KATHOLIEKE
UNIVERSITEIT
LEUVEN

DEPARTEMENT TOEGEPASTE ECONOMISCHE WETENSCHAPPEN

RESEARCH REPORT 0212

'EXTREME SUPPORT FOR UIP' REVISITED: HOW COMES THE DOGS DON'T BARK?

by

P. SERCU
M. VANDEBROEK

D/2002/2376/12

"Extreme Support For UIP" revisited: how comes the dogs don't bark?*

Piet Sercu[†] and Martina Vandebroek[‡]

First draft: March 2001; this version: February 2002
Preliminary — comments very welcome

*We thank Jef Teugels for providing a hitherto unpublished proof, and Yuri Goegebeur for asking pertinent questions and answering some of ours. Thanks also to Geert Dhaene, Tom Vinaimont and participants at a KU Leuven workshop for useful comments and criticisms. All remaining errors are ours.

[†]KU Leuven, Graduate School of Business Studies, Naamsestraat 69, B- 3000 Leuven; Tel: +32 16 32 6756; Fax: +32 16 32 6632; email: piet.sercu@econ.kuleuven.ac.be.

[‡]KU Leuven, Graduate School of Business Studies, Naamsestraat 69, B- 3000 Leuven; Tel: +32 16 32 6975; Fax: +32 16 32 6632; email: martina.vandebroek@econ.kuleuven.ac.be.

Abstract

Conjecturing that, in testing UIP, transaction costs may have obscured the relation between expected exchange-rate changes and forward premia, Huisman *et al.* (1998) focus on days with unusually large cross-sectional variances in forward premia (reflecting, assumedly, episodes with pronounced expectations). They find encouragingly high regression coefficients for those special days—"extreme" support, in short.

We show that, for extreme forward premia to be primarily due to a clear signal rather than loud noise, the signal needs to be thicker-tailed than the noise. Transaction-cost-induced noise seems to have promising properties: percentage deviations from the perfect-markets equilibrium should be (*i*) bounded (that is, they have no tails and, therefore, cannot dominate the extreme forward premia), (*ii*) wide (that is, they may generate betas below 1/2) and (*iii*) U-shaped in distribution, a feature that turns out to make an "extreme" sample quite effective. We derive theoretical and numerical results in the direction of what Huisman *et al.* observe.

JEL F31, G14, G15

"Extreme Support For UIP" revisited: How comes the dogs don't bark?

Introduction

Since the work by Cumby and Obstfeld (1984) and Fama (1984) (henceforth COF), regressions of realized percentage exchange rate changes on beginning-of-the-period forward premia have become the workhorse for tests of uncovered interest parity (UIP). In this paper, we analyze recent results by Huisman *et al.* (1998). They apply panel estimation, singling out days where forward premia have an unusually large cross-sectional variance—hence the "extreme" in the title—and they impose a numeraire-invariance constraint across currencies. As a result, Huisman *et al.* report, the COF slope coefficients improve substantially, to the extent that they even exceed unity when the cross-sectional variance is extreme. Huisman *et al.*'s justification for focusing on large-variance days is that, on those days, exchange-rate changes should tend to have unusually heterogeneous conditional expectations, thus providing a sample with a better signal-to-noise ratio. In their paper, the noise is caused by transaction costs, but the authors do not offer any reason why their logic would not apply to any other factor that is missing from the unbiased-expectations equation, like a risk premium or a genuine market inefficiency.

Articles that provide good news *re* COF regressions are still sufficiently thin on the ground to command attention. The main issues addressed in the present paper stem from the finding by Fama (1984) that most of the variability of the forward premium originates from the missing variable rather than from the conditional expectation. This immediately raises the question why, in the Huisman *et al.* extreme sample, the dogs did not bark: if one selects a subsample where the variance of the forward premia is large, how comes one is not implicitly picking up mostly loud-noise observations rather than clear-signal ones? Our first result is that the Huisman *et al.* attempt to reinforce the signal will work indeed if the signal (the conditional expected change in the exchange rate), despite being the lower-variance variable, is the thicker-tailed of the two. Our second issue then becomes whether the transaction-cost explanation advanced by Huisman *et al.* has the potential to generate a material improvement in the signal-to-noise ratio. The answer is a guarded yes. Interestingly, the mechanism that triggers the improvement turns out to be subtler than just boosting the signal variance. Extreme sampling can, in fact, also lead to a noise reduction in absolute terms rather than just relative to the signal variance. In addition, extreme sampling induces a covariance between noise and signal that further helps eliminating the bias. Thus, extreme sampling is potentially quite effective in dealing with transaction-cost-induced noise in the forward rate—even though the actual extent of success still depends very much on the parameter constellation. In the process of establishing these analytical results, we derive and use new moment conditions to show that the missing variable has not only a high variance and relatively thin tails, but also an uncannily high correlation with the expected exchange-rate change. Thus, unlike the textbook errors-in-variables case, our entire analysis is done in a setting where the error in the regressor is correlated with the true value.

In the remainder of this introductory section we briefly review the literature on the COF regression tests of uncovered interest parity (UIP), and we make a link with the Huisman *et. al.* paper. The empirical failure of UIP in these tests has led to two lines of subsequent research. Some researchers have argued that OLS may be inappropriate or at least inefficient, and have used more advanced estimation techniques. Others have worked on the theoretical side, and have studied what properties the missing variables should have to explain the empirical findings, and what theoretically acceptable model(s) do have these features. Among the missing variables that have been advanced, the (non-constant) risk premium has received most attention, starting with Fama (1984). Others have claimed that the variable that is missing in the empirical tests is a (non-constant) Peso effect, that is, the low-probability jump that is at the back of the market's mind but is rarely, if ever, observed in the data. Still others make a link with transaction costs, which disturb the normal link between the forward premium and the true expectation of the spot-rate change (or, in the presence of a risk premium, its true certainty equivalent). Lastly, for completeness, the missing factor may be the difference between the market's expectation and the true expectation, resulting from an inefficiency.

Huisman *et al.* (1998) contribute to the methodology side (by adopting panel estimation with random time effects and a cross-equation constraint, and by conditioning the COF coefficients on the cross-sectional variation of forward premia), but they also tap into the theory literature, by linking their approach to a particular missing variable. Their prime suspect is transaction costs. If real-world markets are subject to friction, they argue, uncovered interest arbitrage cannot perfectly align expected exchange rates and forward premia. Most of the time, expectations of exchange rate changes are, moreover, so small 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 COF 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 variance of the regressor is 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 COF regression coefficients close to unity, and even substantially above unity if the definition of "large variance" is very strict.

In this paper, we first review the link between the COF beta and its three underlying moments—the variances of the expected change and of the bias, and the covariance between the two—and we derive additional bounds on these, tighter than the Fama (1984) moment condition. In Section 2, we study the effect of sampling from the tails, first analytically and then numerically. Section 3 concludes.

1 Moment Conditions for Missing Variables

Let $\tilde{S}^*$ denote the true expected value, at the beginning of period t and conditional on all then-available information, of the percentage change in the spot rate over that period; let $\tilde{\nu} \stackrel{def}{=} \tilde{S} - \tilde{S}^*$ denote the unexpected spot rate change, and $\tilde{P}$ the forward premium. Let the missing variable or bias (whether it be a risk premium, Peso effect, an inefficiency, or a transaction-cost effect) be denoted and defined by $\tilde{b} \stackrel{def}{=} \tilde{S}^* - \tilde{P}$, so that its sign is that of a conventional risk premium. For simplicity, we omit time subscripts. It is understood that $\tilde{S}^*$, $\tilde{P}$, and $\tilde{b}$ are conditionally non-stochastic; that is, the randomness indicated by their tildes holds unconditionally only. By definition, the prediction error $\tilde{\nu}$ is conditionally (and therefore unconditionally) independent of $\tilde{S}^*$ and $\tilde{b}$, as the latter two are known at the beginning of the period. Then, in the regression of $\tilde{S} \stackrel{def}{=} \tilde{S}^* + \tilde{\nu}$ on $\tilde{P} \stackrel{def}{=} \tilde{S}^* - \tilde{b}$, the slope is determined by the three second moments of $\tilde{S}^*$ and $\tilde{b}$:

$$
\begin{aligned}
\beta &= \frac{\mathrm{cov}(\tilde{S}^*, \tilde{P})}{\mathrm{var}(\tilde{P})} \\
&= \frac{\mathrm{var}(\tilde{S}^*) - \mathrm{cov}(\tilde{S}^*, \tilde{b})}{\mathrm{var}(\tilde{S}^*) - 2\mathrm{cov}(\tilde{S}^*, \tilde{b}) + \mathrm{var}(\tilde{b})} \\
&= \frac{1-w}{1-2w+z}, \qquad (1.1)
\end{aligned}
$$

where, for compactness, the last line uses relative moments, $z \overset{def}{=} \frac{\text{var}(\tilde{b})}{\text{var}(\tilde{S}^*)}$ and $w \overset{def}{=} \frac{\text{cov}(\tilde{S}^*,\tilde{b})}{\text{var}(\tilde{S}^*)}$. With respect to the relative variance, z, Fama(1984) presents a simple but insightful moment condition,

$$\beta \lesseqqgtr \frac{1}{2} \Leftrightarrow z \gtreqqless 1. \tag{1.2}$$

As empirical betas tend to be below 1/2, Fama points out, the missing variable must have a large variance relative to the variability in the expectations—that is, we must have $z > 1$ irrespective of sign or size of the relative covariance, w. It is simple to derive similar bounds on w irrespective of z, and one on w relative to z. It also turns out that a bound with respect to z can be obtained that is tighter than Fama's, simply from the fact that correlations are, at most, equal to unity:

Proposition 1: If $\text{var}(\tilde{b}) > 0$, $\text{var}(\tilde{S}^*) > 0$, and $R^2(\tilde{S}^*, \tilde{b}) < 1$, then

$$\beta \lesseqqgtr 0 \Leftrightarrow w \gtreqqless 1; \tag{1.3}$$

$$\beta \lesseqqgtr 1 \Leftrightarrow w \lesseqqgtr z; \tag{1.4}$$

$$\frac{1}{2} < \frac{1+\sqrt{z}}{1+2\sqrt{z}+z} < \beta < \frac{1-\sqrt{z}}{1-2\sqrt{z}+z} \text{ if } z < 1, \tag{1.5}$$

$$\frac{1-\sqrt{z}}{1-2\sqrt{z}+z} < \beta < \frac{1+\sqrt{z}}{1+2\sqrt{z}+z} < \frac{1}{2} \text{ if } z > 1 \tag{1.6}$$

Proof: provided in Appendix.

Figure 1 shows the feasible values for β, that is, the area bounded by (1.5)-(1.6), for $0 \leq z \leq 7$. On the basis of the Froot and Thaler (1990) meta-average of β, -0.88, one would conclude that the variance ratio z is, at most, 4.5. In addition, one can infer from (1.3) and (1.4) that the covariance-to-variance ratio, w, is somewhere between z and unity. Actually, one can infer even more about the correlation between $\tilde{b}$ and $\tilde{S}^*$, by generalizing (1.5)-(1.6) into

$$\frac{1+\sqrt{z.R^2_{max}}}{1+2\sqrt{z.R^2_{max}}+z} < \beta < \frac{1-\sqrt{z.R^2_{max}}}{1-2\sqrt{z.R^2_{max}}+z} \text{ if } z \leq 1, \tag{1.7}$$

$$\frac{1-\sqrt{z.R^2_{max}}}{1-2\sqrt{z.R^2_{max}}+z} < \beta < \frac{1+\sqrt{z.R^2_{max}}}{1+2\sqrt{z.R^2_{max}}+z} \text{ if } z > 1. \tag{1.8}$$

where R^2_{max} is a tentative upper bound on the squared correlation between $\tilde{b}$ and $\tilde{S}^*$. Suppose, for instance, that one deems R^2_{max} to be at most 0.9. Then the range of z that produces betas of -0.88 is narrowed down to about [1.5 , 3.1], as can be seen from Figure 2. Interestingly, below $R^2_{max} = 0.87$ we cannot even produce any betas as low as -0.88. Thus, if we take the

Froot and Thaler average as our benchmark, then we need to consider a noise-to-signal ratio, z, between 1 to 4.5, and a correlation between $\tilde{b}$ and $\tilde{S}^*$ of at least $\sqrt{0.87} = 0.93$,[1]. surely a disconcertingly high number. The economic issue (not addressed here) is to explain a missing variable that is almost perfectly correlated with the conditional expectation *and* has a larger variance, so that it essentially flips the sign of the signal. For the statistician, the implication is that the covariance between the regressand and the error in the regressor is large relative to the variances. This complicates the issue relative to the standard textbook version of the errors-in-variables problem, where the error is assumed to be pure noise.

