Exchange Rates Under Robustness: An Account of the Forward Premium Puzzle*

Ming Li
San Francisco State University

Aaron Tornell
UCLA

First draft: May 2006
This draft: October 2008

Abstract

We show that robustness against model misspecification can account for the forward premium puzzle through a combination of an exchange rate model and a robustness model under structured uncertainty. In equilibrium, optimizing agents, who hold no misperception about the model, distort their forecasts to attain robustness against potential misspecification. This forecast distortion generates a delayed overreaction of exchange rates to interest rate differential shocks that leads to a negative unconditional correlation between exchange rate changes and interest rate differentials, i.e., a negative Fama coefficient. Using change-of-measure techniques, we derive the familiar uncovered interest rate parity condition–under distorted expectations–and the Fama coefficient in closed-form. We calibrate our model with empirical estimates of key parameters and are able to generate a negative Fama coefficient under sufficient uncertainty-aversion.

JEL Classification: C61, E43, F31, G12.

Keywords: Misspecification, Momentum, Overreaction, Predictability, Robustness, Structured uncertainty, Underreaction.

*We thank Menzie Chinn, Steve Durlauf, Chales Engel, Pierre O. Gourinchas, Jin Hahn, Chris Hellwig, Sebnem Kalemli-Ozcan, Rosa Matzkin, David Pappel, Romain Ranciere, Pierre O. Weil, Ken West, and participants at seminars in UCLA, University of Houston, University of Wisconsin Maddison, and USC.

1. Introduction

We show that a desire for *robustness* against a class of model misspecification accounts for the forward premium puzzle (FPP) in its strong form–a negative Fama coefficient. This important foreign exchange market anomaly implies that the domestic currency tends to appreciate when domestic interest rates are higher than abroad. This pattern generates predictable excess returns as investors can pocket both the interest rate differential as well as the subsequent gains from appreciation.

Our analysis combines robustness and exchange rate models to generate a hump-shaped forecast pattern that accounts for the FPP. The key variable is the differential between the interest rates of two currencies. In our setup this differential is hit by transitory shocks as well as persistent, mean-reverting shocks. Agents do not observe these shocks separately, and constantly try to determine their duration. They have a *baseline model* of the differential that corresponds to the data generating process. However, as in Hansen and Sargent (2007), agents fear that their model is misspecified. To be robust against misspecification, agents overstate the relative importance of transitory shocks; they then forecast exchange rates based on the distorted probability distribution of the interest rate differential process. Given these *robust forecasts*, exchange rates are determined by the familiar no-arbitrage condition, i.e., "robust uncovered interest rate parity" holds. An implication of the equilibrium in this model is delayed overreaction to news. When a shock hits, robust forecasts initially react less than baseline forecasts, and subsequently will need to catch up before mean reverting. The interaction of this hump-shaped forecast pattern and the gradual mean-reversion of interest rate differential shocks generates a negative Fama coefficient because as the robust forecasts catch up, a gradual exchange rate appreciation coexists with a positive interest rate differential.[1]

The guiding principle of our modeling strategy is to find conditions under which robustness induces such a hump-shaped forecast pattern that is then carried over to the equilibrium exchange rate. It turns out that simply invoking robustness against *any* type of misspecification does not generate this hump-shaped pattern. In particular, it is necessary to specify the aspects of the model that agents fear are misspecified.

[1] In the absence of fear of model misspecification this setup does not account for the forward premium puzzle since rational agents cannot be systematically fooled.

This requires a more refined description of model uncertainty than that contained in the so-called unstructured uncertainty. We refine the uncertainty set in several ways so as to capture the different avenues in which the model can be misspecified. We also derive Girsanov-like change-of-measure results that allow for a representation of the no-arbitrage condition in terms of the familiar uncovered interest parity condition albeit under distorted probability measures. This representation, in turn, leads to closed-form solutions of the equilibrium exchange rate path and the Fama coefficient, and allows us to establish a transparent link between classes of misspecification and the empirical FPP literature, which is, to our knowledge, a novel contribution.

Our main result derives from the overstatement of the relative importance of transitory shocks, which is an outcome of the agent's optimization problem given that she fears *misspecification* in the equation that links noisy observations and the unobservable persistent component of the interest rate differential. A negative Fama coefficient arises under this "observational uncertainty" when the interest rate differential is highly persistent and there is a high, but not too high, aversion to model misspecification, i.e., uncertainty-aversion. We find that empirical estimates of key parameters generate a negative Fama coefficient provided there is enough uncertainty-aversion.

To give some intuition, consider the response to a one-time unexpected shock. Suppose that dollar interest rates increase relative to euro interest rates and agents know that there is a high probability this shock is *persistent*. In the absence of robustness, since agents know the nature of the shock, arbitrage would force the dollar to appreciate immediately relative to its long-run value, up to the point where its expected depreciation equals the interest rate differential. The dollar would then gradually revert to equilibrium as the differential vanishes. This is the overshooting or the "forward premium effect" characterized by Dornbusch (1976). Next, for the sake of argument, suppose that as a means to attain robustness, agents treat the shock as highly *transitory*.[2] The robust interest parity condition implies that at impact the dollar needs only to appreciate moderately. In the following period, the dollar interest rate turns out to be higher than the agents' initial forecast. This leads agents to upwardly revise their forecasts about the persistence of the original interest rate shock, triggering

[2] This extreme assumption is not necessary and is made only for the sake of clarity. Proposition 5.2 states the conditions for humped-shaped dynamics.

further appreciation of the dollar. This is the "catch-up effect." If the catch-up effect dominates the forward premium effect, the initial *appreciation* of the dollar will be followed by a subsequent gradual *appreciation*. Eventually, the exchange rate will revert to a depreciating path and converge to its equilibrium value because the interest rate differential is mean-reverting. The key point is that along this path, a *positive* interest rate differential coexists with an *expected appreciation*. It turns out that the intuition for this hump-shaped path *conditional* on a one-time shock carries over to a negative Fama coefficient which reflects a negative *unconditional* correlation between interest rate differential shocks and exchange rate changes.[3]

Next, we provide intuition for distorted forecasts as a means to attain robustness. A robust strategy must work reasonably well within the uncertainty set regardless of the actual data generating process. Thus, in designing her strategy, the agent asks, if the model were misspecified, what would be the costliest direction in which things could go wrong. She then weighs the benefits of dampening the effects of misspecification in this costliest direction against the costs of moving away from optimality under the baseline model. This trade-off is weighted by her degree of uncertainty aversion. In the case of observational uncertainty, the costliest misspecifications occur when the agent's model incorrectly assigns a lower importance to transitory shocks than the importance they actually have. Robustness thus implies an upward distortion of the importance of transitory shocks, as measured by their variance. This distortion increases in the degree of uncertainty aversion.[4]

What about other classes of misspecification? When there is structured uncertainty about the unobservable persistent component of the interest rate differential (either parametric or to the shock process), we find that robustness implies understating the relative importance of transitory shocks. As a result, we cannot account for the FPP –in fact, the Fama regression coefficient would be greater than one. We also consider the unstructured uncertainty case, under which agents fear misspecification of the en-

[3] The *conditional* delayed overreaction –i.e., humped-shaped– response to an unanticipated interest rate shock represents the existence of conditional *momentum* in exchange rates. In the international finance literature it is known as *delayed overshooting* and has been documented by Eichenbaum and Evans (1995)). It stands in contrast to Dornbusch's (1976) overshooting path along which there is an immediate appreciation followed by a gradual depreciation.

[4] The downside of robustness is that if misspecification were inexistent, the estimation of the persistent component of the differential would be inefficiently slow.

tire interest rate differential process, but cannot pinpoint either its nature or location. The result is surprising: robust forecasts are equal to the Bayesian forecasts under the baseline model. Thus, the forecast-distortion mechanism underlying the FPP is not operative. This is because the robust problem reduces to a standard Bayesian problem via the Representation Lemma.

The forecast-distortion mechanism we model is consistent with several papers in the international finance literature. Using survey data Frankel and Froot (1989) find that exchange rate forecast distortions account for an important part of the FPP. Using interest rate survey data Gourinchas and Tornell (2004) find that forecasts overstate the relative importance of transitory shocks and show that the distortion is strong enough to generate a negative Fama coefficient. Burnside, et. al. (2008) find that the carry-trade is profitable and that the main part of its profitability does not reflect a compensation for a large peso-problem or a risk premium. In our equilibrium, the type of misspecifications that agents guard against do not involve a large peso-problem. In fact, when uncertainty aversion is extreme, so that the distortion reflects the worst-case scenario, robustness does not account for the FPP.

The structure of the paper is as follows. In Section 2, we present an intuitive overview. In Section 3, we present the model. In Section 4, we derive the equilibrium exchange rate. In Section 5, we characterize the conditions under which observation uncertainty generates the anomalies. In Section 6, we consider other types of model uncertainty. In Sections 7 and 8, we present a literature review and the conclusions, respectively. Finally, the proofs can be found in the appendix at the end of this paper and in an extended appendix found on our websites.

2. Overview

This section, presents a non-technical roadmap that connects our results with the international finance literature and can be skipped without loss of generality.

The FPP is a violation of the familiar uncovered interest parity condition. In principle, one can account for this violation either via time-varying risk premia or via forecasting distortions. Here, we focus on the latter channel. We use an infinite horizon overlapping generations model where the differential between the domestic and foreign interest rate is the source of misspecification. The interest rate differential process, which is given by (3.1), has a transitory component (v_t) that dies out after one period, and a persistent component (x_t) that dies out only gradually. The investor, however, does not observe them separately.

The agent has a baseline model that corresponds to the data generating process. However, she fears *misspecification* in some parts of the model, which turns out to be the driver in accounting for the FPP. If there was no fear of misspecification, the forecasts would be generated by the Kalman filter and the FPP would not arise.

We derive several change-of-measure lemmas, analogous to the Girsanov theorem, that allow us to represent the no-arbitrage condition in terms of a robust uncovered interest parity condition: the log exchange rate e_t satisfies $E_t^{\theta^*}(e_{t+1}) - e_t = i_t - i_t^f + \zeta_t$. Although this condition has the same form as in familiar international finance models, there are important differences. Expectations $E_t^{\theta^*}(\cdot)$ are taken under a robustly distorted probability measure, and the premium ζ_t reflects not only risk-aversion but also aversion to model uncertainty.

The equilibrium exchange rate takes the familiar linear form (Proposition 4.3)

$$e_t^* = -(i_t - i_t^f) - \sum_{j=1}^{\infty} E_t^{\theta^*}(i_{t+j} - i_{t+j}^f) + \alpha_t^*.$$

The implication of this equation is that the domestic currency appreciates (e_t^* goes down) if there is an increase in current or forecasted interest rate differentials. The summation can be solved in closed-form because in equilibrium the forecasts are given by the sequence $\{a^j \hat{x}_t^{\theta^*}\}$, and the estimate of the persistent component of the differential ($\hat{x}_t^{\theta^*}$) can be computed recursively as a weighted average of the interest rate differential observation and the prior (Lemma 4.2)

$$\hat{x}_t^{\theta^*} = k^{\theta^*}[i_t - i_t^f] + [1 - k^{\theta^*}]a\hat{x}_{t-1}^{\theta^*}.$$

The "gain" k^{θ^*} is endogenously determined and captures the sensitivity of forecasts to news. This gain is the critical determinant of whether or not hump-shaped dynamics arise in equilibrium. We will see that the FPP can be accounted for only if the robust forecasts are less sensitive to news than the Bayesian forecasts associated with the baseline model, i.e. $k^{\theta^*} < k^{Baseline}$. In contrast, if $k^{\theta^*} \geq k^{Baseline}$, the forecast-distortion mechanism cannot account for the FPP.

