and ff in Table 5.

Applying this procedure for each forward premia series, we obtain a currency-specific trend series f_j^{tr} for every currency j .

C Monte-Carlo Standard Deviations

The remaining issue is the reliability (SE) of the estimations. There are two complications. First, the forward premium (and even more so the long-memory component) is non-stationary or nearly so. Second, following Hansen and Hodrick (1980) we do not want to waste information by considering only non-overlapping forward contracts, so we use the weekly observations on one-month forward contracts. For either reason, the conventional standard deviation is underestimated. In this paper Monte Carlo (MC) simulation is employed to calculate the standard deviation.

Monte Carlo method is a stochastic technique that uses random numbers and probability statistics rather than math to discover the distribution of parameter estimates. Variables on both sides of the classical UEH model are expressed in their Moving Average (MA) models. According to the correlogram test, we find that for most currencies the first six lags are significant,¹² so the exchange rate change and the forward premium are expressed as AR(6) models:

$$s_{j,t} = \alpha_1 + \beta_1 s_{j,t-1} + \beta_2 s_{j,t-2} + \dots + \beta_6 s_{j,t-6} + \nu_{j,t}, \quad (3.27)$$

$$f_{j,t} = \alpha_2 + \theta_1 f_{j,t-1} + \theta_2 f_{j,t-2} + \dots + \theta_6 f_{j,t-6} + \xi_{j,t}. \quad (3.28)$$

It turns out that the residuals ν and ξ are non-normally distributed. Ramberg, Dudewicz, Tadikamalla and Mykytka (1979) provide a technique for a non-normal distribution number generator. This technique accommodates a broad class of distributions because it transforms a uniform random number into distribution with any desired set of values for the first four statistical moments (mean, variance, skewness and kurtosis). These four moments, denoted below as μ_1, μ_2, μ_3 and μ_4 , are functions of four parameters $\lambda_1, \lambda_2, \lambda_3$ and λ_4 , as described in

¹²Remember that these are overlapping weekly observations of one-month returns and premia.

the following equations:

$$\mu_1 = \lambda_1 + \frac{A}{\lambda_2}, \quad (3.29)$$

$$\mu_2 = \frac{B - A^2}{\lambda_2^2}, \quad (3.30)$$

$$\mu_3 = \frac{C - 3AB + 2A^3}{\lambda_2^3}, \quad (3.31)$$

$$\mu_4 = \frac{D - 4AC + 6A^2B - 3A^4}{\lambda_2^4}. \quad (3.32)$$

In these equations, the terms A , B , C and D are also functions of $\lambda_1, \lambda_2, \lambda_3$ and λ_4 :

$$A = \frac{1}{1 + \lambda_3} - \frac{1}{1 + \lambda_4}, \quad (3.33)$$

$$B = \frac{1}{1 + 2\lambda_3} + \frac{1}{1 + 2\lambda_4} - 2\mathcal{B}(1 + \lambda_3, 1 + \lambda_4), \quad (3.34)$$

$$C = \frac{1}{1 + 3\lambda_3} - \frac{1}{1 + 3\lambda_4} - 3\mathcal{B}(1 + 2\lambda_3, 1 + \lambda_4) + 3\mathcal{B}(1 + \lambda_3, 1 + 2\lambda_4), \quad (3.35)$$

$$D = \frac{1}{1 + 4\lambda_3} + \frac{1}{1 + 4\lambda_4} - 4\mathcal{B}(1 + 3\lambda_3, 1 + \lambda_4) + 6\mathcal{B}(1 + 2\lambda_3, 1 + 2\lambda_4) - 4\mathcal{B}(1 + \lambda_3, 1 + 3\lambda_4), \quad (3.36)$$

where $\mathcal{B}(u, v)$ is the beta function. To generate the residuals we estimate their first four moments and numerically solve for the corresponding values of the λ 's. The desired non-normal random number \tilde{R} is the following transformation of a unit uniform random number \tilde{p} :

$$R(\tilde{p}; \lambda) = \lambda_1 + \frac{\tilde{p}^{\lambda_3} - (1 - \tilde{p})^{\lambda_4}}{\lambda_2}. \quad (3.37)$$

To test the null of no relation between s and f we generate 1000 numbers with the properties of the observed s_j , and, independently of that, 1000 numbers with the properties of the observed f_j . (The actual number of observations in the real-world sample is between 1030 and 1200.) We then run on this data set all regressions we study in this paper and compute all betas we are interested in. We repeat this 1000 times to get an idea of how the regression output is distributed under the null.