What are the statistical implications of that covariance? Holding constant the covariance, it is easy to show that, like in a classical textbook error-in-variables case, the variance of $\tilde{b}$ still biases β towards zero, and that boosting the variance of the signal weakens that bias:

$$\begin{aligned}\frac{\partial\beta}{\partial \mathrm{var}(\tilde{b})} &= -\frac{\mathrm{var}(\tilde{S}^*) - \mathrm{cov}(\tilde{S}^*,\tilde{b})}{[\mathrm{var}(\tilde{S}^*) - 2\mathrm{cov}(\tilde{S}^*,\tilde{b}) + \mathrm{var}(\tilde{b})]^2}, \\ &= -\frac{\beta}{\mathrm{var}(\tilde{S}^*) - 2\mathrm{cov}(\tilde{S}^*,\tilde{b}) + \mathrm{var}(\tilde{b})} \gtreqless 0 \text{ if } w \gtreqless 1 \Leftrightarrow \beta \lesseqgtr 0, \end{aligned} \tag{1.9}$$

$$\frac{\partial\beta}{\partial \mathrm{var}(\tilde{S}^*)} = \frac{\mathrm{var}(\tilde{b}) - \mathrm{cov}(\tilde{S}^*,\tilde{b})}{[\mathrm{var}(\tilde{S}^*) - 2\mathrm{cov}(\tilde{S}^*,\tilde{b}) + \mathrm{var}(\tilde{b})]^2} \gtreqless 0 \text{ if } w \lesseqgtr z \Leftrightarrow \beta \lesseqgtr 1. \tag{1.10}$$

Boosting the variance of the signal is the stated intention of sampling extreme forward premia. What needs to be determined is (i) how comes that the Huisman *et al.* procedure of picking days with large-variance forward premia seems to be able to produce mainly large-variance signals $\tilde{S}^*$ rather than mainly large-variance $\tilde{b}$s; and (ii) what is the effect of that sampling procedure on the covariance between $\tilde{S}^*$ and $\tilde{b}$? The *cet. par.* effect of that covariance is to bias β towards $1/2$:

$$\frac{\partial\beta}{\partial \mathrm{cov}(\tilde{S}^*,\tilde{b})} = \frac{\mathrm{var}(\tilde{b}) - \mathrm{var}(\tilde{S}^*)}{[\mathrm{var}(\tilde{S}^*) - 2\mathrm{cov}(\tilde{S}^*,\tilde{b}) + \mathrm{var}(\tilde{b})]^2} \gtreqless 0 \text{ if } z \gtreqless 1 \Leftrightarrow \beta \lesseqgtr \frac{1}{2}. \tag{1.11}$$

In the next sections, we therefore review the impact of a sampling-from-the-tails procedure on each of these three moments.

[1]This result can also be obtained directly from fixing beta at -0.88 and writing (1.1) as $-0.88 = \frac{1-\rho\sqrt{z}}{1-2\rho\sqrt{z+2)}}$

2 The effects of sampling from the tails on the (co)variances of $\tilde{b}$ and $\tilde{S}^*$

Huisman *et al.*, who use panel data, select days where the cross-sectional variation in $\tilde{P}$ is large, and expect that in such a subsample the COF beta will be large. In the same spirit, Sercu and Vinaimont (1999) follow the original Bilson (1983) procedure and select, within each time series, the observations where $\mid \tilde{P} \mid$ is large. In either sampling rule the issue is which items on the right-hand-side of

$$\mathrm{var}(\tilde{P}) = \mathrm{var}(\tilde{S}^*) - 2\mathrm{cov}(\tilde{S}^*, \tilde{b}) + \mathrm{var}(\tilde{b}) \tag{2.12}$$

go up most when one exclusively samples from the tails of the distribution of $\tilde{P}$. In this section, we address the issue analytically. Section 3 then provides numerical results.

2.1 Boosting var($\tilde{S}^*$) or var($\tilde{b}$): a matter of relative tail-thickness rather than variance

If one samples extreme value of $\mid \tilde{P} \mid$, why should this primarily be due to extreme values of $\tilde{S}^*$ (as Huisman *et. al.* postulate) rather than of $\tilde{b}$, especially since the latter seems to be the higher-variance summand? It turns out that the predominance of large $\tilde{S}^*$s versus large $\tilde{b}$s in the tails of $\tilde{P}$ depends not so much on the relative variance of the marginal distributions, but primarily on the type of distribution—specifically, the relative tail-thickness of $\tilde{b}$ and $\tilde{S}^*$. We provide an illustration of each of these claims.

First consider two variables with similar distributions but different variances. Specifically, let $\tilde{b}$ and $\tilde{S}^*$ be bivariate normal with constant moments. From the definition $\tilde{P} \stackrel{def}{=} \tilde{S}^* - \tilde{b}$ it follows that $\tilde{S}^*$ and $\tilde{P}$ are also bivariate normal. Thus, for any given value of $\tilde{P}$, whether small or extreme, the best possible conditional forecast of the exchange rate is $\mathrm{E}(\tilde{S}^* \mid P) = a + \beta P$. For this reason, in the bivariate-normal case extreme sampling cannot generate any betas that systematically differ from aselect sampling. This must mean that, in the large-$\mid \tilde{P} \mid$ sample, all second moments of $\tilde{S}^*$ and $\tilde{b}$ have gone up by the same factor: in the bivariate-normal case, sampling extreme Ps does not affect the signal-to-noise ratio. We obtain this result without making any assumption about the relative variance of $\tilde{b}$ and $\tilde{S}^*$. Thus, for a variable to dominate the tails of the sum, a larger variance is surely not sufficient.

Nor is a high variance necessary for that, as the next example shows. Let $\tilde{b}$ be uniform, and $\tilde{S}^*$ normal, both with mean zero and independent of each other. The further one goes into the tails of the sum, the larger its conditional variance. But while there is no bound on the $\tilde{S}^*$-component, in this case there is a limit to what $\tilde{b}$ can add to the sum. That is, even though a sample of extreme $|\ \tilde{P}\ |$s also tends to pick up atypically large $|\ \tilde{b}\ |$s, in the uniform case the conditional variance of $\tilde{b}$ cannot go on rising indefinitely when one goes deeper and deeper into the tail to $|\ \tilde{P}\ |$.

In this second example, a crucial difference between the two distributions seems to be boundedness (for the uniform) versus unboundedness (for the normal). We can generalize, however: even if both $\tilde{S}^*$ and $\tilde{b}$ can assume any value on the real line, in the tails $\tilde{S}^*$ still dominates provided that it has thicker tails. The proposition also generalizes in the sense that it allows for linear dependence between the two summands:

Proposition 2. Let $\tilde{S}^*$ and $\tilde{b}$ be related by $\tilde{S}^* = \mu_S + \theta.\tilde{b} + \tilde{\epsilon}$ with $\mathrm{E}(\tilde{\epsilon} \mid b) = 0$. Then the tails of the sum, $\tilde{P} \stackrel{def}{=} \tilde{S}^* - \tilde{b}$, are dominated by large values $\tilde{S}^*$ rather than of $\tilde{b}$ if

$$\lim_{x \uparrow \infty} \frac{P(\tilde{S}^* \geq x)}{P(\tilde{b} \geq x)} \to 0, \tag{2.13}$$

a sufficient condition for which is that $\tilde{S}^*$ has the lower tail exponent.

Proof: provided in Appendix.

2.2 The $(\pm\delta)$ model for $\tilde{b}$.

The same effects would be observed if $\tilde{b}$ were a bid-ask bounce generated by a Bernouilli process,

$$\tilde{b} = -\delta + 2\delta.\tilde{B}, \text{ where } \tilde{B} = \begin{cases} 1 \ (i.e.\ \tilde{b} = +\delta) \text{ with probability } 0.5, \\ \\ 0 \ (i.e.\ \tilde{b} = -\delta) \text{ with probability } 0.5. \end{cases} \tag{2.14}$$

If $\tilde{S}^*$ and $\tilde{b}$ are both symmetric around zero, then also $\tilde{P}$ is symmetric around zero. Thus, large $|\ \tilde{P}\ |$s would still have as many positive outcomes as negative ones; and the underlying $\tilde{b}$s, being equally likely to be positive or negative, would have the same variance as in an aselect sample. Thus, in this case any increase of $\mathrm{var}(\tilde{P})$ generated by sampling from the tails is not at all due to bunching of noise; rather, it must be entirely due to large $\tilde{S}^*$s sampled from a thicker-tailed distribution, and reduces the bias-toward-zero in β.

Is (2.14) an economically attractive model? The pure bid-ask bounce model for $\tilde{b}$ would seem to make sense if the only friction in the market were a spread in the forward rate, as is the case with futures transaction prices. Assuming that the expectation is measured by the midpoint forward, the observed forward trade price would be the midpoint price plus or minus the half-spread, a perturbation which, at daily frequencies, has no traces of autocorrelation

(Lehman, 1990; Ball, Kothari and Wasley, 1995). However, if one's purpose is to explain the forward puzzle then bid-ask bounce is not a good candidate. In actual practice the forward premia as used in empirical work do not suffer from bid-ask bounce, because they are typically computed from midpoint swap rates or from domestic and foreign interest rates—both midpoint or both ask (like LIBOR). And more fundamentally, in the presence of friction it is no longer obvious that the midpoint forward premium equals the expectation even in the absence of a risk premium.

There are, fortunately, good reasons to believe that the impact of friction is much richer than just a bid-ask-bounce effect. Models with trading costs predict a no-activity zone within which the bias $\tilde{P} - \tilde{S}^*$ can wander without triggering transactions. Thus, $\tilde{b}$ is bounded and therefore probably has a smaller tail exponent than has $\tilde{S}^*$—a feature that is necessary to explain the Huisman *et al.* effect. In addition, in these models the zone is much wider than just the transaction cost. Brennan and Schwarz (1988, 1990) and Baldwin (1990), for instance, point out that the holder of, say, GBP has an American-style perpetual option to switch to, say, USD. It is well known that an American option is rationally exercised not as soon as the exercise value becomes positive, but when the exercise value is sufficiently large. In the same vein, if trading is costly, risk-neutral holders of GBP will not switch to USD as soon as $\tilde{S}^* - \tilde{P}$ exceeds the transaction cost by a minute amount; rather, since there is a cost of switching back if and when the gain disappears, it is optimal to wait until $\tilde{S}^* - \tilde{P}$ has become sufficiently large. The fact that the inactivity zone is wide justifies large values for $\tilde{b}$, which is required if we want to explain betas below 0.5 (Fama, 1984). Transaction cost models of this type however also predict that, while the deviation from UIP stays at the inactivity bounds for relatively long times (see e.g. Constantinides, 1986; Dumas, 1992), the probability of being strictly within the band is of course not zero (as it would have been in the $(\pm\delta)$ case). Thus, from the Brennan-Schwarz-Baldwin argument, the distribution of $\tilde{b}$ would be bounded, but U-shaped rather than having just two separate probability spikes at $\pm\delta$.