It turns out that generating a robust gain k^{θ^*}lower than $k^{Baseline}$ as the outcome of agents' optimization is not trivial. When we first tried to tackle this problem we found that simply invoking robustness against any misspecification does not generate $k^{\theta^*} < k^{Baseline}$. The need for refinement of the uncertainty set has lead us to consider three different classes of misspecification of the process (3.1): uncertainty in the observation equation, uncertainty in the persistent component of the differential, and the so called unstructured uncertainty, under which agents fear misspecification of the entire interest rate differential process but cannot pinpoint either its nature or location. We represent these classes of model misspecification via *sets of underlying probability measures.*

The objective function trades off optimality, under the baseline model, versus robustness against model misspecification, as in Hansen et. al. (2006). It selects in a pessimistic way the probability measure under which expectations are computed. To avoid an extreme worst-case scenario outcome, it includes a penalty for deviations from the baseline model –in addition to the standard primitive utility function. This penalty is a proportion λ of the distance between the baseline and the agent's models, which is captured by the *relative entropy* between the baseline probability measure and the robust measure used to compute expectations. The parameter λ is key because $1/\lambda$ indexes the degree of uncertainty-aversion, which is distinct from the traditional risk-aversion.

For each type of misspecification, the results are derived in three steps. First, we define a set of probability measures that captures that type of misspecification. Second, for each set we establish a one-to-one relationship between sets of probability measures and distortions of probability distributions via change-of-measure lemmas, which are analogous to the Girsanov theorem. These lemmas allow us to convert the optimization problem over unknown probability measures into an optimization over a parameter of a pdf, which is much simpler and allows for closed-form solutions for the forecasts and portfolio strategies. Third, we derive in closed-form the equilibrium exchange rate, the slope coefficient of the Fama regression and the impulse response function to news.

Our main results are the following. First, the gain condition $k^{\theta^*} < k^{Baseline}$ may arise under observation uncertainty (Lemma 4.2), but not under the other types of uncertainty. Second, the FPP ($\beta^{Fama} < 1$) obtains if and only if agents are both risk-averse and uncertainty-averse; the strong FPP ($\beta^{Fama} < 0$) obtains if, in addition, the drift a in (3.1) is high and uncertainty aversion $1/\lambda$ is high, but not excessive (Proposition 5.3). Third, there is delayed overshooting –i.e., conditional momentum– under similar conditions as those for $\beta^{Fama} < 0$ (Proposition 5.2). Fourth, if there is misspecification –either additive or parametric– in the persistent component of (3.1), we find that $k^{\theta^*} > k^{Baseline}$ (Propositions 6.2 and 6.4). The resulting equilibrium exchange rate functions imply that the forecast-distortion mechanism cannot generate β^{Fama} less than one. Fifth, when there is unstructured uncertainty, robust forecasts have the same sensitivity to news than the Bayesian forecasts associated to the baseline model (Proposition 6.6). Thus, the forecast-distortion mechanism underlying the anomalies is not operative. Lastly, we do some simulations for the case of observational uncertainty, and find that empirical estimates of key parameters in the model generate a negative Fama coefficient provided there is enough uncertainty-aversion.

3. The Model

We consider a setup that incorporates robustness considerations into a minimal nominal exchange rate model. There are two one-period bonds: a dollar bond that will pay $\exp(i_t)$ dollars next period and a euro bond that will pay $\exp(i_t^f)$ euros. The source of model misspecification is the interest rate differential. It has two components: observation noise (v_t) and an unobservable persistent component, that is stationary, and which we will refer to as the trend (x_t).

$$\begin{aligned} i_t - i_t^f &\equiv y_t = x_t + v_t \\ x_t &= a x_{t-1} + w_{t-1}, \qquad x_0 = 0, \quad a \in (0,1) \end{aligned} \tag{3.1}$$

We allow for both risk and Knightian uncertainty. In the literature, risk refers to the unique probability distribution of the relevant random variables –$\{w_j\}$ and $\{v_j\}$ in our case, which agents either know or can learn. Knightian uncertainty refers to a potential misspecification of the model.

As in Hansen et. al. (2006), we introduce both dimensions of uncertainty by assuming that the representative agent is endowed with a *baseline model* of the interest rate differential. However, she fears *model misspecification* and takes the baseline model simply as an approximation to the data-generating model. In order to measure the distance between models in a simple way we represent model uncertainty in terms of sets of *underlying probability measures*. To this end we will call 'model θ' a probability measure θ defined on a measurable space $(\Omega, \mathcal{B}(\Omega))$, where Ω is a compact metric space and $\mathcal{B}(\Omega)$ is the Borel $\sigma-$ algebra on the space Ω. The agent is endowed with a baseline model, which we will denote by θ', but she takes θ' simply as an approximation and allows for the possibility that the true model θ lies in an *uncertainty set* Θ^j.

In order to obtain the standard Bayes estimator in the special case where agents do not care about robustness, we impose the following condition.

Assumption 1. The baseline model θ' corresponds to the true data-generating process, under which the interest rate differential follows (3.1) and the shock processes are i.i.d. normal random variables

$$w_t \overset{\theta'}{\sim} N\left(0, \sigma_w^2\right), \qquad v_t \overset{\theta'}{\sim} N\left(0, \sigma_v^2\right). \tag{3.2}$$

The robust control literature considers two classes of model uncertainty: structured and unstructured. The former specifies the location and nature of the misspecification. In contrast, under unstructured uncertainty one does not have this refined information. In Sections 4 and 5, we consider a structured uncertainty set that captures misspecification in the observation equation, while keeping the rest of the interest rate differential process (3.1) unchanged. In Section 6 we consider other types of uncertainty. The observation-uncertainty set is given by

$$\Theta^v = \left\{ \theta \in P(\Omega) : \frac{d\theta}{d\theta'} = \exp\left(-\frac{1}{2}\left(\frac{1}{\tilde{\sigma}_v^2} - \frac{1}{\sigma_v^2}\right)(y_t - x_t)^2 \right) \cdot \sqrt{\frac{\sigma_v^2}{\tilde{\sigma}_v^2}}, \ \ \tilde{\sigma}_v^2 \in [\omega, \infty], \ \omega > 0 \right\} \tag{3.3}$$

This is an infinitely large set of probability measures that is generated by letting the parameter $\tilde{\sigma}_v^2$, in the Radon-Nikodim derivative, take on values on the extended positive real line. The lower bound ω is an arbitrarily small number[5].

[5] We set ω instead of 0 as the lower bound for $\tilde{\sigma}_v^2$ so that the set Θ^v is closed. The Gilboa and Schmeidler (1989) type of utility function that we use below requires a closed set of prior probability measures.

There are overlapping generations of two-period lived investors that can take any long and short positions in the dollar bond and the euro bond described above. A representative young investor is willing to sacrifice profitability under the baseline model in exchange for some degree of robustness against misspecification–within the uncertainty set. To formalize this problem, as in Epstein and Schneider (2004), Gilboa and Schmeidler (1989) and Hansen, et. al. (2006), we consider a utility function that pessimistically selects a robust probability measure from the uncertainty set Θ^j.

$$U_t = \inf_{\theta \in \Theta^j} \left\{ E_t^\theta \left[-\exp\left(-\gamma W_{t+1}\right) + \lambda \cdot \Re(\theta || \theta') \right] \right\}, \quad \lambda \geq 0, \ \gamma \geq 0, \tag{3.4}$$

$$\text{where} \quad W_{t+1} \equiv b_t \left[(i_t - i_t^f) - (e_{t+1} - e_t) \right], \quad e \equiv \log(E) \tag{3.5}$$

E denotes the dollar-euro exchange rate, i.e., the number of dollars per euro; and b_t is the young agent's position in the dollar bond. The bracketed term in W_{t+1} is the familiar excess rate of return of the carry trade: the dollar-euro interest rate differential minus the dollar's depreciation rate.[6]

The set Θ^j is a closed and convex set of probability measures and $\Re(\theta||\theta')$ is the *relative entropy* of probability measure θ with respect to the baseline measure θ' :

$$\Re(\theta||\theta') = \begin{cases} \int_\Omega \log\left(\frac{d\theta}{d\theta'}\right) d\theta & \text{if } \theta \text{ is absolutely continuous w.r.t. } \theta' \\ \infty & \text{otherwise.} \end{cases} \tag{3.6}$$

The relative entropy can be thought of as the distance between the baseline model θ' and the alternative model θ.

We consider an equilibrium concept analogous to a rational expectations equilibrium, in which the price reveals all the information available to agents. As is standard in the asset pricing literature we endow the representative agent with a conjecture of the log exchange rate function and a prior about the persistent component of the interest rate

[6] Consider a young representative investor that forms a zero-cost portfolio by taking a position b_t in dollar bonds and a position $-\frac{b_t}{E_t}$ in euro bonds. Next period her payoff, in dollar terms, will be $\exp(i_t)b_t - \frac{E_{t+1}}{E_t}\exp(i_t^f)b_t$. We show in the extended appendix that one obtains W_{t+1} by taking the Taylor expansion of this expression. The second order term is close to zero when i_t and $e_{t+1} - e_t + i_t^f$ are small, which is the case for monthly and quarterly data across the major currency pairs considered in the literature.

differential. The agent's conjecture is

$$e_t^{conj} = \alpha_t + \beta_1 \hat{x}_t^{\theta_t} + \beta_2 \left(i_t - i_t^f \right), \tag{3.7}$$

where $\{\alpha_t, \beta_1, \beta_2\}$ are undetermined coefficients and $\hat{x}_t^{\theta_t} \equiv E_t^{\theta_t}(x_t)$ is the estimate, under probability measure θ_t, of the mean of the interest rate differential's persistent component, conditional on the information available at time $t : I_t = \{y_1, ..., y_t, e_1, ...e_t\}$.

At time t, the prior of the young t-agent is that x_{t-1} is Normally distributed with mean $\hat{x}_{t-1}^{\theta_{t-1}}$ and variance $\sigma_{t-1}^2 = \frac{\left(a^2\sigma_{t-2}^2 + \sigma_w^2\right)\sigma_v^2}{a^2\sigma_{t-2}^2 + \sigma_w^2 + \sigma_v^2}$.[7] Given this prior and conjecture (3.7), the young t-agent solves the following problem.

$$\sup_{b_t \in \mathcal{R}} \inf_{\theta_t \in \Theta^j} \left\{ E_t^{\theta_t} \left[-\exp\left(-\gamma W_{t+1}(b_t,\ e_t,\ e_{t+1}^{conj}) \right) + \lambda \cdot \Re(\theta_t || \theta') | I_t \right] \right\},\ I_t = \{y_1, ...y_t, e_1, ...e_t\} \tag{3.8}$$

That is, using the information I_t the agent updates her estimate of x_t and x_{t+1} under the robust probability measure θ_t, that solves (3.8), and chooses her portfolio b_t. We will denote the solution to the representative agent's problem (3.8) by $b_t^* \left(e_t \mid \alpha_{t+1}^*, \alpha_t^*, \beta_1^*, \beta_2^* \right)$. Since young agents have measure one, b_t^* is the aggregate demand for domestic bonds.