As we do not know the specific functional form of the U-distribution for $\tilde{b}$, we initially work with the Bernouilli-based distribution, where analytical results are quite transparent; and we then verify numerically that our analytical results do generalize to a U-shaped distribution. In that analytical work we immediately add another feature to the model, namely correlation between $\tilde{b}$ and $\tilde{S}^*$. In the binary case, (2.14), the relationship $\mathrm{E}(\tilde{S}^* \mid b)$ can always be written as a linear equation. In short, for most of the remaining analytical work we assume that

$$\tilde{S}^* = \mu_S + \theta.\tilde{b} + \tilde{\epsilon}, \text{ with } \begin{cases} \mathrm{E}(\tilde{\epsilon} \mid b) = 0, \\ \tilde{\epsilon} \text{ is IID, finite-variance, symmetric and unimodal,} \\ \tilde{b} = \begin{cases} +\delta \text{ with probability } 0.5, \\ -\delta \text{ with probability } 0.5. \end{cases} \\ 0 < \theta < 1, \end{cases} \tag{2.15}$$

The constraint $0 < \theta \overset{def}{=} \mathrm{cov}(\tilde{S}^*, \tilde{b})/\mathrm{var}(\tilde{b}) = w/z < 1$ is motivated by (1.3)-(1.4) and the empirical observation that $\beta < 0$. Graphically, this works as follows. If $b = 0$, the joint distribution of period-by-period expectations and forward premia P plots on the 45-degree ray: in the absence of a bias, both variables are of course identical. When $\tilde{b} = \pm\delta$, the S^* that is associated with a particular P is shifted rightward/downward or leftward/upward relative to the 45-degree ray. The parameter θ tells us how much of the shift is vertical versus horizontal. In particular, when $\theta = 0$ the shift is all horizontal, which corresponds to the pure textbook errors-in-variables case. When $\theta = 1$, in contrast, the shift is all vertical, in which case $\tilde{b}$ is similar, in effect, from a pure white-noise prediction error. The situation relevant to us is somewhere in between.

As a matter of notation, we let "hi" and "lo" (as subscripts or as conditioning events in a conditional distribution) refer to the events "$\mid \tilde{P} \mid \geq \chi$" and "$\mid \tilde{P} \mid < \chi$", respectively, where χ is the percentile value for $\mid \tilde{P} \mid$ that produces a desired split of the sample, like the 5% most extreme forward premia. For example, $\mathrm{E}(\tilde{b} \mid hi)$ is shorthand for $\mathrm{E}(\tilde{b} \mid \mid \tilde{P} \mid \leq \chi)$, and β_{hi} refers to a beta from a sample consisting solely of extreme $\mid \tilde{P} \mid$s.

In model (2.15), we can use the equalities $\tilde{S}^* = \mu_S + \theta.\tilde{b} + \tilde{\epsilon}$ and $\tilde{P} = \tilde{S}^* - \tilde{b}$ to specify β_{hi} and the regular (unconditional) beta as, respectively,

$$\begin{aligned} \beta_{hi} &= \frac{\theta.(\theta-1)\mathrm{var}(\tilde{b} \mid hi) + (2\theta-1)\mathrm{cov}(\tilde{b}, \tilde{\epsilon} \mid hi) + \mathrm{var}(\tilde{\epsilon} \mid hi)}{(\theta-1)^2\mathrm{var}(\tilde{b} \mid hi) + 2(\theta-1)\mathrm{cov}(\tilde{b}, \tilde{\epsilon} \mid hi) + \mathrm{var}(\epsilon \mid \tilde{h}i)} \\ &= 1 - \frac{(1-\theta)\mathrm{var}(\tilde{b} \mid hi) - \mathrm{cov}(\tilde{b}, \tilde{\epsilon} \mid hi)}{(\theta-1)^2\mathrm{var}(\tilde{b} \mid hi) + 2(\theta-1)\mathrm{cov}(\tilde{b}, \tilde{\epsilon} \mid hi) + \mathrm{var}(\tilde{\epsilon} \mid hi)}, \text{ and} \end{aligned} \tag{2.16}$$

$$\beta = 1 - \frac{(1-\theta)\mathrm{var}(\tilde{b})}{(1-\theta)^2\mathrm{var}(\tilde{b}) + \mathrm{var}(\tilde{\epsilon})}. \tag{2.17}$$

Thus, again, the bias in the marginal beta falls if the noise variance is reduced relative to the variance of the premium.

Let us recapitulate. We found that the signal/noise ratio improves if $\tilde{b}$ has thinner tails (or no tails), which is quite likely under the transaction-cost story. In fact, the apparent success of the procedure, in Huisman *et al.*, suggests that we do have that ordering of tail-thicknesses. We have also proposed one simple model, (2.15), that has some of the crucial features one would expect from a transaction-cost model: boundedness and a large probability mass on or near the bounds. The model is able to generate a wide grid of w and z values. A weakness is that (2.15) assumes away any probability mass strictly within the bounds; but in Section 2.5 we show numerically that such mass has only a small effect on the outcome. Thus, for the above model we now present our results on the items $\text{var}(\tilde{b} \mid hi)$ and $\text{cov}(\tilde{b} \mid hi)$ in (2.16), respectively. From the simplified model we show, first, that if the exchange rates have zero unconditional drift—or, more in general, if extreme samples are equally likely to come from either tail—then the variance of the noise is not affected by extreme sampling. Thus, the signal-to-noise ratio improves, which probably was the hunch underlying the Huisman *et al.* paper. This orthodox effect is reinforced by two less obvious ones. One effect is that, if the exchange rates have nonzero unconditional drift—or, more in general, if extreme samples are more likely to come from one particular tail—then the variance of the noise actually drops in absolute terms. The second reinforcing effect is that the covariance term in (2.16) is negative.

2.3 The effect of extreme sampling on $\text{var}(\tilde{b} \mid hi)$.

Proposition 3. In model (2.15),

- if $\mu_S = 0$ (or, more generally, if upper and lower tails of S^* are equally present in the extreme sample), then $\text{var}(\tilde{b} \mid hi) = \text{var}(\tilde{b}) = \text{var}(\tilde{b} \mid lo)$.
- if $\mu_S \neq 0$ (or, more generally, if upper and lower tails of S^* are *not* equally present in the extreme sample), then $\text{var}(\tilde{b} \mid hi) < \text{var}(\tilde{b})$ and $\text{var}(\tilde{b} \mid lo) < \text{var}(\tilde{b})$.

  For $\text{var}(\tilde{b} \mid hi)$, this effect increases the higher the cut-off value χ. For $\text{var}(\tilde{b} \mid lo)$, the effect is U-shaped in χ.

Proof: provided in Appendix.

Table 1 illustrates this effect for a simple distribution. In each panel of that table, $\tilde{S}^*$ and $\tilde{b}$ are independent and have the marginal distributions indicated in the top row and first column, respectively, of the panel. Panel A assumes a zero mean for $\tilde{S}^*$, while in Panels B and C the expectation of $\tilde{S}^*$ equals unity. In panels A and B, "high" outcomes are defined as $\mid \tilde{S}^* - \tilde{b} \mid \geq 3$, while in panel C the cut-off value is 4 rather than 3. The body of each panel

Table 1: Examples of a simple joint distribution of S^* and b

Panel A: $\mu_S = 0$, cutoff at $\chi = \pm 3$					
	$S^* = -2$ prob 0.1	$S^* = -1$ prob 0.2	$S^* = 0$ prob 0.4	$S^* = 1$ prob 0.2	$S^* = 2$ prob 0.1
$b = +2$, prob 1/2	*-4 (0.05)*	*-3 (0.10)*	-2 (0.20)	-1 (0.10)	0 (0.05)
$b = -2$, prob 1/2	0 (0.05)	1 (0.10)	2 (0.20)	*3 (0.10)*	*4 (0.05)*
Panel B: $\mu_S = 1$, cutoff at $\chi = \pm 3$					
	$S^* = -1$ prob 0.1	$S^* = 0$ prob 0.2	$S^* = 1$ prob 0.4	$S^* = 2$ prob 0.2	$S^* = 3$ prob 0.1
$b = +2$, prob 1/2	*-3 (0.05)*	-2 (0.10)	-1 (0.20)	0 (0.10)	1 (0.05)
$b = -2$, prob 1/2	1 (0.05)	2 (0.10)	*3 (0.20)*	*4 (0.10)*	*5 (0.05)*
Panel C: $\mu_S = 1$, cutoff at $\chi = \pm 4$					
	$S^* = -1$ prob 0.1	$S^* = 0$ prob 0.2	$S^* = 1$ prob 0.4	$S^* = 2$ prob 0.2	$S^* = 3$ prob 0.1
$b = +2$, prob 1/2	-3 (0.05)	-2 (0.10)	-1 (0.20)	0 (0.10)	1 (0.05)
$b = -2$, prob 1/2	1 (0.05)	2 (0.10)	3 (0.20)	*4 (0.10)*	*5 (0.05)*

Key to Table 1. In each panel of that table, $\tilde{S}^*$ and $\tilde{b}$ are mutually independent and have the marginal distributions indicated in the top row and first column, respectively, of the panel. Panel A assumes a zero mean for $\tilde{S}^*$, while in Panels B and C the expectation of $\tilde{S}^*$ equals unity. In panels A and B, "high" outcomes are defined as $\mid \tilde{S}^* - \tilde{b} \mid \geq 3$, while in panel C the cut-off value is 4 rather than 3. The body of each panel shows the premia and the corresponding joint probabilities. If a cell falls in the *hi* subsample, the probability is printed in italics.

shows the premia and the corresponding joint probabilities. If a cell falls in the *hi* subsample, the probability is printed in italics. We easily see that, in Panel A, the distribution of $\tilde{b}$ given a *hi* event remains symmetric, thus preserving a variance of $4.2^2.(1/2)(1-1/2) = 4$. In Panel B, in contrast, the *hi* events consist predominantly of outcomes where $\tilde{b}$ equals $-\delta$: given a high $\mid \tilde{P} \mid$, the probability of $\tilde{b} = +\delta$ now drops from 0.5 to $0.05/0.40 = 0.125$. Thus, $\text{var}(\tilde{b} \mid hi)$ drops from 4 to $4.2^2.(0.125 \times 0.875) = 1.75$.[2]

More in general, if $\mu_S > 0$, the distribution of $\tilde{P}$ has more mass in the positive domain, so that large $\tilde{P}$s tend to be positive rather than negative. From the relation $\tilde{P} \stackrel{def}{=} \tilde{S}^* - \tilde{b}$, this means a predominance of negative $\tilde{b}$s in the high-$\mid \tilde{P} \mid$ sample, and vice versa. Likewise, if $\mu_S < 0$, $\tilde{P}$ tends to be negative and large $\mid \tilde{P} \mid$s tend to be associated with positive values for $\tilde{b}$. If sampling is thus predominantly from one side of the distribution of $\tilde{b}$, then $\pi(\tilde{b} = -\delta \mid hi)$ no longer equals the unconditional probability, 0.5. Any deviation from $\pi = 0.5$ lowers the

[2]The table may raise the issue that asymmetric sampling not only lowers the variance of the noise, but also the variance of the entire sum and, at first blush, potentially also the variance of the expectation. However, Proposition 2 would still mean that the thicker-tailed variable tends to dominate in the tails. Case C in the table, where the cut-off values are set at ±4 rather than ±3, illustrates this very clearly.

variance, as can be seen from $\text{var}(\tilde{b} \mid hi) = 4\delta^2.\pi.(1-\pi)$. Thus, if $\mu_S \neq 0$, the error variance is reduced. If, in addition, δ is large and the "*hi*" criterion is sufficiently strict, the sample of high $\mid \tilde{P} \mid$s comes almost entirely from events where $\tilde{b}$ has the same sign. For example, in Panel C of Table 1, raising the hurdle $\mid \tilde{P} \mid$ from 3 to 4 means that all *hi* events are now generated by $\tilde{b} = -2$. The closer one gets to such situations, the smaller $\text{var}(\tilde{b} \mid hi)$, implying that β_{hi} approaches unity.

2.4 The Effect of extreme sampling on $\text{cov}(\tilde{S}^*, \tilde{b} \mid hi)$

Unconditionally, $\tilde{\epsilon}$ is independent of $\tilde{b}$. But as $\tilde{P}$ depends on $\tilde{b}$ and $\tilde{S}^*$, and $\tilde{S}^*$ depends on $\tilde{b}$ and $\tilde{\epsilon}$, selecting a sample on the basis of $\mid \tilde{P} \mid$ means that sampling is not random with respect to $\tilde{\epsilon}$ and $\tilde{b}$.

Proposition 4. in model (2.15), $\text{cov}(\tilde{\epsilon}, \tilde{b} \mid hi) < 0$.

Proof: provided in Appendix.

The intuition follows immediately from the equations $\tilde{P} \stackrel{def}{=} \tilde{S}^* - \tilde{b}$ and $\tilde{S}^* = \mu_S + \tilde{b} + \tilde{\epsilon}$ where $\theta < 1$. Then $\tilde{P} = \mu_S + \tilde{\epsilon} - (1-\theta)\tilde{b}$ with $1 - \theta > 0$. As $\tilde{b}$ can assume only two values, namely $\pm\delta$, a large value of $\mid \tilde{P} \mid$ is not due to an unusually large $\mid \tilde{b} \mid$; rather, a large $\mid \tilde{P} \mid$ tends to mean either a large value of $\mid \tilde{\epsilon} \mid$, or opposite signs for $\tilde{\epsilon}$ and $\tilde{b}$, or both. Thus, conditional on $\mid \tilde{P} \mid$ being large, we expect that $\tilde{\epsilon}$ and $(1-\theta)\tilde{b}$ to be of the opposite sign more often than what would expected on the basis of the zero unconditional covariance.

This effect is immediately visible in the example of Table 1. In the zero-mean example of Panel A, cases where $\tilde{P}$ falls below -3 occur when $\tilde{S}^*$ is negative and $\tilde{b}$ positive; and cases where $\tilde{P}$ exceeds 3 are found where $\tilde{S}^*$ is positive and $\tilde{b}$ negative. With $\text{E}(\tilde{b} \mid hi)$ and $\text{E}(\tilde{S}^* \mid hi)$ both being zero and every term in $\tilde{S}^*.\tilde{b}$ being negative, the covariance is obviously negative, and turns out to be -2.667. In Panel B, the unconditional mean is of $\tilde{S}^*$ is positive rather than zero, which, as we have seen, makes $\text{var}(\tilde{b} \mid hi)$ drop (from 4 to 1.75). Not surprisingly, then also $\text{cov}(\tilde{b}, \tilde{S}^* \mid hi)$ then shrinks towards zero, but it remains negative (at -1.125).

Thus, if the cut-off values are not set symmetrically around the unconditional mean, one might theoretically succeed in getting a sample that is entirely from one of the two 45-degree lines only. The Huisman *et al.* procedure may actually produce such asymmetric samples: since their definition of "*hi*" is not currency-specific, they may very well end up with lots of lower-tail observations from weak currencies, and lots of upper-tail observations from strong ones. As we saw, such asymmetric samples are predominantly drawn from just one 45-degree

Table 2: Parameter values used in the simulations

	z = 1.1	z=4	z=7	z=10
$\beta = 0.2,$	w = 0.967, var($\tilde{P}$) = .167 θ= 0.88	w = 0, var($\tilde{P}$) = 5.000 θ= 0.00	w = -1, var($\tilde{P}$) = 6.000 θ= -0.14	w = -2, var($\tilde{P}$) = 7.000 θ= -0.20
$\beta =$ -0.2	w = 1.014 var($\tilde{P}$) = .072 θ= 0.92	w = 1.429 var($\tilde{P}$) = 2.142 θ= 0.36	w = 1.857 var($\tilde{P}$) = 4.286 θ= 0.28	w = 2.286 var($\tilde{P}$) = 7.286 θ=0.23
$\beta =$ -0.6	w = 1.027 var($\tilde{P}$) = .046 θ= 0.93	w = 1.818 var($\tilde{P}$) = 1.364 θ=0.46	w = 2.636 var($\tilde{P}$) = 2.728 θ= 0.38	(no solution)
$\beta = \beta_{min}(z)$	($\beta = -1$) w=1.033 var($\tilde{P}$) = 0.033 θ= 0.94	β_{min}= -0.997, w =1.999 var($\tilde{P}$) = 1.002 θ= 0.50	β_{min} = -0.607, w = 2.644 var($\tilde{P}$) = 2.712 θ= 0.38	β_{min} = -0.462, w = 3.161 var($\tilde{P}$) = 4.678 θ=0.32

Key to Table 2: From the preset grid $z = \{1.1, 4.\ 7.\ 10\}$ and $\beta = \{0.2, \text{-}0.2, \text{-}0.6\}$ we compute w using (2.18). To get the parameter values in the bottom line we set $R^2(\tilde{S}^*, \tilde{b})$ equal to 0.999 and compute the lowest value of β as well as the implied value for w. var($\tilde{P}$) is computed assuming var($\tilde{S}^*$) = 1, i.e. var($\tilde{P}$) $= 1 - 2w + z$. θ is implied as w/z.

line, with upper-tail observations from strong currencies being mostly on the rightmost/lower line, and lower-tail observations from weak currencies mostly on the leftmost/upper line.[3]

2.5 Numerical Confirmation of the Results for a U-shaped distribution of $\tilde{b}$

What remains to be done is to numerically verify whether the theoretical effects are sufficiently important to explain the effects actually observed, especially if the distribution of $\tilde{b}$ is U-shaped rather than the discrete $\pm\delta$ one.

We proceed as follows. We select parameter values so as to the generate twelve cases on a pre-set grid of $\beta = \{0.2, \text{-}0.2, \text{-}0.6\}$ and $z = \{1.1, 4, 7, 10\}$. We choose $z = 1.1$ instead of $z = 1$ because, at $z = 1$, β equals 1/2 irrespectively of w and because, close to $z = 1$, the bounds on β are hypersensitive to minute changes in the parameters. For each (β, z) combination, the corresponding values for the relative covariance, w, are then derived from the relation

$$w \stackrel{def}{=} \frac{\text{cov}(\tilde{S}^*, \tilde{b})}{\text{var}(\tilde{S}^*)} = \frac{(z+1)\beta - 1}{2\beta - 1}. \tag{2.18}$$

One combination on this grid, case ($\beta = -0.6, z = 10$), is incompatible with the unit upper

[3]Note also that the Huisman *et al.* beta is common across currencies but the intercepts are not. This avoids the downward bias that would have arisen if samples from different 45-degree lines had been pooled together.

bound on R^2. To get results near the bounds of feasibility for $z = \{1.1, 4, 7, 10\}$, we set $R^2(\tilde{S}^*, \tilde{b})$ equal to 0.999 and take the lower of the two corresponding β, from (1.8). All this produces the set of parameter values in Table 2. The distribution of $\tilde{\epsilon}$, the noise in $\tilde{S}^* = \mu_S + \theta.\tilde{b} + \tilde{\epsilon}$, is Gaussian, and that of $\tilde{b}$ is a mixture of a uniform on $[-\delta, \delta]$ and a Bernouilli. We chose how much of the probability distribution of $\tilde{b}$ is strictly within the bounds, and how much at the ends. From this parameter and w we then compute $\theta = w/z$. In turn, this θ (together with z) then implies a value for var($\tilde{\epsilon}$). More details are provided in the Appendix.

Table 3: Simulated betas (1): Bernouilli-distributed $\tilde{b}$; $\mu_S = 0$

	$z = 1.1$, $\beta = 0.2$		$z = 4$, $\beta = 0.2$		$z = 7$, $\beta = 0.2$		$z = 10$, $\beta = 0.2$	
split	β_{lo}	β_{hi}	β_{lo}	β_{hi}	β_{lo}	β_{hi}	β_{lo}	β_{hi}
60.0 / 40.0	0.12	0.21	-0.21	0.35	-0.03	0.33	0.06	0..31
80.0 / 20.0	0.14	0.22	-0.00	0.42	0.08	0.39	0.12	0.35
90.0 / 10.0	0.15	0.25	0.09	0.47	0.13	0.43	0.16	0.39
95.0 / 05.0	0.17	0.27	0.13	0.51	0.16	0.46	0.17	0.42
97.5 / 02.5	0.17	0.30	0.16	0.54	0.18	0.49	0.18	0.44
	$z = 1.1$, $\beta =$ -0.2		$z = 4$, $\beta =$ -0.2		$z = 7$, $\beta =$ -0.2		$z = 10$, $\beta =$ -0.2	
split	β_{lo}	β_{hi}	β_{lo}	β_{hi}	β_{lo}	β_{hi}	β_{lo}	β_{hi}
60.0 / 40.0	-0.30	-0.18	-0.86	0.02	-0.63	0.01	-0.51	-0.00
80.0 / 20.0	-0.27	-0.16	-0.52	0.14	-0.42	0.11	-0.37	0.08
90.0 / 10.0	-0.25	-0.13	-0.38	0.22	-0.33	0.17	-0.29	0.14
95.0 / 05.0	-0.24	-0.10	-0.30	0.28	-0.28	0.23	-0.25	0.19
97.5 / 02.5	-0.23	-0.06	-0.26	0.33	-0.25	0.27	-0.23	0.23
	$z = 1.1$, $\beta =$ -0.6		$z = 4$, $\beta =$ -0.6		$z = 7$, $\beta =$ -0.6		—	
split	β_{lo}	β_{hi}	β_{lo}	β_{hi}	β_{lo}	β_{hi}	—	—
60.0 / 40.0	-0.75	-0.57	-1.20	-0.31	-0.53	-0.94		
80.0 / 20.0	-0.72	-0.53	-0.91	-0.18	-0.50	-0.16		
90.0 / 10.0	-0.69	-0.47	-0.77	-0.09	-0.47	-0.42		
95.0 / 05.0	-0.66	-0.42	-0.70	-0.01	-0.45	-0.42		
97.5 / 02.5	-0.64	-0.38	-0.66	0.04	-0.43	-0.37		
	$z = 1.1$, $\beta =$ -1		$z = 4$, $\beta =$ -.997		$z = 7$, $\beta =$ -.607		$z = 10$, $\beta =$ -.462	
split	β_{lo}	β_{hi}	β_{lo}	β_{hi}	β_{lo}	β_{hi}	β_{lo}	β_{hi}
60.0 / 40.0	-1.22	-0.98	-1.04	-0.94	-0.62	-0.58	-0.47	-0.44
80.0 / 20.0	-1.17	-0.92	-1.02	-0.91	-0.61	-0.57	-0.47	-0.44
90.0 / 10.0	-1.13	-0.84	-1.01	-0.89	-0.61	-0.57	-0.47	-0.44
95.0 / 05.0	-1.10	-0.77	-1.00	-0.88	-0.61	-0.56	-0.46	-0.43
97.5 / 02.5	-1.07	-0.70	-1.00	-0.86	-0.61	-0.55	-0.46	-0.42

Table 3 displays the result of the computations for a case we studied analytically (that is, $\tilde{b} = \pm\delta$, with no mass in the middle), for an unconditionally no-drift distribution. In every case, β_{hi} rises above the marginal beta listed in the heading of the cell, and *a fortiori* above β_{lo}. Also, within every cell the β_{hi}s rise the higher one sets the cut-off values χ. While,