A "robust linear equilibrium" is an exchange rate function $e_t^* = \alpha_t^* + \beta_1^* \hat{x}_t^{\theta_t^*} + \beta_2^* \left(i_t - i_t^f \right)$ that is consistent with conjecture (3.7) and that clears the market, i.e., $b_t^* \left(e_t^* | \alpha_{t+1}^*, \alpha_t^*, \beta_1^*, \beta_2^* \right) = b_t^s$.

3.1. Discussion of the Setup

In order to investigate how robustness accounts for the FPP, we need a tractable model with a closed-form solution of the Fama regression coefficient. To this end we have considered a minimal setup. One could think of other setups where the structure of the economy is more complicated. However, we believe that considering a minimal model that yields tractable solutions will allow us to better characterize the forecast-distortion mechanism that generates the FPP.

The utility function given in (3.4) distinguishes *risk aversion*–parametrized by γ– from *uncertainty aversion*, which is parametrized by $1/\lambda$. As we will show, both *risk aversion* and *uncertainty aversion* are necessary to account for the anomalies.

[7] This formula comes from the Kalman filter under baseline measure θ'.

To interpret utility function (3.4) we can think of nature as choosing the true probability measure θ in a malevolent way, so as to minimize the agent's utility.[8] Note, however, that by increasing the distance between model θ and the baseline model θ'–proxied by $\Re(\theta||\theta')$–nature incurs a penalty $\lambda \cdot \Re(\theta||\theta')$. The greater the uncertainty aversion $(1/\lambda)$, the lower the penalty for deviating from the baseline model. In one extreme, if $1/\lambda$ were zero, (3.4) would reduce to the familiar expected utility function $-E_t^{\theta'} \exp(-\gamma W_{t+1})$. In this case, the agent would not allow for model misspecification because choosing any model θ different from θ' would make (3.4) infinite. In the other extreme, if $1/\lambda$ were infinity, the agent would choose the worst case model among all possible models. This worst-case objective is too paranoid and leads to overly conservative strategies. In these two extreme cases, $\lambda = \{0, \infty\}$, robustness against misspecification does not generate the hump-shaped dynamics necessary for the FPP.

The inclusion of the relative entropy in the objective is frequently used in the literature on the theory of large deviations and information theory. We can think of relative entropy as a distance between two probability measures on $P(\Omega)$, the set of all probability measures on Ω. It is always non-negative, and it is zero if and only if $\theta = \theta'$.[9] Moreover, if we consider a new measure $\eta = \tau\theta + (1-\tau)\theta'$, then $\Re(\eta||\theta') \leq \Re(\theta||\theta')$.[10]

Gilboa and Schmeidler (1989) show that in the presence of Knightian uncertainty, preferences can be represented by a utility function of the form: $U = \inf_{\theta \in \Theta} u$, where Θ is a closed and convex set of *underlying probability measures* and u is an affine function. The objective function in (3.4) belongs to this class of functions because the sets Θ^j of probability measures that we consider are closed and convex under the weak* topology, and the term in brackets is a standard expected utility function as $-E_t^{\theta} \exp(-\gamma W_{t+1}) + \lambda \Re(\theta||\theta') = E_t^{\theta}\left[-\exp(-\gamma W_{t+1}) + \lambda \log(d\theta/d\theta')\right]$.

We have considered two-period lived agents. If agents had infinite horizons, we could not have a constant λ as this would generate time inconsistency. Epstein and Schneider (2003) extend to a multiperiod setup the framework of Gilboa and Schmeidler (1989), and discuss the issues involved in considering long-lived agents.

[8] Obviously, nature does not care about our forecasts. This device is simply a useful way to induce robustness.

[9] Define $\theta' = \theta$ if $\theta'(A) = \theta(A)$ for all $A \in \mathcal{B}(\Omega)$.

[10] The relative entropy is not a metric on the space $\mathcal{P}(\Omega)$ because it does not satisfy the triangle inequality.

Although our agents' baseline model corresponds to the data generating process, when making portfolio decisions they do not estimate the state using this process because of robustness considerations. Therefore, it would seem to an outside observer that our agents refuse to learn the parameters of the data generating process.

Finally, the reader might recognize a discrepancy as we have described the intuition in Section 2 via distortions of the gain k (equivalently, misspecifications to the probability distributions of disturbances), while the objective function we consider is defined in terms of underlying probability measures. This former representation is attractive because, arguably, distortions to means and variances of probability distributions can be estimated by an econometrician, they are intuitive, and it is the way in which the empirical evidence is presented in the literature. In contrast, underlying probability measures are abstract, unobservable concepts. However, they are conceptually more convenient because they allow us to treat different types of uncertainty in a unified framework as well as define the agent's objective concisely. We can do this translation because we establish a one-to-one mapping between probability measures and the gain k. For instance, under observational uncertainty every probability measure $\theta \in \Theta^v$ can be represented in terms of a specific value of the variance of the observational noise, and there is a one-to-one mapping between this variance and the gain k. An attractive property of the uncertainty set Θ^v is that it captures observational uncertainty in a particularly useful way that allows for the derivation of a tractable closed-form solution.

4. Derivation of the Equilibrium

We derive the equilibrium in two steps. First, we solve the agent's problem (3.8) for a given exchange rate e_t. Then, we derive the exchange rate function that equilibrates the market and that is consistent with the conjecture (3.7).

4.1. The Joint Forecasting-portfolio Problem

The investor's problem (3.8) cannot be solved by applying the well known separation principle under which the forecasts are derived using Bayes law, independently of portfolio optimization. Instead, we need to consider a joint forecasting-portfolio problem. We will solve this joint problem by treating it as a zero-sum game between the investor and nature. The investor makes her estimate of the interest rate differential trend and chooses her portfolio (b_t) accounting for the strategy of nature. Conditional on the choice of the investor nature chooses the probability measure θ_t in a malevolent way.

This problem is rather complicated as the investor must optimize over a set of unknown probability measures. We convert this problem into a parametric problem that determines a variance distortion of a probability distribution rather than a robust probability measure. We do this transformation by deriving a change-of-measure lemma, which is analogous to the Girsanov theorem.[11] For a given random variable, the lemma: (i) establishes a one-to-one mapping between the variance of the probability distribution and the underlying probability measure; and (ii) provides a parametric formula for the relative entropy $\Re(\theta||\theta')$.[12]

Lemma 4.1 (Change of Measure I). *If under baseline probability measure θ' the random variables in the interest rate differential process (3.1) are distributed as $x_t|I_{t-1} \overset{\theta'}{\sim} N(a\hat{x}_{t-1}, \sigma_w^2)$, $y_t|x_t \overset{\theta'}{\sim} N(x_t, \sigma_v^2)$ and $x_{t-1} \overset{\theta'}{\sim} N\left(\hat{x}_{t-1}, \sigma_{t-1}^2\right)$, then for any probability measure $\theta \in \Theta^v$:*

1. *The distribution of the observation satisfies $y_t|x_t \overset{\theta}{\sim} N\left(x_t, \tilde{\sigma}_v^2\right)$, while the distributions of $x_t|I_{t-1}$ and x_{t-1} are preserved.*

2. *The relative entropy of probability measure θ with respect to the baseline θ' equals*

$$\Re(\theta||\theta') = \frac{1}{2}\left(\frac{\tilde{\sigma}_v^2}{\sigma_v^2} - \log\left(\frac{\tilde{\sigma}_v^2}{\sigma_v^2}\right) - 1\right) \qquad \text{for any } \theta \in \Theta^v \tag{4.1}$$

The first part of Lemma 4.1 says that if under baseline model θ' the interest rate differential process is given by (3.1), and random variables v_t and w_t are normally distributed as in (3.2), then under an alternative model $\theta \in \Theta^v$ only the variance of the

[11] The Girsanov theorem applies to a change of drift. Here we are interested in a change of variance.

[12] A random variable v is a measurable function that maps an abstract underlying measurable space $(\Omega, \mathcal{I})$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, where $\mathcal{B}(\mathbb{R})$ is the Borel σ-algebra. An underlying probability measure θ on $(\Omega, \mathcal{I})$ determines the probability of v belonging to a set $A \in \mathcal{B}(\mathbb{R})$, which is defined by $\Pr(v \in A) \equiv \theta(\{\omega \in \Omega : v(\omega) \in A\} \in \sigma(v))$. In our setup, there exists a one-to-one correspondence between the probability distribution and the underlying measure on the space we consider for the following reasons. First, the distribution function of random variable v is a non-decreasing function $F : \mathbb{R} \to [0,1]$ such that $\Pr(v \in A) = \int_A dF(x)$ for every set $A \in \mathcal{B}(\mathbb{R})$. By definition the induced distribution function F is determined by the underlying probability measure θ almost everywhere. Second, if we know the distribution function of v, we can retrieve the underlying probability measure θ with the formula $\theta(B) = \int_{v(B)} dF(x)$, $B \in \sigma(v)$. This measure is unique except in sets with zero measure.

observation equation is altered from σ_v^2 to $\tilde{\sigma}_v^2$, while the rest of the process remains unchanged. Equation (4.1) says that the relative entropy between measures θ and θ' is zero when there is no variance distortion, and that it increases at an increasing rate as the distortion grows in either direction. This equation will prove quite useful because it defines the distance between models only in terms of the variance distortion.

Solution to the Investor's Problem

Using the change-of-measure Lemma 4.1, we show in the appendix that problem (3.8) is equivalent to the following problem.

$$\max_{b_t \in R} \inf_{\tilde{\sigma}_{v,t}^2 \in (\varepsilon,\infty)} \left[-\exp\left(-\gamma b_t E_t^{\theta_t}\left(J_{t+1}\right) + \frac{\left(\gamma b_t\right)^2}{2} Var_t^{\theta_t}\left(J_{t+1}\right) \right) + \tfrac{\lambda}{2}\left(\frac{\tilde{\sigma}_{v,t}^2}{\sigma_v^2} - \log\frac{\tilde{\sigma}_{v,t}^2}{\sigma_v^2} - 1 \right) \right]. \tag{4.2}$$

$J_{t+1} \equiv (i_t - i_t^f) - (e_{t+1} - e_t)$ is the log excess return, $E_t^{\theta_t}\left(J_{t+1}\right)$ is the conditional mean and $V_t^{\theta_t}\left(J_{t+1}\right)$ is the conditional variance of J_{t+1} under probability measure $\theta_t \in \Theta^v$. They are given by (8.1) in the appendix. The attractive property of (4.2) is that it has converted an optimization problem over unknown probability measures to a much simpler parametric one over the variance distortion $\tilde{\sigma}_{v,t}^2$. The solution to this problem is given by the following Lemma.