Table 4: Simulated betas (2): H-distributed $\tilde{b}$; $\mu_S = 0$

	z = 1.1, β = 0.2		z = 1.1, β = 0.2		z = 7, β = 0.2		z = 10, β = 0.2	
split	β_{lo}	β_{hi}	β_{lo}	β_{hi}	β_{lo}	β_{hi}	β_{lo}	β_{hi}
60.0 / 40.0	0.15	0.21	-0.09	0.27	0.02	0.27	0.09	0.25
80.0 / 20.0	0.16	0.23	0.03	0.35	0.10	0.33	0.13	0.30
90.0 / 10.0	0.18	0.25	0.10	0.40	0.14	0.37	0.16	0.33
95.0 / 05.0	0.18	0.27	0.14	0.43	0.16	0.40	0.18	0.36
97.5 / 02.5	0.19	0.30	0.16	0.47	0.18	0.43	0.19	0.39
	z = 1.1, β = 0.2		z = 4, β = -0.2		z = 7, β = -0.2		z = 10, β = -0.2	
split	β_{lo}	β_{hi}	β_{lo}	β_{hi}	β_{lo}	β_{hi}	β_{lo}	β_{hi}
60.0 / 40.0	-0.29	-0.19	-0.68	-0.07	-0.53	-0.09	-0.43	-0.10
80.0 / 20.0	-0.27	-0.16	-0.46	0.03	-0.37	0.01	-0.34	-0.01
90.0 / 10.0	-0.25	-0.14	-0.35	0.12	-0.31	0.08	-0.28	0.05
95.0 / 05.0	-0.23	-0.11	-0.30	0.17	-0.25	0.14	-0.24	0.10
97.5 / 02.5	-0.23	-0.07	-0.25	0.23	-0.23	0.18	-0.22	0.14
	z = 1.1, β = -0.6		z = 4, β = -0.6		z = 7, β = -0.6		—	
split	β_{lo}	β_{hi}	β_{lo}	β_{hi}	β_{lo}	β_{hi}	—	—
60.0 / 40.0	-0.74	-0.58	-1.05	-0.44	-0.64	-0.57		
80.0 / 20.0	-0.70	-0.55	-0.86	-0.31	-0.63	-0.54		
90.0 / 10.0	-0.67	-0.51	-0.75	-0.22	-0.62	-0.51		
95.0 / 05.0	-0.65	-0.47	-0.69	-0.14	-0.61	-0.49		
97.5 / 02.5	-0.64	-0.42	-0.65	-0.09	-0.60	-0.47		
	z = 1.1, β = -1		z = 4, β = -.997		z = 7, β = -.607		z = 10, β = -.462	
split	β_{lo}	β_{hi}	β_{lo}	β_{hi}	β_{lo}	β_{hi}	β_{lo}	β_{hi}
60.0 / 40.0	-1.18	-0.98	-1.02	-0.98	-0.62	-0.60	-0.47	-0.46
80.0 / 20.0	-1.15	-0.92	-1.01	-0.95	-0.62	-0.59	-0.47	-0.45
90.0 / 10.0	-1.10	-0.87	-1.01	-0.93	-0.61	-0.58	-0.47	-0.44
95.0 / 05.0	-1.07	-0.81	-1.00	-0.91	-0.61	-0.57	-0.46	-0.44
97.5 / 02.5	-1.05	-0.77	-1.00	-0.90	-0.61	-0.57	-0.46	-0.43

qualitatively, all this is as expected, we also note that the differences between β_{hi} and β_{lo} are substantially smaller than what Huisman *et al.* obtain. Nor do we see the β_{hi}s come anywhere near (or above) unity. Especially the numbers in the lower half of the table, from parameter constellations that produce marginal betas in the Froot-Thaler ballpark, are quite disappointing: all β_{hi}s but one remain negative, and the lone exception (0.04, for z = 4 and β = -0.6 , split 97.5%/02.5%) is quite unimpressive.

Table 4 demonstrates that the results obtained thus far are not very sensitive to the absence of probability mass within $[-\delta, +\delta]$. On a priori grounds, a U-shaped distribution, with some mass in the middle, is more likely indeed. But when we mix the Bernouilli with a uniform on $[-\delta, +\delta]$ and give the latter 25% probability weight, the results hardly change; if anything, the changes are for the worse. Results for 50% probability weight in the center are available on

Table 5: Simulated betas (3): H-distributed $\tilde{b}$; $\mu_S = 0.25$

	$z = 1.1, \beta = 0.2$		$z = 1.1, \beta = 0.2$		$z = 7, \beta = 0.2$		$z = 10, \beta = 0.2$	
split	β_{lo}	β_{hi}	β_{lo}	β_{hi}	β_{lo}	β_{hi}	β_{lo}	β_{hi}
60.0 / 40.0	0.18	0.23	-0.08	0.29	0.03	0.27	0.10	0.25
80.0 / 20.0	0.17	0.26	0.04	0.36	0.10	0.33	0.14	0.30
90.0 / 10.0	0.18	0.29	0.11	0.41	0.14	0.38	0.16	0.34
95.0 / 05.0	0.18	0.32	0.15	0.45	0.17	0.41	0.18	0.37
97.5 / 02.5	0.19	0.34	0.17	0.48	0.18	0.44	0.19	0.39
	$z = 1.1, \beta = 0.2$		$z = 4, \beta = \text{-}0.2$		$z = 7, \beta = \text{-}0.2$		$z = 10, \beta = \text{-}0.2$	
split	β_{lo}	β_{hi}	β_{lo}	β_{hi}	β_{lo}	β_{hi}	β_{lo}	β_{hi}
60.0 / 40.0	-0.22	-0.13	-0.63	-0.06	-0.50	-0.08	-0.43	-0.10
80.0 / 20.0	-0.22	-0.07	-0.44	0.05	-0.37	0.02	-0.33	-0.01
90.0 / 10.0	-0.22	0.01	-0.34	0.13	-0.30	0.09	-0.28	0.06
95.0 / 05.0	-0.22	0.13	-0.28	0.20	-0.26	0.15	-0.25	0.11
97.5 / 02.5	-0.21	0.21	-0.25	0.25	-0.23	0.20	-0.23	0.15
	$z = 1.1, \beta = \text{-}0.6$		$z = 4, \beta = \text{-}0.6$		$z = 7, \beta = \text{-}0.6$		—	
split	β_{lo}	β_{hi}	β_{lo}	β_{hi}	β_{lo}	β_{hi}	—	—
60.0 / 40.0	-0.58	-0.46	-0.97	-0.42	-0.62	-0.57		
80.0 / 20.0	-0.62	-0.37	-0.81	-0.27	-0.62	0.95		
90.0 / 10.0	-0.62	-0.23	-0.72	-0.16	-0.61	0.98		
95.0 / 05.0	-0.62	-0.07	-0.67	-0.08	-0.61	0.99		
97.5 / 02.5	-0.62	0.16	-0.64	-0.02	-0.60	0.99		
	$z = 1.1, \beta = \text{-}1$		$z = 4, \beta = \text{-}.997$		$z = 7, \beta = \text{-}.607$		$z = 10, \beta = \text{-}.462$	
split	β_{lo}	β_{hi}	β_{lo}	β_{hi}	β_{lo}	β_{hi}	β_{lo}	β_{hi}
60.0 / 40.0	-0.96	-0.80	-1.01	-0.96	-0.61	-0.60	-0.47	-0.45
80.0 / 20.0	-1.02	-0.59	-1.01	0.96	-0.61	0.98	-0.47	0.99
90.0 / 10.0	-.105	-0.43	-1.01	0.99	-0.61	1.00	-0.46	1.00
95.0 / 05.0	-1.04	-0.25	-1.00	0.99	-0.61	1.00	-0.46	1.00
97.5 / 02.5	-1.03	-0.21	-1.00	1.00	-0.61	1.00	-0.46	1.00

request; they confirm that the Bernouilli case tends to provide an overly optimistic picture, but only marginally so. The fact that the amount of mass put between $-\delta$ and $+\delta$ makes so little difference also suggests that our adoption of a H-shaped distribution for $\tilde{b}$ (instead of the U-shape that we should have had) is not likely to be material.

Our theoretical results showed that there should be a gain in effectiveness if sampling is asymmetric, so that the extreme observations tend to come predominantly from one particular tail and have similar values for $\tilde{b}$. The Huisman *et al.* procedure is likely to produce this feature: "weakness" and "strength"—i.e. high or low interest rates—are highly persistent features, so that that extreme-variance days are likely to have abnormally many date where strong currencies are at their strongest and weak currencies at their weakest. To produce such asymmetric sampling in our experiments, we set the mean of $\tilde{S}^*$ and $\tilde{P}^*$ equal to 0.25, keeping the cut-offs symmetric around zero at $\pm\chi$. To put this number in perspective: the standard deviation of the conditional expectations, a number that in reality is probably rather small, is standardized at unity in the simulations. That is, we can think of a mean of 0.25 as a small number. As predicted, both conditional variances (for *hi* and *lo*) are lower than the marginal variance; and var($\tilde{b} \mid hi$) drops steadily as the cut-off value χ is increased, while var($\tilde{b} \mid lo$) is U-shaped in χ. The impact of all this on β can be quite strong for selected parameter configurations, notably in cells ($z = 7, \beta = -0.6$) and ($z = 4, \beta = -1$) where the extreme betas get quite close to unity. When the unconditional mean is set equal to twice the variance of $\tilde{S}^*$ (results not shown), we get such betas also for an adjacent cell. Thus, the procedure does have the potential to generate quite high betas, and specifically for parameter constellations that are in the Froot-and-Thaler ballpark.[4]

3 Concluding remarks

The empirical failure of the UIP model suggests that a variable (b) needs to be added to the forward premium P. Fama (1983) showed that a β below 1/2 means that the missing variable shows more variability than the signal S^*. We provide some additional information. For instance, a negative β means that also the covariance between signal and noise is larger than the signal variability. And the Froot and Thaler meta-average β of -0.88 caps the noise-to-signal variance ratio at 4.5, and implies a whopping 0.9 lower bound on the correlation between signal and noise.

One of the possible explanations behind the failure of the unbiased expectations hypothesis is that the relation between expected exchange-rate changes and forward premia is obscured by transaction costs effects. Huisman *et al.* (1998) treat this transaction-cost effect as an errors-in-the-regressor problem, and note that the resulting bias toward zero should be reduced in samples with a better signal-to-noise ratio. Occasionally, expected exchange-rate changes can be very pronounced, leading to unusually large forward premia. Accordingly, Huisman *et al.* focus on days where forward premia tend to be exceptionally large, and find that in these "extreme samples" the bias toward zero is reduced (and perhaps even reversed). But there is a possible flaw here: given unusually large expectations, one expects larger forward

[4]True, the effects actually observed by Huisman *et al.* are much more extreme what is predicted here, but within our setting that must have been pure luck: there is no way a model with a true slope of unity can generate an expected coefficient way above unity. Huisman *et al.* do not provide t-tests on the deviation from unity.

premia indeed, but that does not necessarily mean that, given unusually large premia, also expectations tend to be outsized. For the second part of the statement to be true on average, the distribution of the conditional expectations must be thicker-tailed than the density of the missing variable. Interestingly, a transaction-cost model is likely to generate that property (Constantinides, 1986, and Dumas, 1992): transaction costs lead to a no-activity cone around the perfect-markets price, and the price tends to spend unusually long times at the border relative to, say, a uniform distribution. Also, the cone is far wider than the equilibrium price increased/decreased with the percentage transaction cost: if switching to and fro between currencies is costly, one wants currency B to move deep into the money before converting one's A-holdings into B in the first place. In short, in this type of model the percentage deviations from the perfect-markets equilibrium should be (*i*) bounded (that is, they cannot dominate the extreme forward premia), (*ii*) wide (that is, they can generate betas below .5) and (*iii*) U-shaped in density. We approximate the U-distribution by a H-distribution and derive theoretical results for a special case. From the numerical verifications we find that, for parameters in line with Froot and Thaler, extreme sampling can eliminate a lot of the transaction-cost-induced noise in forward premia. The mechanisms are more subtle than one would expect at first blush. For example, extreme sampling not only boosts the signal variance, but also induces a signal-noise covariance that improves beta, and may even lower the noise variance in absolute terms. Also, extreme sampling seems to be at its best when, for a given currency, large premia tend to come from from one end of the expectations distribution only. It is quite likely that this mechanism was active in the Huisman tests, too.

Appendix

Proof of Proposition 1

To prove (1.5)-(1.6), set $\frac{\mathrm{cov}(\tilde{S}^*,\tilde{b})}{\mathrm{var}(\tilde{S}^*)} = w$ and $\frac{\mathrm{var}(\tilde{b})}{\mathrm{var}(\tilde{S}^*)} = z$ so that

$$\beta = \frac{1-w}{1-2w+z}, \tag{A.1}$$

provided that $1-2w+z \neq 0$. As the correlation between $\tilde{S}^*$ and $\tilde{b}$ is below unity, and as $R^2(\tilde{S}^*,\tilde{b}) = w^2/z$, we know that w is bounded by $\sqrt{z.R^2}$ with $R^2 < 1$. For any given value of z, the value of w that maximizes or minimizes β s.t. $w^2 \leq z$ always is a corner solution, because

$$\frac{\partial}{\partial w}\frac{1-w}{1-2w+z} = \frac{1+z}{(1-2w+z)^2}$$

$$\neq \; 0 \text{ for } -\sqrt{z} < w < \sqrt{z}. \tag{A.2}$$

Thus, provided that $1-2w+z \neq 0$, the extrema for β, given z and given $R^2 < 1$, are obtained by setting w at the bounds, $\pm\sqrt{z}$. This leads to (1.5) and (1.6). **QED**

Proof of Proposition 2.

The relations $\tilde{P} = \tilde{S}^* - \tilde{b}$ and $\tilde{S}^* = \mu_S + \theta.\tilde{b} + \tilde{\epsilon}$ immediately imply that $\tilde{S}^* - \mu_S = (\theta-1)\tilde{b} + \tilde{\epsilon}$ with $\tilde{b}$ and $\tilde{\epsilon}$ independent. In a more standard notation, we are considering a two independent variables, X and Y, and we want to know to what extent the upper- tail probability of the sum, $P(X+Y > x)$ with x a large positive number, is determined by the upper tail probabilities of X and Y, and similarly for the lower tail. Let X be the thicker-tailed variable. Obviously, a high (low) value for $X+Y$ is not generated by extreme Ys when Y is negative (positive), so we are primarily interested in Y's contribution to the upper tail when Y is positive, and in Y's contribution to the lower tail when Y is negative. In the Lemma below, a to the best of our knowledge unpublished result kindly provided by Jef Teugels, it is shown that even when Y is positive it does not contribute to the upper-tail distribution of the sum and vice versa.

Lemma (Teugels): Let (X, Y) be a randomly drawn vector from a joint distribution, and let X be the thicker-tailed of the two in the sense that[5]

$$\lim_{x\uparrow\infty} \frac{P(Y \geq x)}{P(X \geq x)} \to 0. \tag{A.3}$$

Let X have dominatedly varying tails, i.e. for all $a > 0$ and some non-negative α we have[6]

$$\frac{P(X \geq ax)}{P(X \geq x)} \leq Ka^{-\alpha}. \tag{A.4}$$

Then even when $Y > (<)0$ the upper (lower) tail of the distribution of the sum is unaffected by Y:

$$\lim_{x\uparrow\infty} \frac{P(X+Y \geq x \mid Y > 0)}{P(X \geq x)} \;\to\; 1, \tag{A.5}$$

$$\lim_{x\downarrow-\infty} \frac{P(X+Y \leq x \mid Y < 0)}{P(X \leq x)} \;\to\; 1. \tag{A.6}$$

Proof: We prove the case where $Y > 0$. We immediately have

---

[5] A sufficient condition for this is that both variables are Pareto distributed with X having the lower tail exponent.

[6] This class comprises the Pareto-type class of distributions (notably when $K = 1$) and *a fortiori* the classical Pareto (when $P(X > x) = x^{-\alpha}$). See Bingham *et al.* (1987).

$$P(X+Y>x \mid Y>0) \geq P(X>x), \tag{A.7}$$

implying that

$$\lim_{x\uparrow\infty} \frac{P(X+Y>x \mid Y>0)}{P(X>x)} \geq 1. \tag{A.8}$$

To get an inequality in the other direction, we first note that, in the case we are interested in (i.e. deep in the right-hand tail), x is positive. We then use

$$\begin{aligned} P(X+Y>x \mid Y>0) &\leq P(X>x(1-a)) + P(X \leq x(1-a), Y>ax) \\ &\leq P(X>x(1-a)) + P(Y>ax). \end{aligned}$$

Dividing both sides by $P(X>x)$ we get

$$\begin{aligned} \frac{P(X+Y>x \mid Y>0)}{P(X>x)} &\leq \frac{P(X>x(1-a))}{P(X>x)} + \frac{P(Y>ax)}{P(X>x)} \\ &= \frac{P(X>x(1-a))}{P(X>x)} + \frac{P(Y>ax)}{P(X>ax)} \cdot \frac{P(X>ax)}{P(X>x)} \\ \lim_{x\uparrow\infty} \frac{P(X+Y>x \mid Y>0)}{P(X>x)} &\leq \lim_{x\uparrow\infty} \frac{P(X>x(1-a))}{P(X>x)} + 0 \cdot Ka^{-\alpha} \\ &\leq \lim_{x\uparrow\infty} \frac{P(X>x(1-a))}{P(X>x)}, \end{aligned} \tag{A.9}$$

where the third line uses (A.4) and (A.3). Taking a to be arbitrarily small, results (A.8) and (A.9) imply (A.5). **QED.**

The proof of proposition 2 follows immediately. If $\tilde{S}^*$ has the higher tail exponent relative to $\tilde{b}$, the tails of the density of the forward premium is determined by the density of $\tilde{S}^*$. **QED**

Proof of Proposition 3.

In the table below, we set out the critical values for $\tilde{\epsilon}$'s conditional probabilities and partial means in each of the four cases when $\mid \tilde{P} \mid$ (*i.e.* $\mid \tilde{S}^* - \tilde{b} \mid$) exceeds χ—depending on the sign of $\tilde{b}$, and, for each of these cases, first for the right and then the left tail:

partitioning of the $\tilde{S}^*$ high-$\mid \tilde{P} \mid$ events	probability conditional on $\tilde{b}$	partial mean of S^*
case 1: $\tilde{S}^* - \tilde{b} > \chi$ with $\tilde{b} = +\delta$ i.e. $\tilde{\epsilon} > \chi - \mu_S + (1-\theta)\delta \overset{def}{=} \epsilon_1$	$\pi_1 \overset{def}{=} \int_{\epsilon_1}^{\infty} f(\epsilon)\mathrm{d}\epsilon$	$\pi_1.(\mu_S + \theta\delta) + P_1, P_1 \overset{def}{=} \int_{\epsilon_1}^{\infty} \epsilon f(\epsilon)\mathrm{d}\epsilon$
case 2: $\tilde{S}^* - \tilde{b} < -\chi$ with $\tilde{b} = +\delta$ i.e. $\tilde{\epsilon} < -\chi - \mu_S + (1-\theta)\delta \overset{def}{=} \epsilon_2$	$\pi_2 \overset{def}{=} \int_{-\infty}^{\epsilon_2} f(\epsilon)\mathrm{d}\epsilon$	$\pi_2.(\mu_S + \theta\delta) + P_2, P_2 \overset{def}{=} \int_{-\infty}^{\epsilon_2} \epsilon f(\epsilon)\mathrm{d}\epsilon$
case 3: $\tilde{S}^* - \tilde{b} > \chi$ with $\tilde{b} = -\delta$ i.e. $\tilde{\epsilon} > \chi - \mu_S - (1-\theta)\delta \overset{def}{=} \epsilon_3$	$\pi_3 \overset{def}{=} \int_{\epsilon_3}^{\infty} f(\epsilon)\mathrm{d}\epsilon$	$\pi_3.(\mu_S - \theta\delta) + P_3, P_3 \overset{def}{=} \int_{\epsilon_3}^{\infty} \epsilon f(\epsilon)\mathrm{d}\epsilon$
case 4 $\tilde{S}^* - \tilde{b} < -\chi$ with $\tilde{b} = -\delta$ i.e. $\tilde{\epsilon} < \chi - \mu_S - (1-\theta)\delta \overset{def}{=} \epsilon_4$	$\pi_4 \overset{def}{=} \int_{-\infty}^{\epsilon_4} f(\epsilon)\mathrm{d}\epsilon$	$\pi_4.(\mu_S - \theta\delta) + P_4, P_4 \overset{def}{=} \int_{-\infty}^{\epsilon_4} \epsilon f(\epsilon)\mathrm{d}\epsilon$

The familiar solution for the conditional variances of a Bernouilli $\tilde{b}$ is

$$\mathrm{var}(\tilde{b} \mid hi) = 4\delta^2.\pi(\tilde{b} = \delta \mid hi).[1 - \pi(\tilde{b} = \delta \mid hi)], \tag{A.10}$$

and likewise for $\mathrm{var}(\tilde{b} \mid lo)$. Note that $\pi.(1-\pi)$ is maximal for $\pi = 1/2$. With the above definitions of the events, the probability of $\tilde{b} = +\delta$, conditional on a high $\mid \tilde{P} \mid$, is given by

$$\pi(\tilde{b} = \delta \mid hi) = \frac{\pi_1 + \pi_2}{(\pi_1 + \pi_2) + (\pi_3 + \pi_4)}. \tag{A.11}$$

The sign of $\pi(\tilde{b} = \delta \mid hi) - 1/2$ is the sign of $(\pi_1 + \pi_2) - (\pi_3 + \pi_4)$ or, equivalently, of $(\pi_2 - \pi_4) - (\pi_3 - \pi_1)$. The latter can be interpreted as

$$(\pi_2 - \pi_4) - (\pi_3 - \pi_1) = \int_{-\chi-\mu-(1-\theta)\delta}^{-\chi-\mu+(1-\theta)\delta} f(\epsilon)\mathrm{d}\epsilon - \int_{\chi-\mu-(1-\theta)\delta}^{\chi-\mu+(1-\theta)\delta} f(\epsilon)\mathrm{d}\epsilon, \tag{A.12}$$

which is negative (Figure A1): both areas are $2(1-\theta)\delta$ wide, but in a symmetric, unimodal distribution, if $\mu_S > 0$ then the area above $[-\chi - \mu_S - (1-\theta)\delta, -\chi - \mu_S + (1-\theta)\delta]$, being more off-center, is smaller than the area above $[\chi - \mu_S - (1-\theta)\delta, \chi - \mu_S + (1-\theta)\delta]$. Thus, $\pi(\tilde{b} = \delta \mid hi) < 1/2$. Similarly, when μ_S is negative, we have $\pi(\tilde{b} = \delta \mid hi) < 1/2$. Therefore $\mathrm{var}(\tilde{b} \mid hi) < \delta^2 = \mathrm{var}(\tilde{b})$.

Also from proposition 2, the probability of a positive $\tilde{b}$ when $\mid \tilde{P} \mid$ is low is given by

$$\pi(\tilde{b} = \delta \mid lo) = \frac{1 - (\pi_1 + \pi_2)}{[1 - (\pi_1 + \pi_2)] + [1 - (\pi_3 + \pi_4)]}. \tag{A.13}$$

If $\mu_S > 0$, then $\pi(\tilde{b} = \delta \mid lo) > 1/2$, meaning likewise that $\mathrm{var}(\tilde{b} \mid lo) < \delta^2 = \mathrm{var}(\tilde{b})$. Initially, this phenomenon becomes more important the higher the cut-off value. However, for large χ, the high-$\mid \tilde{P} \mid$ sample becomes so small that more and more of the "$\tilde{b} = -\delta$" observations end up in the low-$\mid \tilde{P} \mid$ sample. That is, all π_is shrink. In the limit, with a very strict definition of "*hi*", almost the entire distribution of $\tilde{P}$ is classified as low-$\mid \tilde{P} \mid$, and $\pi(\tilde{b} = \delta \mid lo)$ again approaches the marginal mean, 1/2. Thus, for increasingly stricter definitions of "*hi*", $\mathrm{var}(\tilde{b} \mid lo)$ first falls and then inches up again to $\mathrm{var}(\tilde{b}) = \delta^2$. **QED.**

Proof of Proposition 4.