Lemma 4.2 (Solution to the Portfolio-Forecasting Problem). *In the presence of observational uncertainty, i.e., $\theta_t \in \Theta^v$, Problem (4.2) has a solution only if uncertainty aversion $1/\lambda$ is not greater than a threshold $1/\lambda_t^v$ given by (8.4). In this case:*

1. *The estimate of the unobservable component of the interest rate differential is given by the Kalman filter under the robust probability measure θ_t*

$$\hat{x}_{t+1}^{\theta_t} \equiv E_t^{\theta_t}\left(x_{t+1}\right) = a\hat{x}_t^{\theta_t}, \qquad \text{with } \hat{x}_t^{\theta_t} = \left(1 - k_t^{\theta_t}\right) a\hat{x}_{t-1}^{\theta_{t-1}} + k_t^{\theta_t}\left(i_t - i_t^f\right). \tag{4.3}$$

The 'gain' of the filter $k_t^{\theta_t}$ is $k_t^{\theta_t} \equiv \frac{a^2\sigma_{t-1}^2+\sigma_w^2}{a^2\sigma_{t-1}^2+\sigma_w^2+\tilde{\sigma}_{v,t}^2}$, with $\sigma_{t-1}^2 = \frac{\left(a^2\sigma_{t-2}^2+\sigma_w^2\right)\sigma_v^2}{a^2\sigma_{t-2}^2+\sigma_w^2+\sigma_v^2}$. The distorted observational variance $\tilde{\sigma}_{v,t}^2$ is given by (8.5).

2. *The robust forecast of the interest rate differential are less sensitive to news than the baseline forecasts: $k(\tilde{\sigma}_{v,t}^2) < k(\sigma_v^2)$ for any $\lambda \in (\lambda_t^v, \infty)$.*

3. *The demand for the domestic bond is ($Var_t^{\theta_t}(J_{t+1})$ below is given by (8.1))*

$$b_t(e_t, \tilde{\sigma}_{v,t}^2) = \frac{\left(i_t - i_t^f\right) - \left(E_t^{\theta_t} e_{t+1} - e_t\right)}{\gamma \cdot Var_t^{\theta_t}(J_{t+1})}. \tag{4.4}$$

4. *There is no forecast distortion if either the primitive utility function is risk neutral or there is no aversion to uncertainty, i.e., $\tilde{\sigma}_{v,t}^2 \to \sigma_v^2$ if $\gamma \to 0$ or $\lambda \to \infty$.*

The solution is quite intuitive. Part 1 says that the forecasts are given by the celebrated Kalman filter, which gives a weight $k_t^{\theta_t}$ to the current interest rate differential and $1 - k_t^{\theta_t}$ to the prior estimate. Part 2 implies that under observation uncertainty the robust gain $k_t^{\theta_t}$ must be smaller than the Bayesian gain associated with the baseline model θ'. This result holds because the variance of the observation noise is distorted upwards under the robust probability measure θ_t: $\tilde{\sigma}_{v,t}^2 > \sigma_v^2$.[13] As we will see, this forecast distortion is the key to accounting for the anomalies. The form of the demand in part 3 is standard in the literature. The numerator is the expected excess rate of return, and the denominator is the risk aversion coefficient times the variance of returns. A non-standard aspect of the demand is that the moments of returns are computed under the robust probability measure θ_t, which need not be the same as the baseline measure θ'. Part 4 says that both risk aversion and uncertainty aversion are necessary for distorted forecasts. In particular, a desire for robustness will not generate the FPP if agents have a risk neutral primitive utility function. Finally, the lemma implies that in general the effects of misspecification can be ameliorated though not entirely eliminated. This is because for any degree of uncertainty aversion greater than $1/\lambda_t^v$, the game between the agent and nature ceases to be convex, and so there is no solution to the portfolio-forecasting problem.

4.2. Equilibrium Exchange Rate

Here, we compute the equilibrium. From the perspective of the FPP, the main question is whether the distortion in the representative investor's forecasts carries over to the equilibrium exchange rate.

[13] The result $\tilde{\sigma}_{v,t}^2 > \sigma_v^2$ means that the robust agent considers a probability density function for y_t with fatter tails, but the same mean, than the standard Normal. This result does not depend on the particular value of the baseline variance σ_v^2.

If we substitute the market clearing condition $b_t(e_t^*, \tilde{\sigma}_{v,t}^2) = b_t^s$ in the first order condition for the holdings of the bond b_t, we obtain the *robust interest parity condition.* Namely, the depreciation rate expected by robust agents equals the interest rate differential $(i_t - i_t^f)$ minus an uncertainty premium on domestic assets (ζ_t)

$$\begin{aligned} E_t^{\theta_t^*}(e_{t+1}) - e_t &= (i_t - i_t^f) - \zeta_t, \qquad (4.5) \\ \text{where } \zeta_t &\equiv \gamma b_t^s Var_t^{\theta_t^*}(J_{t+1}) = \gamma b_t^s \left(\frac{a k_{t+1}^{\theta_{t+1}^*}}{1-a} + 1\right)^2 (a^2 k_t^{\theta_t^*} \tilde{\sigma}_{v,t}^{2*} + \sigma_w^2 + \tilde{\sigma}_{v,t}^{2*}) \end{aligned}$$

Notice that (4.5) has been characterized only in terms of the distorted variance $\tilde{\sigma}_{v,t}^2$. The abstract robust probability measure θ_t does not enter (4.5) directly, but only through the distorted variance $\tilde{\sigma}_{v,t}^2$. The reason for this simplicity is that there is a one-to-one map between $\tilde{\sigma}_{v,t}^2$ and θ_t established by Lemma 4.1. This transformation will prove useful because it allows us to describe the robust interest parity condition simply in terms of distorted variances, which can be linked to the empirical literature on the FPP.

To obtain the market clearing exchange rate, we substitute conjecture (3.7) in (4.5), and impose the fixed point condition that the market clearing exchange rate equals the conjecture.

Proposition 4.3. *In a robust linear equilibrium, the log exchange rate is*

$$e_t^* = -\left(i_t - i_t^f\right) - \frac{a}{1-a}\hat{x}_t^{\theta_t^*} + \alpha_t^*, \qquad \alpha_t^* = \alpha_{t+1}^* + \gamma b_t^s Var_t^{\theta_t^*}(J_{t+1}). \qquad (4.6)$$

The robust forecast of the interest rate differential $\hat{x}_{t+1}^{\theta_t^*} = a\hat{x}_t^{\theta_t^*}$ is given by the filter (4.3) and the variance of returns $Var_t^{\theta_t^*}(J_{t+1})$ is given by (8.1).

The exchange rate function (4.6) is intuitive. The first two terms imply that the domestic currency appreciates (i.e., there is a fall in the number of Dollars per Euro) if there is an increase in either the current interest rate differential $(i_t - i_t^f)$ or its forecasts $(\sum_{i=1}^{\infty} \hat{x}_{t+i}^{\theta_t^*} = \sum_{i=1}^{\infty} a^i \hat{x}_t^{\theta_t^*} = \frac{a}{1-a}\hat{x}_t^{\theta_t^*})$. The third term α_t^* can be interpreted as the long-run exchange rate.[14]

For the purpose of accounting for the anomalies, the key aspect of (4.6) is that the distortion of individual forecasts is carried over to the equilibrium exchange rate. This is

[14] In order for α_t^* to be bounded it is necessary that $lim_{t\to\infty}\gamma b_t^s Var_t^{\theta_t^*}(J_{t+1}) = 0$.

because the estimate of the unobservable trend ($\hat{x}_t^{\theta_t^*}$), that enters (4.6), is derived under the robust probability measure θ_t^*, not under the baseline measure θ'.

4.3. Intuition for Distorted Forecasts: the Robustness Principle

This subsection, provides intuition for distorted forecasts as a means to attain robustness and can be skipped without loss of continuity.

Why should robustness against observation uncertainty imply a forecast distortion? There is, what we might term, a "robustness principle" at work. In designing her strategy the agent asks: if things were to go wrong, what would be the costliest direction in which they could go wrong? She then trades off the benefits of dampening the effects of misspecification in this costliest direction, against the costs of moving away from "optimality under the baseline model." This trade-off is weighted by the degree of uncertainty aversion $1/\lambda$.

To determine the distortion that agents will introduce we need to apply this robustness principle to a specific uncertainty set. If agents fear misspecification in the equation that links observations and the unobservable trend of the interest rate differential, the costliest misspecification occurs when the agent's model incorrectly sets the observation variance lower than what it actually is. Therefore, under observation uncertainty robustness entails distorting upwards the observation variance, which in turn implies less sensitivity to news.

To illustrate the core of the mechanism that generates a lower sensitivity to news, let us decouple the forecast problem from the portfolio problem. Consider a one-period filtering problem where the observation is $y = x + v$, the unobservable component follows $x = ac + w$, the shocks w and v are independently distributed random variables with variances d and σ^2, and $var(c) = 0$. In this example, the agent's primitive objective is to find an estimator $\hat{x}$ to minimize the mean squared error (MSE) of her state estimate $E^\theta (x - \hat{x})^2$. The solution to this filtering problem entails an estimator $\hat{x}$ of the form $\hat{x} = ky + (1 - k)\, ac$. Thus, the true mean squared error (MSE) equals $E^\theta(x - \hat{x})^2 = k^2 \cdot \sigma^2 + (1 - k)^2 \cdot d$.[15] Here, θ is the probability measure that the agent chooses. If the agent fears no model uncertainty, then $\theta = \theta'$ and we obtain the celebrated Kalman

[15] We don't include the term $(1 - k)^2 a^2 var(c)$ because $var(c) = 0$.

gain: $k^{\theta'} = \frac{d}{d+\sigma^2}$.[16] If the agent fears misspecification of the observation equation (i.e., $\theta \in \Theta^v$), what gain k shall she choose? Since the misspecification is unknown, the agent considers all potential misspecifications within the uncertainty set. This is why the structure of the uncertainty is key. When $\theta \in \Theta^v$, the agent can make two type of errors. Error I occurs when she sets σ too high and so chooses k lower that what it should be under the true model. Error II occurs when she sets σ too low. The agent then asks, if things were to go wrong, which error is more costly? The key point is that when $\theta \in \Theta^v$, it is more harmful in terms of the MSE to mistakenly believe that σ is *low* than to mistakenly believe σ is *high*.[17] Therefore, the robust agent focuses on the type II error and guards against the possibility of understating σ leading her to set the gain lower than the baseline gain $k^{\theta'}$ so as to be robust against observational uncertainty.

The above argument explains why there is a lower sensitivity to news. But how far shall a robust agent underreact? The answer depends on the penalty for deviating from the baseline $\lambda \cdot \Re(\theta||\theta')$. To see the intuition consider the case in which the agent is infinitely uncertainty averse ($\lambda = 0$), so her estimation objective is to minimize $\sup_{\theta\in\Theta^v} E^\theta(x - \hat{x})^2$, and suppose for a moment that σ can take values only on $[\underline{\sigma}, \bar{\sigma}]$. In this case the agent would set the gain to $k = \frac{d}{d+\bar{\sigma}^2} < k^{\theta'}$. This way, she would bound the maximum MSE regardless of the value of σ and would actually attain the lowest upper bound on the MSE.[18] Notice, however, that the uncertainty set Θ^v does not impose a hard upper bound on σ, and allows σ to take any value in the extended positive real line. Thus, the simplistic worse-case analysis would lead the agent to guard against $\bar{\sigma} = \infty$ and set $k = 0$. This belief is quite paranoid. The role of the cost $\lambda \cdot \Re(\theta||\theta')$ is to dampen this paranoia. Instead of imposing hard bounds on σ, we penalize the fictitious malevolent nature for choosing a probability measure θ different from the baseline θ'. When $\theta \in \Theta^v$, this cost grows at an increasing rate as the distorted variance moves away

[16] In a Gaussian setup Bayes law leads to the same result.