The sign of $\mathrm{cov}(\tilde{\epsilon}, \tilde{b} \mid hi)$ is the same as the sign of $\gamma_{hi} \stackrel{def}{=} \frac{\mathrm{cov}(\tilde{\epsilon}, \tilde{S}^* \mid hi)}{var(\tilde{b} \mid hi)}$, the conditional regression coefficient of $\tilde{\epsilon}$ on $\tilde{b}$. As $\tilde{b}$ can assume only two values, this regression coefficient can be written as the slope of the line through the points $(\delta, \mathrm{E}(\tilde{\epsilon} \mid hi \cap \tilde{b} = \delta))$ and $(-\delta, \mathrm{E}(\tilde{\epsilon} \mid hi \cap \tilde{b} = -\delta))$.

Using the definitions of the table in Proposition 3, the events "$hi \cap \tilde{b} = \delta$" and "$hi \cap \tilde{b} = -\delta$" can be rewritten as "$\tilde{\epsilon} \leq \epsilon_2 \cup \tilde{\epsilon} \geq \epsilon_1$" and "$\tilde{\epsilon} \leq \epsilon_4 \cup \tilde{\epsilon} \geq \epsilon_3$", respectively. Thus,

$$\gamma_{hi} \stackrel{def}{=} \frac{\text{cov}(\tilde{\epsilon}, \tilde{S}^* \mid hi)}{var(\tilde{b} \mid hi)} = \frac{\text{E}(\tilde{\epsilon} \mid \tilde{\epsilon} \leq \epsilon_2 \cup \tilde{\epsilon} \geq \epsilon_1) - \text{E}(\tilde{\epsilon} \mid \tilde{\epsilon} \leq \epsilon_4 \cup \tilde{\epsilon} \geq \epsilon_3)}{2\delta}. \tag{A.14}$$

We now show that the numerator on the right hand side is negative. Let

$f(\epsilon)$ = the density of $\tilde{\epsilon}$, symmetric and unimodal with mean zero

$F(\epsilon) = \int_{-\infty}^{\epsilon} f(\epsilon)\text{d}f(\epsilon)$

$\pi_b = F(\epsilon_b)$, the probability that $\tilde{\epsilon} \leq \epsilon_b$

$\overline{\pi}_a = 1 - F(\epsilon_a)$, the probability that $\tilde{\epsilon} \geq \epsilon_a$

$$\begin{aligned} \phi(\epsilon; \epsilon_a, \epsilon_b) &= \text{the density of } \tilde{\epsilon} \text{ given that } \tilde{\epsilon} \text{ is in the tails, } i.e.\ (\tilde{\epsilon} \leq \epsilon_b \cup \tilde{\epsilon} \geq \epsilon_a) \\ &= \begin{cases} \frac{f(\epsilon)}{\overline{\pi}_a + \pi_b} & \text{if } \tilde{\epsilon} \leq \epsilon_b, \\ 0 & \text{if } \epsilon_b < \tilde{\epsilon} < \epsilon_a, \\ \frac{f(\epsilon)}{\overline{\pi}_a + \pi_b} & \text{if } \tilde{\epsilon} \geq \epsilon_a, \end{cases} \end{aligned}$$

$\Phi(\epsilon; \epsilon_a, \epsilon_b) = \int_{-\infty}^{\epsilon} \phi(\epsilon; \epsilon_a, \epsilon_b)\text{d}\phi(\epsilon)$, the corresponding (cumulative) probability function.

We easily establish the relationship between $F(\epsilon)$ and $\Phi(\epsilon; \epsilon_a, \epsilon_b)$—see Figure A2:

- for $\tilde{\epsilon} \leq \epsilon_b$, we obviously have $(\overline{\pi}_a + \pi_b).\Phi(\epsilon; \epsilon_a, \epsilon_b) = F(\epsilon)$;
- on $[\epsilon_b, \epsilon_a]$, $(\overline{\pi}_a + \pi_b).\Phi(\epsilon; \epsilon_a, \epsilon_b)$ remains flat while $F(\epsilon)$ rises from π_b to $1 - \overline{\pi}_a$, *i.e.* by an additional $(1 - \overline{\pi}_a - \pi_b)$;
- beyond ϵ_a, $(\overline{\pi}_a + \pi_b).\Phi(\epsilon; \epsilon_a, \epsilon_b)$ again rises in step with $F(\epsilon)$ but from a level that is $(1 - \overline{\pi}_a - \pi_b)$ below $F(\epsilon)$.

In short,

$$\Phi(\epsilon; \epsilon_a, \epsilon_b) = \begin{cases} \frac{F(\epsilon)}{\overline{\pi}_a + \pi_b} & \text{for } \tilde{\epsilon} \leq \epsilon_b, \\ \frac{F(\epsilon_b)}{\overline{\pi}_a + \pi_b} & \text{for } \epsilon_b < \tilde{\epsilon} < \epsilon_a, \\ \frac{F(\epsilon) - (1 - \overline{\pi}_a - \pi_b)}{\overline{\pi}_a + \pi_b} = \frac{F(\epsilon) - 1}{\overline{\pi}_a + \pi_b} + 1 & \text{for } \tilde{\epsilon} \leq \epsilon_a. \end{cases} \tag{A.15}$$

We need to rank $\text{E}(\tilde{\epsilon} \mid \tilde{\epsilon} \leq \epsilon_2 \cup \tilde{\epsilon} \geq \epsilon_1)$ versus $\text{E}(\tilde{\epsilon} \mid \tilde{\epsilon} \leq \epsilon_4 \cup \tilde{\epsilon} \geq \epsilon_3)$ where $\epsilon_4 < \epsilon_2 < \epsilon_1$ and $\epsilon_4 < \epsilon_3 < \epsilon_1$. The expectation of $\tilde{\epsilon}$ given that $\tilde{\epsilon} \leq \epsilon_b$ or $\tilde{\epsilon} \geq \epsilon_a$ can be derived from

$$\text{E}(\tilde{\epsilon} \mid \tilde{\epsilon} \leq \epsilon_b \cup \tilde{\epsilon} \geq \epsilon_a) = \int_{-\infty}^{+\infty} \epsilon \phi(\epsilon; \epsilon_a, \epsilon_b)\text{d}f(\epsilon)$$

$$= \int_{-\infty}^{+\infty} [1 - \Phi(\epsilon; \epsilon_a, \epsilon_b)] \mathrm{d}f(\epsilon), \tag{A.16}$$

where the second line, a familiar result, follows from partial integration. We now establish that over the entire integration range we have $\Phi(\epsilon; \epsilon_1, \epsilon_2) > \Phi(\epsilon; \epsilon_3, \epsilon_4)$, which is a sufficient condition for the ranking $\mathrm{E}(\tilde{\epsilon} \mid \tilde{\epsilon} \leq \epsilon_2 \cup \tilde{\epsilon} \geq \epsilon_1) < \mathrm{E}(\tilde{\epsilon} \mid \tilde{\epsilon} \leq \epsilon_4 \cup \tilde{\epsilon} \geq \epsilon_3)$. We start with the case $\mu_S > 0$ and use the result from Proposition 3 that,

$$\text{if } \mu_S > 0 \text{ then } \pi_1 + \pi_2 < \pi_3 + \pi_4. \tag{A.17}$$

From this it immediately follows that, below ϵ_4 and beyond ϵ_1, we have $\Phi(\epsilon; \epsilon_1, \epsilon_2) > \Phi(\epsilon; \epsilon_3, \epsilon_4)$ (Figure A3):

- on $[-\infty, \epsilon_4]$, both Φs rise in $\tilde{\epsilon}$, with

$$\Phi(\epsilon; \epsilon_1, \epsilon_2) = \frac{F(\epsilon)}{\pi_1 + \pi_2} > \frac{F(\epsilon)}{\pi_3 + \pi_4} = \Phi(\epsilon; \epsilon_3, \epsilon_4); \tag{A.18}$$

- on $[\epsilon_1, \infty]$, both Φs also rise in $\tilde{\epsilon}$, with

$$\Phi(\epsilon; \epsilon_1, \epsilon_2) = \frac{F(\epsilon) - 1}{\pi_1 + \pi_2} + 1 > \frac{F(\epsilon) - 1}{\pi_3 + \pi_4} + 1 = \Phi(\epsilon; \epsilon_3, \epsilon_4). \tag{A.19}$$

Equations (A.18) and (A.19) trivially imply

$$\Phi(\epsilon_4; \epsilon_1, \epsilon_2) > \Phi(\epsilon_4; \epsilon_3, \epsilon_4), \tag{A.20}$$

$$\Phi(\epsilon_1; \epsilon_1, \epsilon_2) > \Phi(\epsilon_1; \epsilon_3, \epsilon_4). \tag{A.21}$$

It then follows that also on $[\epsilon_4, \epsilon_1]$ we have $\Phi(\epsilon; \epsilon_1, \epsilon_2) > \Phi(\epsilon; \epsilon_3, \epsilon_4)$. This is because, as illustrated in Figure A2,

- on $[\epsilon_4, \epsilon_2]$,$\Phi(\epsilon; \epsilon_1, \epsilon_2)$ goes on rising (from a higher starting point, see (A.18)), while $\Phi(\epsilon; \epsilon_3, \epsilon_4)$ initially stays flat;
- on $[\epsilon_3, \epsilon_1]$, $\Phi(\epsilon; \epsilon_3, \epsilon_4)$ starts rising, but in light of of (A.19) it never quite catches up with $\Phi(\epsilon; \epsilon_1, \epsilon_2)$ even in the range where the latter is flat.

Thus, when $\mu_S > 0$, then everywhere we have $\Phi(\epsilon; \epsilon_1, \epsilon_2) > \Phi(\epsilon; \epsilon_3, \epsilon_4)$, implying, from (A.16), that $\mathrm{E}(\tilde{\epsilon} \mid \tilde{\epsilon} \leq \epsilon_2 \cup \tilde{\epsilon} \geq \epsilon_1) < \mathrm{E}(\tilde{\epsilon} \mid \tilde{\epsilon} \leq \epsilon_4 \cup \tilde{\epsilon} \geq \epsilon_3)$. This establishes the proof when $\mu_S > 0$.

Next suppose that $\mu_S < 0$. Then

$$\pi(\tilde{\epsilon} \leq \epsilon_2 \cup \tilde{\epsilon} \geq \epsilon_1) > \pi(\tilde{\epsilon} \leq \epsilon_4 \cup \tilde{\epsilon} \geq \epsilon_3) \tag{A.22}$$

We now use the symmetry-around-zero property for $f(\epsilon)$ to restate the problem:

$$
\begin{aligned}
\mathrm{E}(\tilde{\epsilon} \mid \tilde{\epsilon} \leq \epsilon_a \cup \tilde{\epsilon} \geq \epsilon_b) &= -\mathrm{E}(\tilde{\epsilon} \mid \tilde{\epsilon} \geq -\epsilon_a \cup \tilde{\epsilon} \leq -\epsilon_b) \\
&= -\mathrm{E}(\tilde{\epsilon} \mid \tilde{\epsilon} \geq \tilde{\epsilon}_a^* \cup \tilde{\epsilon} \leq \tilde{\epsilon}_b^*), \qquad \text{(A.23)}
\end{aligned}
$$

where $\tilde{\epsilon}_a^* \stackrel{def}{=} \epsilon_b$ and $\tilde{\epsilon}_b^* \stackrel{def}{=} \epsilon_a$. Thus, the ranking $\mathrm{E}(\tilde{\epsilon} \mid \tilde{\epsilon} \leq \epsilon_2 \cup \tilde{\epsilon} \geq \epsilon_1) < \mathrm{E}(\tilde{\epsilon} \mid \tilde{\epsilon} \leq \epsilon_4 \cup \tilde{\epsilon} \geq \epsilon_3)$ is equivalent to the ranking $\mathrm{E}(\tilde{\epsilon} \mid \tilde{\epsilon} \geq \tilde{\epsilon}_1^* \cup \tilde{\epsilon} \leq \tilde{\epsilon}_2^*) > \mathrm{E}(\tilde{\epsilon} \mid \tilde{\epsilon} \geq \tilde{\epsilon}_3^* \cup \tilde{\epsilon} \leq \tilde{\epsilon}_4^*)$, and a sufficient condition for the latter is that, everywhere, $\Phi(\epsilon; \epsilon_1^*, \epsilon_2^*) < \Phi(\epsilon; \epsilon_3^*, \epsilon_4^*)$. Figure A4 pictures the critical points in the original and the restated problem. We see that the critical range for the lower-probability event is now to the right of the range of the higher-probability event, as was the case when $\mu_S > 0$. Thus, *mutatis mutandis* the previous proof still applies.