[17] Because the growth rate of the MSE is $\frac{dMSE}{d\sigma^2} = k^2 = \left(\frac{d}{d+\sigma^2}\right)^2$, the MSE grows at a lower rate when a higher σ^2 is chosen. That is, for $\sigma_1 < \sigma_2$, we have $\frac{dMSE}{d\sigma^2}(\sigma_1) > \frac{dMSE}{d\sigma^2}(\sigma_2)$. Therefore, it is less harmful to chose σ_2 rather than σ_1.

[18] To clarify the mechanics suppose for a moment that σ is drawn from a pdf $f(\sigma)$ with support $[\underline{\sigma}, \bar{\sigma}]$. The robust agent disregards the information contained in $f(\sigma)$ and sets $k = \frac{d}{d+\bar{\sigma}}$. Notice that even if the baseline were $\bar{\sigma} - \varepsilon$ and $\bar{\sigma}$ were far away from $\underline{\sigma}$, the agent would choose k for any $\varepsilon > 0$. This is because her objective is to bound the worst-case damage.

in either direction from the baseline.[19] Since the benefit to nature grows linearly at rate k^2 in terms of a greater MSE, nature faces a concave objective and thus sets σ at a bounded level, but greater than its baseline which is determined by λ.

Lemma 4.2 considers a more complex problem where the forecast and portfolio problems must be solved jointly. However, the gist of the argument is the same as the one we have used to illustrate the intuition. Although the formula for the variance distortion $\tilde{\sigma}^2_{v,t}$ is more complex, it has the same two properties as the distortion σ in the simple example above: $\tilde{\sigma}^2_{v,t}$ is greater than the baseline level and it is inversely related to the degree of uncertainty aversion $1/\lambda$. In fact, $\tilde{\sigma}^2_{v,t}$ tends to infinity as λ approaches the lowest permissible bound λ^v_t. Notice that in general λ^v_t cannot be zero and so absolute robustness against observation uncertainty cannot be attained. This is because below the threshold λ^v_t the objective function ceases to be convex in $\tilde{\sigma}^2_{v,t}$, and so the problem has no solution.

5. Foreign Exchange Market Anomalies

There are two channels through which the equilibrium exchange rate (4.6) might generate the FPP and hump-shaped dynamics: forecast-distortions and time-varying risk premia. Here, we focus on the first channel and we shut off the latter channel. We find that when there is no uncertainty aversion, forecasts are generated by Bayes law under the baseline model, the Fama coefficient is one and there is no delayed overshooting. In contrast, when uncertainty aversion is high, a negative Fama coefficient arises under observation uncertainty provided the interest rate differential is highly persistent.

To concentrate on the forecast-distortion mechanism we consider the case where the gain k_t in (4.3) has converged and α^*_t in (4.6) is deterministic. To do so we set the supply of domestic bonds equal to a constant up to some time T, where T can be very large.

[19] We know from Lemma 4.1 that if $\theta \in \Theta^v$, the relative entropy equals $\Re\left(\theta||\theta'\right) = \frac{1}{2}\left(\frac{\tilde{\sigma}^2_v}{\sigma^2_v} - \log\left(\frac{\tilde{\sigma}^2_v}{\sigma^2_v}\right) - 1\right)$. Since $\frac{d\Re\left(\theta||\theta'\right)}{d\frac{\tilde{\sigma}^2_v}{\sigma^2_v}} = \frac{1}{2}(1 - \frac{1}{\frac{\tilde{\sigma}^2_v}{\sigma^2_v}})$, the growth rate of the relative entropy is increasing in the ratio $\frac{\tilde{\sigma}^2_v}{\sigma^2_v}$. Therefore, the punishment to nature increases at a faster rate as $\tilde{\sigma}^2_v$ moves further away from σ^2_v.

Assumption 2. The supply of domestic bonds follows

$$b_t^s = \begin{cases} \bar{b}, & t \leq T \\ -\bar{b}\exp(T-t)\,\phi_t, & t \in (T,\infty), \quad \phi_t \equiv \left(V_t^{\theta_t}(J_{t+1})\right)^{-1} \end{cases} \tag{5.1}$$

The next Lemma combines this assumption with the well known fact in the control literature that the gain $k_t^{\theta^*}$ of the filter (4.3) converges rather fast if the underlying coefficients $(a, \tilde{\sigma}_v^2)$ are constant.

Lemma 5.1. *If the bonds' supply follows (5.1) and initial time is in the infinite past, then on $t \in [0,T]$ the equilibrium exchange rate (4.6) has a deterministic α_t^* (given by (8.10)), and a constant gain in $\hat{x}_t^{\theta^*}$, given by*

$$k^{\theta^*} = \frac{a^2\xi^* + \sigma_w^2}{a^2\xi^* + \sigma_w^2 + \tilde{\sigma}_v^{2*}}, \qquad \xi^* = \frac{-\left(\sigma_w^2 + \sigma_v^2 - a^2\sigma_v^2\right) + \sqrt{\left(\sigma_w^2 + \sigma_v^2 - a^2\sigma_v^2\right)^2 + 4a^2\sigma_w^2\sigma_v^2}}{2a^2}. \tag{5.2}$$

The distorted variance $\tilde{\sigma}_v^{2*}$ is a constant given by (8.7) and λ^v, the lower bound for λ, is given by (8.4).

Figure 1 makes clear the dependence of k^{θ^*} on the degree of uncertainty aversion $(1/\lambda)$. We can see that the higher uncertainty aversion (lower λ), the higher the distorted observational variance $\tilde{\sigma}_v^{2^*}$ relative to the baseline σ_v^2, and the lower the robust gain relative to the baseline gain.

5.1. Delayed Overshooting (Conditional Momentum)

As a preliminary step in explaining the FPP, we investigate the conditions under which the exchange rate exhibits a hump-shaped response conditional on the occurrence of a once-and-for-all shock to the interest rate differential, i.e., the forward premium.[20] That is, a positive forward premium shock generates an initial appreciation of the domestic currency, which is followed by further appreciation for several periods afterwards before reverting to a depreciating path. This "delayed overshooting" pattern, or conditional

[20] The covered interest parity condition implies that the interest rate differential equals the forward premium: $f_t - e_t = i_t - i_t^f$, where f_t is the log one period ahead forward exchange rate and e_t is the log spot exchange rate.

momentum, is consistent with the FPP because there is a period during which an exchange rate appreciation coexists with a positive interest rate differential.[21]

To compute the impulse response to a random forward premium shock, suppose that at time $t = 1$ the representative agent observes a forward premium realization $y_1 = 1$ and suppose no shocks occur afterwards. She knows that the forward premium shock is generated by a combination of a transitory shock $v_1 = \kappa$ and a persistent shock $w_0 = \varepsilon$ such that $y_1 = \varepsilon + \kappa = 1$, but she does not observe the particular values of ε and κ.

Since the data is generated by baseline model θ', we define the 'average' impulse response at time t to an initial y-shock as follows

$$e_t^{av} \equiv E^{\theta'}\left(\hat{e}_t\left(\varepsilon,\kappa\right)\right)|_{y_1=\varepsilon+\kappa=1}, \quad \text{with } \hat{e}_t\left(\varepsilon,\kappa\right) \equiv e_t\left(\varepsilon,\kappa\right) - e_t(0,0). \tag{5.3}$$

The term $\hat{e}_t(\varepsilon,\kappa)$ is the response of the log exchange rate at time t to an initial persistent shock ε and an initial transitory shock κ. That is, the time t response to shock sequences $w^s = \{\varepsilon, 0, ..., 0\}_{1\times t}$ and $v^s = \{\kappa, 0, ..., 0\}_{1\times t}$. We must consider the average response to account for delayed overshooting and not just a response to a persistent shock because the agent and the econometrician cannot condition on whether the shocks are transitory or persistent.

Since in equilibrium the exchange rate and the forecasts are linear in the initial shocks ε and κ, we can express the average impulse response (5.3) as a weighted average of the responses to a persistent shock $\hat{e}_t(\varepsilon,0)$ and to a transitory shock $\hat{e}_t(0,\kappa)$:

$$e_t^{av} = q^{\theta'} \cdot \hat{e}_t(\varepsilon,0) + [1 - q^{\theta'}] \cdot \hat{e}_t(0,\kappa), \qquad q^{\theta'} \equiv \frac{1}{1+\frac{\sigma_v^2}{\sigma_w^2}} \tag{5.4}$$

Lemma 8.1 in the appendix shows that the weight $q^{\theta'}$ is the expected value under the data generating model θ' of the persistent shock ε conditional on $y_1 = 1$. The greater $q^{\theta'}$, the greater the share of persistent shocks in the data. As we can see, $q^{\theta'}$ is decreasing in $\frac{\sigma_v^2}{\sigma_w^2}$, the noise-to-signal ratio of the data generating model.

The next Proposition states the conditions under which the average impulse response

[21] This pattern stands in contrast to the overshooting pattern of Dornbusch (1976), and has been found in the data by Eichenbaum and Evans (1995). Empirically, it is more difficult to find delayed overshooting than the FPP. To identify delayed overshooting, it is necessary to determine when a shock to the interest rate differential occurs, an event that is hard to single out in the data.

appreciates at time $t = 1$, and continues appreciating until some time τ after which it reverts back to its long run level. That is, whether there exists an integer $\tau \geq 2$, such that $e_t^{av} - e_{t-1}^{av} < 0$ for all integers $t \in (1, \tau]$ and $e_{t+1}^{av} - e_t^{av} > 0$ for all $t \geq \tau + 1$ impulse response of the exchange rate to an interest rate differential shock at $t = 1$

Proposition 5.2 (Delayed Overshooting). *In the robust equilibrium, the average satisfies:* $e_1^{av} - e_0 = -\frac{1-a\left(1-k^{\theta^*}\right)}{1-a}$ *and for* $t \geq 1$:

$$e_{t+1}^{av} - e_t^{av} = \frac{a^{t-1}}{1-a}\left[\left(k^{\theta^*} - q^{\theta'}\right) a\left(1 - a\left(1 - k^{\theta^*}\right)\right)\left(1 - k^{\theta^*}\right)^{t-1} + (1-a)\, q^{\theta'}\right] \tag{5.5}$$

i. *There is delayed overshooting if the trend of the interest rate differential is highly persistent (a is high), agents are risk-averse ($\gamma > 0$) and uncertainty aversion ($1/\lambda$) is large, but smaller than $1/\lambda^v$, so that the robust gain k^{θ^*} is positive and lower than the share of persistence shocks in the data $q^{\theta'}$.*

ii. *Delayed overshooting occurs up to a time τ, after which mean-reversion takes place. Time τ is the smallest integer that satisfies*

$$\tau \geq 1 + \frac{\log\left(q^{\theta'}[\frac{1}{a} - 1]\right) - \log\left([1 - a(1 - k^{\theta^*})][q^{\theta'} - k^{\theta^*}]\right)}{\log\left(1 - k^{\theta^*}\right)} \tag{5.6}$$

To see the intuition recall that the exchange rate is a function of the estimate of the hidden trend $\hat{x}_t$, and that the response of $\hat{x}_t$ to forward premium news is determined by the gain k^{θ^*} because $\hat{x}_t = k^{\theta^*} y_t + (1 - k^{\theta^*}) a \hat{x}_{t-1}$. When a positive forward premium shock occurs, the agent's robust estimate initially reacts on average less to the news because $k^{\theta^*} < k^{\theta'}$, and so the exchange rate does not appreciate much at $t = 1$. As a result, forecasts will latter on have to catch up. This catch-up will be strong enough to generate momentum in the exchange rate if persistent shocks are long lasting (a is high) and sensitivity to news is low enough ($k^{\theta^*} < q^{\theta'}$). If instead a were small, even persistent shocks would disappear fast, so the initial small reaction would not lead to momentum. If k^{θ^*} were not small enough, most of the reaction would occur initially, and so subsequent forecast revisions would be very small and would be dominated by the inherent mean reversion of the forward premium. In particular, it is necessary that $k^{\theta^*} < q^{\theta'}$. Since momentum cannot be generated from transitory shocks, the initial forecast reaction to the shock $\left(k^{\theta^*}\right)$ must be less than the expected value of the persistent shock $q^{\theta'}$.[22]

Robustness against observation uncertainty generates a lower gain k^{θ^*} by inducing agents to distort upwards the variance of observation shocks relative to the data generating process θ' (i.e., $\tilde{\sigma}_v^{2*} > \sigma_v^2$). This variance distortion is in turn determined by the degree of uncertainty aversion $1/\lambda$. If there is no uncertainty aversion, $1/\lambda = 0$, there is no distortion as $\tilde{\sigma}_v^{2*}$ must be equal to σ_v^2 and the robust gain k^{θ^*} must equal the baseline Bayes gain $k^{\theta'}$. In this case there cannot be momentum because $k^{\theta'}$ is necessarily greater than $q^{\theta'}$.[23] In other words, rational agents with no fear of misspecification cannot systematically be fooled. In contrast, we can ensure that $k^{\theta^*} < q^{\theta'}$ by letting λ go towards its lower bound λ^v because $\tilde{\sigma}_v^{2*} \to \infty$ and so $k^{\theta^*} \to 0$.

To illustrate the results in Proposition 5.2, Figure 2 exhibits the impulse response functions to various shocks to the interest rate differential. As we can see, under the baseline model the average impulse response function appreciates on impact and immediately reverts to a depreciating path. In contrast, under the robust model there is delayed overshooting as the exchange rate continues appreciating after the initial jump.

5.2. The Forward Premium Puzzle (FPP)

Consider the "Fama regression" $e_{t+1} - e_t = \alpha + \beta^{Fama}(i_t - i_t^f) + u_t$. Under the null of uncovered interest parity and rational expectations the slope (Fama) coefficient β^{Fama} is one. However, many empirical studies find that $\hat{\beta}^{Fama} < 1$ (weak FPP) and often $\hat{\beta}^{Fama} < 0$ (strong FPP). The next proposition states the conditions under which the forecast-distortion mechanism generates a theoretical $\beta^{Fama} < 0$, given that the interest rate differential is generated by baseline model θ', but agents use the robust model θ^* that solves their portfolio problem.

[22] The response to a purely transitory shock is $\hat{e}_1(0,\kappa) = -\kappa - \frac{a}{1-a}k^{\theta^*}\kappa$ for $t = 1$ and $\hat{e}_t(0,\kappa) = -\frac{a^t k^{\theta^*}\left(1-k^{\theta^*}\right)^{t-1}}{1-a}\kappa$ for $t \geq 2$. This response does not exhibit delayed overshooting for any $a \in (0,1)$. The response to a purely persistent shock is $\hat{e}_t(\varepsilon, 0) = -\frac{a^{t-1}}{1-a}\left(1 - a\left(1-k^{\theta^*}\right)^t\right)\varepsilon$ for $t \geq 1$. It follows that for any $k^{\theta^*} \in (0,1)$ there exists a sufficiently large a such that this response exhibits delayed overshooting. For instance, $e_2(\varepsilon,0) < e_1(\varepsilon,0) < e_0$ provided $a \geq \frac{3}{4}$ and $k^{\theta^*} \in \left(\frac{2a-1-\sqrt{4a-3}}{2a}, \frac{2a-1+\sqrt{4a-3}}{2a}\right)$. Since this interval converges to $(0,1)$ as $a \to 1$, it follows that for any $k^{\theta^*} \in (0,1)$ there exists a sufficiently large a such that $e_2(\varepsilon,0) < e_1(\varepsilon,0)$.

[23] A simple computation shows that $k^{\theta'} = \frac{a^2\xi^* + \sigma_w^2}{a^2\xi^* + \sigma_w^2 + \sigma_v^2} > \frac{1}{1+\frac{\sigma_v^2}{\sigma_w^2}} = q^{\theta'}$.

Proposition 5.3 (Forward Premium Puzzle). *Under observational uncertainty, the Fama regression coefficient converges in plim to*

$$\beta^{Fama} = 1 - \frac{\left(k^{\theta'} - k^{\theta^*}\right) a \left(a\left(1+a\right) k^{\theta^*} + \left(1-a^2\right)\right) \left(\frac{1}{\left(1-a^2\left(1-k^{\theta'}\right)\right)\left(1-a^2\left(1-k^{\theta^*}\right)\right)} + \frac{\sigma_v^2}{\sigma_w^2}\right)}{\left(1-a^2\right)\frac{\sigma_v^2}{\sigma_w^2} + 1} \tag{5.7}$$

1. Weak FPP. β^{Fama} *is less than one if and only if agents are both uncertainty-averse and risk-averse, so that* $k^{\theta^*} < k^{\theta'}$.

2. Strong FPP. β^{Fama} *is negative if, in addition, the unobservable component of the interest rate differential is highly persistent (i.e.,* a *is large) and uncertainty aversion* $(1/\lambda)$ *is large, but smaller than* $1/\lambda^v$ *(so that* k^{θ^*} *is strictly positive).*

3. β^{Fama} *cannot be negative if the drift* a *is small.*

Part 1 says that if agents overstate the relative importance of transitory shocks, so that $k^{\theta^*} < k^{\theta'}$, the asymptotic value of the Fama coefficient is strictly smaller than one and the forward premium is a biased predictor of the future rate of depreciation. Part 2 states that the strong form of the FPP (negative β^{Fama}) results if both persistence a and undersensitivity to news $k^{\theta^{'}} - k^{\theta^*}$ are large. Part 3 says that β^{Fama} can be negative only if the trend of the interest rate differential is highly persistent.

Before we describe the intuition, we note that the FPP implies predictable excess returns. To see this, notice that in our model "predictable excess returns under the data generating model θ'" equal the forward premium minus the expected–under the baseline model–exchange rate depreciation

$$\Lambda_t^{\theta'} := y_t - E_t^{\theta'}(e_{t+1} - e_t) \tag{5.8}$$

where $y_t \equiv i_t - i_t^f$. Using (4.5) to substitute for y_t, we show in the appendix that in the robust equilibrium

$$\Lambda_t^{\theta'} = \left[E_t^{\theta'}\left(x_{t+1}\right) - E_t^{\theta^*}\left(x_{t+1}\right)\right]\left[1 + \frac{ak^{\theta^*}}{1-a}\right] + \zeta_t \tag{5.9}$$

That is, under the data generating model θ', predictable excess returns equal the expectational distortion plus the uncertainty premium.[24] Let us now analyze the regression coefficient $\beta^{Fama} = p\lim_{t\to\infty} \frac{cov^{\theta'}(\Delta e_{t+1}, y_t)}{var^{\theta'}(y_t)}$. Using (5.8) and (5.9) we obtain $E_t^{\theta'}\Delta e_{t+1} = y_t - \Lambda_t^{\theta'}$, which can be expressed as

$$E_t^{\theta'}\Delta e_{t+1} = y_t - \left(1+\frac{ak^{\theta^*}}{1-a}\right)\left(k^{\theta'}-k^{\theta^*}\right)y_t + a\left(1+\frac{ak^{\theta^*}}{1-a}\right)\left(k^{\theta'}-k^{\theta^*}\right)\left(E_{t-1}^{\theta^*}(x_t) - E_{t-1}^{\theta'}(x_t)\right) + \zeta_t \tag{5.10}$$

The first term in (5.10) is the direct effect of the forward premium on average depreciation. The second term is the catch-up effect on average depreciation, and it is the source of the FPP. The third term captures the effect of past forecast errors. The last term is the uncertainty premium, which is constant in the equilibrium we are considering.

Consider an increase in y_t and ignore the third term in (5.10) for a moment. Equation (5.10) shows that if forecasts are less sensitive to news than baseline forecasts (i.e., $k^{\theta^*} < k^{\theta'}$), expected depreciation responds by less than the change in the forward premium. The result is a Fama coefficient being less than one, as stated in part 1 of Proposition 5.3. To derive part 2 notice that if both the persistence parameter a and the undersensitivity to news $k^{\theta'} - k^{\theta^*}$ are large, the catch-up effect in (5.10) captured by $\left(1+\frac{ak^{\theta^*}}{1-a}\right)\left(k^{\theta'}-k^{\theta^*}\right)$ can dominate the direct effect of one. As a result, the initial mispricing is so large that it requires the currency to appreciate further in the future. When a is large, this future upward revision in beliefs will have a large effect on the exchange rate because agents will expect high domestic interest rates to persist far into the future. This mispricing results in an extreme scenario where a high domestic interest rate coexists with an appreciating currency. Hence, under the data generating model θ', the forward premium and the expected depreciation tend to move in opposite directions on average, generating a negative regression coefficient β^{Fama}. Consider now the third term in (5.10). Since the gain differential $(k^{\theta'} - k^{\theta^*})$ is a positive constant, the sign of the third term equals the sign of $a\left(E_{t-1}^{\theta^*}(x_t) - E_{t-1}^{\theta'}(x_t)\right)$.[25] Furthermore, since $E_{t-1}^{\theta^*}(x_t)$ reacts less to y_{t-1} than $E_{t-1}^{\theta'}(x_t)$, and y_{t-1} is positively correlated with y_t, it follows that on average $E_{t-1}^{\theta^*}(x_t)$ is smaller(larger) than $E_{t-1}^{\theta'}(x_t)$ when y_{t-1} is positive(negative).

[24] In contrast, condition (4.5) implies that predictable excess returns under the agents' robust model θ^* (i.e., $(i_t - i_t^f) - E_t^{\theta^*}(e_{t+1} - e_t)$) equal simply the uncertainty premium ζ_t.

[25] Notice that $aE_{t-1}^{\theta^*}(x_t)$ and $aE_{t-1}^{\theta'}(x_t)$ are the estimates of y_t at time $t-1$.

Thus, $E_{t-1}^{\theta^*}(x_t) - E_{t-1}^{\theta'}(x_t)$ and y_t are negatively correlated. Therefore, the error in past forward premium estimates makes β^{Fama} smaller.

To see the intuition for part 3, notice that the coefficient β^{Fama} is decreasing in a because a more persistent trend implies that any forward premium expectational error will lead to a more severe mispricing of the exchange rate. When a is small, the downward bias in the robust gain ($k^{\theta^*} < k^{\theta'}$) does not translate into a significant undersensitivity of forecasts to interest rate news because persistent trend shocks become very similar to transitory observation shocks.

Proposition 5.3 implies that the strong FPP does not arise in an environment where agents are extremely uncertainty averse, so that they only guard against the worst-case scenario. In this extreme case, robust agents set the variance distortion to infinity ($\tilde{\sigma}_v^{2^*} = \infty$), which leads them to set the robust gain to zero ($k^{\theta^*} = 0$). A negative β^{Fama} requires a low, but strictly positive k^{θ^*}.[26]

Figure 3 depicts the Fama coefficient for the case of a persistent interest rate differential. As we can see, β^{Fama} is decreasing in the degree of aversion $(1/\lambda)$, and it becomes negative for high $1/\lambda$.