Technical Note on the Simulations

Define $w \stackrel{def}{=} \mathrm{cov}(\tilde{S}^*, \tilde{b})/\mathrm{var}(\tilde{S}^*)$ and $z \stackrel{def}{=} \mathrm{var}(\tilde{b})/\mathrm{var}(\tilde{S}^*)$. Consider the regression $\tilde{S}^* = \beta \tilde{P} + \nu$ where

$$\beta = \frac{1-w}{1-2w+z}. \qquad \text{(A.24)}$$

We work with the grid of βs and zs explained in the main text (Table 2), and determine the implied values of w, θ, and $\mathrm{var}(\tilde{\epsilon})$. We then generate a simple U-shaped distribution for $\tilde{b}$, as follows. Let $\tilde{U} = \tilde{U}(-\alpha, +\alpha)$ be uniformly distributed on $[-\alpha, \alpha]$ where $\alpha > \delta$. Then $\pi(-\delta \leq \tilde{U} \leq +\delta) = \frac{\delta}{\alpha}$, which then implies

$$\pi(\tilde{U} < -\delta) = \pi(\tilde{U} > +\delta) = \frac{1}{2}\left(1 - \frac{\delta}{\alpha}\right). \qquad \text{(A.25)}$$

We generate $\tilde{b}$ as

$$\tilde{b} = \begin{cases} -\delta & \text{if } \tilde{U} \leq -\delta \quad , \text{prob} = \frac{1}{2}\left(1-\frac{\delta}{\alpha}\right), \\ \tilde{U} & \text{if } -\delta < \tilde{U} < +\delta \quad , \text{density} = \frac{1}{2\alpha}, \\ +\delta & \text{if } \tilde{U} \geq +\delta \quad , \text{prob} = \frac{1}{2}\left(1-\frac{\delta}{\alpha}\right). \end{cases} \qquad \text{(A.26)}$$

This implies a mean and variance given by

$$
\begin{aligned}
\mathrm{E}(\tilde{b}) &= 0, && \text{(A.27)} \\
\mathrm{var}(\tilde{b}) = \mathrm{E}(\tilde{b}^2) &= \frac{1}{2}\left(1 - \frac{\delta}{\alpha}\right).[\delta^2 + (-\delta)^2] + \frac{1}{2\alpha}\int_{-\delta}^{+\delta} b^2 \mathrm{d}b && \text{(A.28)} \\
&= \left(1 - \frac{\delta}{\alpha}\right).\delta^2 + \frac{\delta^3}{3\alpha} && \text{(A.29)} \\
&= \delta^2\left(1 - \frac{2}{3}\frac{\delta}{\alpha}\right). && \text{(A.30)}
\end{aligned}
$$

The above implies that we need to pre-set one additional variable, which we chose to be δ/α. This ratio indicates how much of the mass of $\tilde{b}$ is on the bounds, $\pm\delta$. The limiting cases are (i) $[\alpha \to \infty, \delta > 0]$ (implying $\delta/\alpha = 0$), a binomial distribution with variance δ^2; and (ii) $\alpha = \delta$, a continuous uniform distribution with variance $\delta^2/3$. The implied value of δ is found from

$$\text{var}(\tilde{b}) = z = \delta^2 \left(1 - \frac{2}{3}\frac{\delta}{\alpha}\right) \tag{A.31}$$

implying

$$\delta = \sqrt{\frac{z}{1 - \frac{2}{3}\frac{\delta}{\alpha}}}. \tag{A.32}$$

References

Baldwin, E. R., 1990, Reinterpreting the Failure of the Foreign Exchange Market Efficiency Tests: Small Transaction Costs, *CEPR Discussion Papers*, 407

Ball, R., S. P. Kothari, and C.E. Wasley, 1995, "Can We Implement Research on Stock Trading Rules?", *Journal of Portfolio Management* **21**, 54-63

Bansal, R., 1997, An Exploration of the forward Premium Puzzle in Currency Markets, *Review of Financial Studies* **10**, 369-403

Bilson, J.F.O., 1981, The Speculative Efficiency Hypothesis, *Journal of Business* **54**, 435-451

Bingham, N.H., C.H. Goldie and J.L. Teugels (1987), *Regular Variation*, Cambridge University Press.

Brennan, M.J., and E. Schwartz, 1988, Optimal arbitrage strategies under basis variability, Studies in banking and Finance 5, 167-180

Brennan, M.J., and E. Schwartz, 1990, Arbitrage in stock index futures, *Journal of Business* 63, S7-S31

Constantinides, G. M., 1986, Capital Market Equilibrium with Transaction Costs, *Journal of Political Economy* 94(4), August 1986, pages 842-62.

Cumby, R. and M. Obstfeld, 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.

Dooley, M.P., and J. Schafer, 1983, Analysis of Short-Run Exchange Rate behavior: March 1973 to November 1981, in Bigman, D. and T. Taya (eds.), *Exchange Rate Instability: Causes, Consequences and Remedies*, Ballinger, Cambridge, MA

Dumas, B. (1992): Dynamic Equilibrium and the Real Exchange Rate in a Spatially Separated World, *The Review of Financial Studies* 5, 153-180.

Fama, E., 1984, Forward and Spot Exchange Rates, *Journal of Monetary Economics* **14**, 319-338

Froot, K. and R. Thaler, 1990, Anomalies: Foreign Exchange, *Journal of Economic Perspectives* **4**, 179-192

Hansen, L. P. and R. J. Hodrick, 1980, Forward Exhange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis, *Journal of Political Economy* **88**(5), 829-853

Hodrick, R. J. and S. Srivastava, 1986, The Covariation of Risk Premiums and Expected Future Spot Exchange Rates, *Journal of International Money and Finance* **5**, 5-22

Hollifield, B., and R. Uppal, 1997, An Examination of Uncovered Interest Parity in Segmented International Commodity Markets, *Journal of Finance* **52**, 2145-2170

Huisman, R., K. Koedijk, C. Kool, F. Nissen, 1998, Extreme Support for Uncovered Interest Parity, *Journal for International Money and Finance* **17**, 211-228.

Lehmann, B.N., 1990, "Fads, Martingales, and Market Efficiency." *Quarterly Journal of Economics* , 1-28

Nissen, F., 1997, *International Financial Markets Dynamics. An Empirical Investigation of Exchange Rates, Interest Rates, and Stock returns*, Doctoral Dissertation, Universiteit Maastricht, 168 p

Sercu, P. and T. Vinaimont, 2001, Peso Effects in the Forward Bias: Evidence from the Private ECU, KU Leuven DTEW Research Report 2001/2376/36

Figure 1: bounds on β for various $z \stackrel{def}{=} \frac{\text{var}(\tilde{b})}{\text{var}(\tilde{S}^*)}$

The figure shows the highest and lowest possible values of beta for various levels of the noise-to-signal variance ratio.

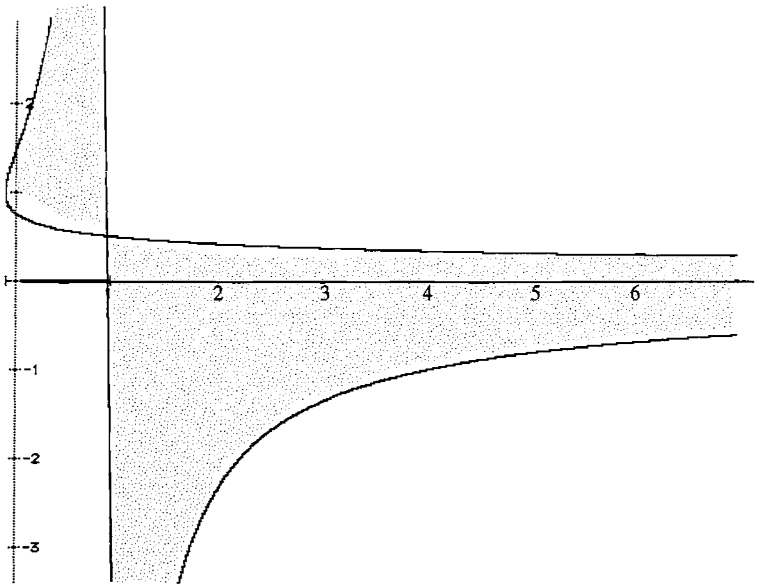

Figure 2: bounds on β for various $z \stackrel{def}{=} \frac{\text{var}(\tilde{b})}{\text{var}(\tilde{S}^*)}$ for the $R^2_{max} = .9$

The figure shows the highest and lowest possible values of beta for various levels of the noise-to-signal variance ratio when the R^2 between signal and noise is at most .90.

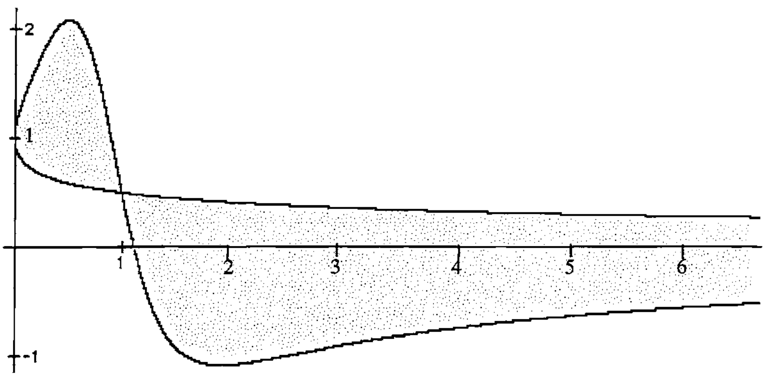

Figure A1

In a symmetric, unimodal distribution, the area below the curve of width $2(1-\theta)\delta$ is larger than another one with the same width is the former is closer to the center.

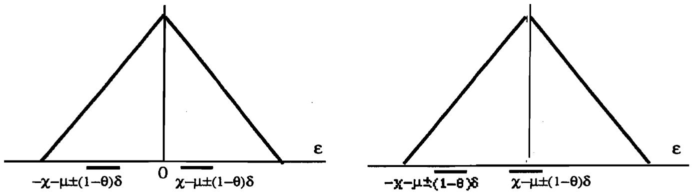

Figure A2: The link between $F(\epsilon)$ and $\Phi(\epsilon \mid \epsilon_a, \epsilon_b)$

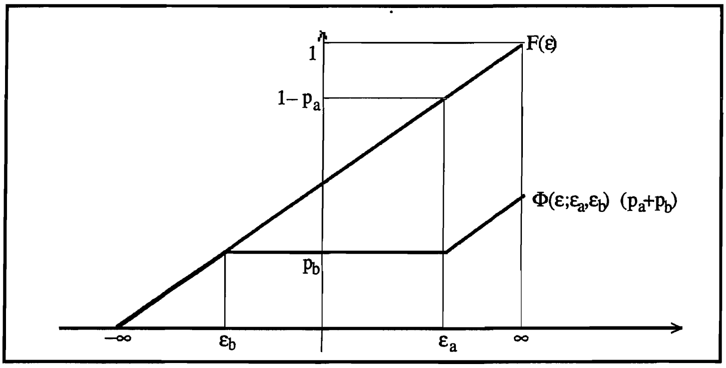

Figure A3

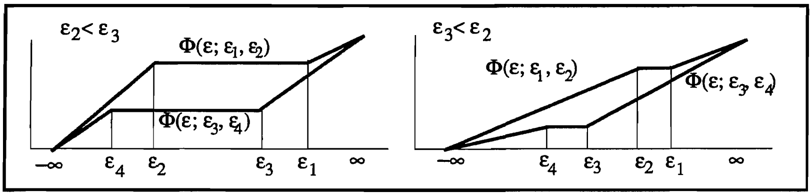

Figure A4: Restating the problem when $\mu < 0$.