Finally, notice that in our robust equilibrium all deviations of β^{Fama} from one are generated by the forecast-distortion mechanism because the uncertainty premium ζ_t is deterministic.[27]

Simulations

Four parameters in our model determine the sign of the Fama coefficient: the persistence parameter a and the noise-to signal ratio σ_v^2/σ_w^2 that characterize the data

[26] When $k^{\theta^*} = 0$, $\bar{\beta}^F$ is positive: $\bar{\beta}^F = \frac{(1-a^2)\frac{\sigma_v^2}{\sigma_w^2}\left(1-ak^{\theta'}\right)+(1-a)\left(1+a\left(1-k^{\theta'}\right)\right)\left(1-a^2\left(1-k^{\theta'}\right)\right)^{-1}}{(1-a^2)\frac{\sigma_v^2}{\sigma_w^2}+1} > 0.$

[27] To see this decompose the realized log exchange rate change ($\Delta e_{t+1} := e_{t+1} - e_t$) into its robust forecast $\mathcal{E}_t^{\theta^*}(\Delta e_{t+1})$ and a forecast error v_{t+1} :

$$\Delta e_{t+1} = E_t^{\theta^*}(\Delta e_{t+1}) + v_{t+1}, \qquad E_t^{\theta^*}(v_{t+1}) = 0;$$

where we use the fact that the forecast error is zero under the robust measure θ^*. Using the above equation and (4.5) we have that

$$\beta^{Fama} = \frac{cov^{\theta'}(\Delta e_{t+1}, y_t)}{var^{\theta'}(y_t)} = 1 + \frac{cov^{\theta'}(v_{t+1}, y_t)}{var^{\theta'}(y_t)} - \frac{cov^{\theta'}(\zeta_t, y_t)}{var^{\theta'}(y_t)}$$

Since ζ_t is deterministic, $cov^{\theta'}(\zeta_t, y_t) = 0$.

generating process (3.1), and the degrees of absolute risk-aversion (γ) and uncertainty-aversion ($1/\lambda$) that characterize the utility function (3.4). We use empirical estimates found in the literature for (γ, σ_v^2, σ_w^2, a) and pin down the values of λ that generate a negative slope coefficient in the Fama regression.

Gourinchas and Tornell (2004) consider an interest rate differential process similar to the one in this paper. Using data on interest rate differentials and their forecasts for the G7 countries, they find that the drift of the differential a is high both in the actual data and the forecasts, but the noise-to signal ratio σ_v^2/σ_w^2 of the actual data is much lower than that implicit in the forecasts. Their estimates of a range over the interval $(0.95, 1)$ and their estimates of σ_v^2/σ_w^2 are near zero. Using the aggregate value of various risky and risk-free assets in the US., Bodie et. al. (2008) find that a coefficient of absolute risk-aversion γ of 2.6. According to Friend and Blume (1975) and Grossman and Shiller (1981) γ takes values on the interval $[2, 4]$. In our baseline simulations we set $\gamma = 2.6$, $a = 0.99$, $\sigma_v^2 = 1$ and $\sigma_w^2 = 100$ (so that, as in the data, the noise-to-signal ratio of the data generating process is close to zero: $\sigma_v^2/\sigma_w^2 = 0.01$).

For each set of parameter values (γ, σ_v^2, σ_w^2, a, λ) we obtain the values of the distorted variance $\tilde{\sigma}_v^{2*}$ and the robust gain k^{θ^*} that solve the system of equations (8.6)-(8.7) in the appendix.[28] We then compute the Fama coefficient using formula (5.7). Figure 4 sets (γ, σ_v^2, σ_w^2) equal to their baseline values and pins down the region (a, λ) that generates a negative Fama coefficient. As we can see, consistent with Proposition 5.3, a negative β^{Fama} results only if both a and $1/\lambda$ are large. Figure 5 shows the sensitivity to the degree of risk-aversion by letting γ range over the interval $[2, 4]$. As we can see, the lower γ, the smaller the set of (a, λ) pairs that generate negative β^{Fama}. This pattern is consistent with the result in Proposition 5.3 that β^{Fama} cannot be negative if $\gamma = 0$, regardless of the degree of uncertainty aversion.

As we noted earlier, a negative Fama coefficient is associated with a hump-shaped response to a one-time interest rate differential shock of unknown duration (i.e., delayed overshooting). Figure 6 exhibits the pairs (a, λ) that generate delayed overshooting paths of various lengths (five through twenty periods). In these simulations we set (γ, σ_v^2, σ_w^2) equal to their baseline values and, for each pair (a, λ), compute the number of periods the exchange rate continues appreciating following a shock to the forward

[28] Notice that for each set (γ, σ_v^2, σ_w^2, a, λ) there corresponds a unique pair ($\tilde{\sigma}_v^{2\theta*}$, k^{θ^*}).

premium of unknown duration (this number is given by formula (5.6)). As we can see, the (a, λ) regions over which there is delayed overshooting have the same shape as the region that generates a negative Fama coefficient in Figure 4.

6. Other Types of Uncertainty

We have seen that structured uncertainty about the link between the observations of the interest rate differential and its unobservable persistent component –i.e., the trend– can generate both the FPP and delayed overshooting when the trend is highly persistent. In this section we consider two other types of uncertainty: structured uncertainty about the trend process x_t, and unstructured uncertainty under which agents fear misspecification of the entire interest rate differential process but cannot pinpoint either its nature or location. In the first case we find that forecasts are more sensitive to news than baseline forecasts, which leads to $\beta^{Fama} > 1$. In the second case, the result is surprising: robust forecasts have the same sensitivity to news than the Bayesian forecasts under the baseline model. Thus, the forecast-distortion mechanism underlying the anomalies is not operative.

6.1. Structured Uncertainty in the Trend Equation

We consider two types of structured uncertainty in the trend equation (3.1): about the shock process w_t and about the parameter a. In each case, we follow the same steps as in Section 4. First, we define the uncertainty set, and present the change of measure lemma that links the set to a parameter distortion. Then we solve the investor's problem, and derive the equilibrium. The following uncertainty set captures misspecification in the shock process of the trend equation (3.1)

$$\Theta^w = \left\{ \theta \in P(\Omega) : \frac{d\theta}{d\theta'} = \exp\left(-\frac{1}{2} \left(\frac{1}{\tilde{\sigma}_w^2} - \frac{1}{\sigma_w^2} \right) (x_t - a x_{t-1})^2 \right) \cdot \sqrt{\frac{\sigma_w^2}{\tilde{\sigma}_w^2}}, \tilde{\sigma}_w^2 \in [\omega, \infty],\ \omega > 0 \right\} \tag{6.1}$$

The condition $\omega > 0$ ensures that set Θ^w is closed and convex. The next Lemma shows that under any measure in the set Θ^w, the interest rate differential process is given by baseline model (3.1), except that the variance of trend shocks has a distorted value $\tilde{\sigma}_w^2$ instead of the baseline σ_w^2.

Lemma 6.1 (Change of Measure II). *If under the baseline probability measure θ' the random variables in (3.1) are distributed as $x_t|I_{t-1} \overset{\theta'}{\sim} N\left(a\hat{x}_{t-1}, \sigma_w^2\right)$, $y_t|x_t \overset{\theta'}{\sim} N\left(x_t, \sigma_v^2\right)$ and $x_{t-1} \overset{\theta'}{\sim} N\left(\hat{x}_{t-1}, \sigma_{t-1}^2\right)$, then:*

1. *Under any probability measure $\theta \in \Theta^w$, $x_t|I_{t-1} \overset{\theta}{\sim} N\left(a\hat{x}_{t-1}, \tilde{\sigma}_w^2\right)$, while the distributions of $y_t|x_t$ and x_{t-1} are preserved.*

2. *The relative entropy of measure θ with respect to baseline measure θ' is*

$$\Re\left(\theta||\theta'\right) = \frac{1}{2}\left(\frac{\tilde{\sigma}_w^2}{\sigma_w^2} - \log\frac{\tilde{\sigma}_w^2}{\sigma_w^2} - 1\right) \qquad \text{for any } \theta \in \Theta^w \tag{6.2}$$

As in Section 4, the representative investor solves Problem (3.8). The next proposition shows that the solution to this problem and the equilibrium are similar to the case of observational uncertainty.

Proposition 6.2 (Equilibrium Under Trend-Shock Uncertainty). *Under uncertainty set Θ^w there exists a robust linear equilibrium if and only if the degree of uncertainty aversion $1/\lambda$ is lower than a threshold $1/\lambda_t^w$ given by (8.18).*

1. *The log exchange rate is ($Var_t^{\theta_t^*}\left(J_{t+1}\right)$ below is given by (8.17))*

$$e_t^* = -\left(i_t - i_t^f\right) - \frac{a}{1-a}\hat{x}_t^{\theta_t^*} + \alpha_t^*, \quad \alpha_t^* = \alpha_{t+1}^* + \gamma b_t^s Var_t^{\theta_t^*}\left(J_{t+1}\right). \tag{6.3}$$

2. *The robust forecast of the interest rate differential is given by (4.3) with a gain*

$$k_t^{\theta_t^*} \equiv \frac{a^2\sigma_{t-1}^2 + \tilde{\sigma}_{w,t}^{2*}}{a^2\sigma_{t-1}^2 + \sigma_v^2 + \tilde{\sigma}_{w,t}^{2*}}$$

3. *The trend-shock variance is necessarily distorted upwards ($\tilde{\sigma}_{w,t}^{2*} > \sigma_w^2$) for any $\gamma > 0$ and $\lambda > \lambda_t^w$. Thus, the robust forecasts of the forward premium are more sensitive to news than baseline forecasts: $k_t^{\theta_t^*} > k_t^{\theta'} = \frac{a^2\sigma_{t-1}^2+\sigma_v^2}{a^2\sigma_{t-1}^2+\sigma_v^2+\sigma_w^2}$.*

This proposition shows that robustness against misspecification in the trend-shock process generates a robust gain $k_t^{\theta_t^*}$ which is greater than the baseline gain. This is because robustness leads agents to introduce an upward distortion in the variance of the trend-equation shock. From the perspective of the anomalies the bad news is that since forecasts are more sensitive to news than baseline forecasts, the equilibrium exchange rate generates neither the FPP nor delayed overshooting. To see this note that the equilibrium exchange rate (6.3) has the same form as that under observational uncertainty in (4.6). It follows from the formula of the Fama coefficient (5.7) that we cannot account for the FPP if $\theta \in \Theta^w$. In fact, since the robust gain is strictly greater than the baseline gain, the regression coefficient in the Fama regression β^{Fama} would be greater than one. Similarly, the average impulse response function (5.5) implies that if $\theta \in \Theta^w$, there cannot be delayed overshooting. Following a positive forward premium shock the exchange rate appreciates initially and immediately reverts to a depreciating path. To see this note that (5.5) is strictly positive because $q^{\theta'}$ is necessarily lower than k_t^{θ} for any $\theta \in \Theta^w$. That is, there is no delayed overshooting because the agents' gain takes trend shocks to be more abundant that what they actually are in the data.

The upward distortion in the variance of the trend shock follows from applying the robustness principle to uncertainty set Θ^w. If agents fear misspecification in the equation of the unobservable trend of the forward premium, the costliest misspecification occurs when the agent's model wrongly sets the variance of the trend shock lower than what it actually is. Therefore, under trend equation uncertainty, robustness entails distorting upwards the variance of trend shocks, which in turn implies more sensitivity to news. To see the intuition consider a fictitious game between the agent and nature. The agents tries to minimize the mean squared error (MSE) of her estimate, while nature acts malevolently and tries to maximize it. Since the estimator takes the observer form $\hat{x} = (1-k)\, a\hat{c} + ky$, the MSE is $E^{\theta \in \Theta^w}(x - \hat{x})^2 = a^2(1-k)^2\sigma_c^2 + (1-k)^2\tilde{\sigma}_w^2 + k^2$. We can see that given the estimator (i.e., the gain k), nature can increase the MSE by choosing a larger $\tilde{\sigma}_w^2$. To counteract this potential misspecification the robust estimator sets the gain k greater than what Bayes law indicates. The degree of oversensitivity is determined by the degree of aversion to model uncertainty $1/\lambda$. The higher $1/\lambda$, the greater $\tilde{\sigma}_w^{2*}$ and thus the greater $k_t^{\theta_t^*}$. This is because the higher $1/\lambda$, the lower the penalization nature suffers when it sets $\tilde{\sigma}_w^2$ farther away from its baseline level. This penalization is given by $\lambda \cdot \Re\left(\theta || \theta'\right) = \frac{\lambda}{2}\left(\frac{\tilde{\sigma}_w^2}{\sigma_w^2} - \log\frac{\tilde{\sigma}_w^2}{\sigma_w^2} - 1\right)$.

6.2. Drift uncertainty

Here, we consider uncertainty about the drift of the unobservable trend of the forward premium. In the agent's baseline model the drift is a, but she fears that the true drift is $a+\delta$. We allow the misspecification δ to take values on $[-a, 1-a)$ so that the true drift is mean reverting and non-negative: $a+\delta \in [0,1)$. The following set of probability measures captures such drift uncertainty, while keeping the rest of the interest rate differential process unchanged.

$$\Theta^a = \left\{\theta \in \Omega : \ \frac{d\theta}{d\theta'} = \exp\left[-\frac{\left(x_{t-1}\delta\right)^2 - 2x_{t-1}\delta\left(x_t - ax_{t-1}\right)}{2\sigma_w^2}\right], \ \delta \in [-a, 1-a), \ a \geq 0\right\} \tag{6.4}$$

The next Lemma, which is a version of Girsanov's Theorem, establishes a one-to-one relationship between the set Θ^a and distortions to the drift.

Lemma 6.3 (Change of Measure III). *If under the baseline probability measure θ' the random variables in (3.1) are distributed as $x_t|I_{t-1} \overset{\theta'}{\sim} N\left(a\hat{x}_{t-1}, \sigma_{t-1}^2\right)$, $y_t|x_t \overset{\theta'}{\sim} N\left(x_t, \sigma_v^2\right)$ and $x_{t-1} \overset{\theta'}{\sim} N\left(\hat{x}_{t-1}, \sigma_w^2\right)$, then:*

1. *Under any probability measure $\theta \in \Theta^a$, $x_t|I_{t-1} \overset{\theta}{\sim} N\left(\left(a+\delta\right)\hat{x}_{t-1}, \sigma_{t-1}^2\right)$, while the distributions of $y_t|x_t$ and x_{t-1} are preserved.*

2. *The relative entropy of measure θ with respect to baseline measure θ' is*

$$\Re\left(\theta||\theta'\right) = \frac{1}{2}\frac{\delta^2}{\sigma_w^2}E_t^{\theta_t}(x_t^2) \qquad \text{for any } \theta \in \Theta^a \tag{6.5}$$

At time t, the representative young agent solves problem (3.8). Her conjecture of next period's exchange rate is (3.7), and her prior is that x_{t-1} is normally distributed with mean $\hat{x}_{t-1}^{\theta_{t-1}}$ and variance $\sigma_{t-1}^2 = \frac{\left(a^2\sigma_{t-2}^2+\sigma_w^2\right)\sigma_v^2}{a^2\sigma_{t-2}^2+\sigma_w^2+\sigma_v^2}$. The next proposition characterizes the equilibrium.

Proposition 6.4 (Equilibrium Under Drift Uncertainty). *Under uncertainty set Θ^a there exists a robust linear equilibrium if and only if the degree of uncertainty aversion $1/\lambda$ is lower than a threshold $1/\lambda_t^a$ given by (8.21). In this equilibrium:*

1. *The log exchange rate is ($Var_t^{\theta_t^*}(J_{t+1})$ below is given by (8.19)).*

$$e_t^* = -\left(i_t - i_t^f\right) - \frac{a+\delta_t^*}{1-(a+\delta_t^*)}\hat{x}_t^{\theta_t^*} + \alpha_t^*, \quad \alpha_t^* = \alpha_{t+1}^* + \gamma b_t^s Var_t^{\theta_t^*}(J_{t+1}). \quad (6.6)$$

2. *The robust estimate of the persistent component of the interest rate differential is*

$$\begin{aligned} \hat{x}_t^{\theta_t} &= k_t^{\theta_t^*}\left(i_t - i_t^f\right) + \left(1 - k_t^{\theta_t^*}\right)(a+\delta_t^*)\hat{x}_{t-1}^{\theta_{t-1}^*}, \quad (6.7) \\ \text{with} \quad k_t^{\theta_t^*} &= \frac{(a+\delta_t^*)^2\sigma_{t-1}^2 + \sigma_w^2}{(a+\delta_t^*)^2\sigma_{t-1}^2 + \sigma_w^2 + \sigma_v^2} \text{ and } \sigma_t^2 = \frac{\left((a+\delta_t^*)^2\sigma_{t-1}^2 + \sigma_w^2\right)\sigma_v^2}{(a+\delta_t^*)^2\sigma_{t-1}^2 + \sigma_w^2 + \sigma_v^2} \end{aligned}$$

3. *The drift distortion is necessarily positive: $\delta_t^* > 0$. Thus, the robust forecasts of the forward premium are more sensitive to news than baseline forecasts: $k_t^{\theta_t^*} > k_t^{\theta'} = \frac{a^2\sigma_{t-1}^2+\sigma_v^2}{a^2\sigma_{t-1}^2+\sigma_v^2+\sigma_w^2}$.*

This proposition says that in the presence of drift uncertainty, the forecasts are more sensitive to news than baseline forecasts. This is because the gain is increasing in the drift distortion δ, which is positive in equilibrium. Equation (6.6) shows that this oversensitivity ($k_t^{\theta_t} > k_t^{\theta'}$) is carried over to the exchange rate. As a result the forecast-distortion mechanism generates neither the FPP nor delayed overshooting.

To see the intuition for why the drift distortion δ must be positive, consider a static filtering problem in which the agent is confident about her model linking noisy observations and trend, but is uncertain about the *drift* of the unobservable trend component. That is, let $y = x+v$ and $x = (a+\delta)\,c+w$, where $c \sim N\left((a+\delta)\,\hat{c}, \sigma_c^2\right)$ is the prior distribution of x, and the disturbances follow distributions $w \sim N\left(0, \sigma_w^2\right)$ and $v \sim N\left(0, \sigma_v^2\right)$. Then, the estimator of x given y is $\hat{x} = (a+\delta)(1-k)\,\hat{c} + ky$. Therefore, under drift uncertainty the MSE is $E^{\theta\in\Theta^a}(x-\hat{x})^2 = (a+\delta)^2(1-k)^2\sigma_c^2 + (1-k)^2\sigma_w^2 + k^2\sigma_v^2$.[29] Since the true MSE is increasing in the true drift, and the true drift is restricted to be positive and mean reverting (i.e., $a+\delta \in [0,1)$), the costliest misspecification occurs when the agent's model wrongly sets the drift lower than what it actually is. Therefore, if her objective is to bound the effects of misspecification on the mean squared error of her estimator, she should distort upwards the drift. This distortion implies more sensitivity

[29] To obtain the MSE notice that $\hat{x} - x = (a+\delta)(1-k)\,\hat{c} + (k-1)(a+\delta)\,c + (k-1)\,w + kv$.

to news. In other words, when she observes a large realization of the interest rate differential she should fear that it is more likely to come from a change in the trend than what her baseline model implies. Thus, her best response is to put a higher weight on the observations than the baseline weight. Hence, she is more sensitive to news. The degree of oversensitivity depends on the value of the uncertainty aversion coefficient $1/\lambda$.

6.3. Unstructured Uncertainty

Under unstructured uncertainty the investor does not know the nature of the misspecification, and does not know whether it is located in the observation equation, or in the trend equation or in both. Here, we define the unstructured uncertainty set as the set of all probability measures on the measurable space $(\Omega, \mathcal{B}(\Omega))$, where $\mathcal{B}(\Omega)$ is the Borel $\sigma-$algebra.[30]

$$\Theta^u = P(\Omega) \equiv \{\theta : \mathcal{B}(\Omega) \to [0,1] \text{ and } \theta(\Omega) = 1\} \tag{6.8}$$

This set allows for a truly worst-case scenario! Optimizing over the set Θ^u seems a daunting task. Fortunately a result of the theory of large deviations–the Representation Lemma of Dupuis and Ellis (1997)–implies that the problem of the investor simplifies significantly.

Lemma 6.5 (Representation Lemma). *Under unstructured uncertainty, the robust forecasting-portfolio problem (3.4) reduces to the following Bayesian problem under the unique baseline probability measure θ'.*

$$\Gamma_t = \max_{b_t} \inf_{\theta \in P(\Omega)} \left\{E_t^\theta \left[u(W_{t+1}) + \lambda \cdot \Re(\theta||\theta')\right]\right\} = \max_{b_t} \left\{-\lambda \log \left(E_t^{\theta'} \exp\left(-\frac{1}{\lambda} u(W_{t+1})\right)\right)\right\} \tag{6.9}$$

This result is striking. It says that the robust agent's problem reduces to an expected utility maximization problem under a unique probability measure. Moreover, this unique probability measure is the baseline measure θ'. This is a Bayesian problem, similar to the ones considered in rational expectations models. Under the equivalent representation the relative entropy disappears and $\inf_{\theta \in P(\Omega)} E_t^\theta u(W_{t+1})$ is replaced by

[30] The set of all probability measures on Ω is compact by Alaoglu's theorem (Folland (2001), pp 169).

the so called risk-sensitive utility function $-\lambda \log\left(E_t^{\theta'} \exp\left(-\frac{1}{\lambda} u\left(W_{t+1}\right)\right)\right)$. This function keeps the baseline probability measure unchanged and, because of the exponential function, captures the desire for robustness by putting more weight on the tails of the distribution. Hence a risk-sensitive agent is more concerned about tail events than a typical risk-averse agent.

From the perspective of accounting for the exchange rate anomalies the bad news is that in this problem the "separation principle" applies: Expectations can be computed independently of the portfolio strategy. The investor forms her expectations using Bayes law under the baseline probability measure θ' and based on these expectations she then chooses her portfolio. This separation implies that the gain that will appear in the equilibrium exchange rate function is the Bayesian gain $k^{\theta'}$. Thus, we will not get the lower sensitivity of forecasts to news ($k^{\theta^*} < k^{\theta'}$) that underpins the explanation for delayed overshooting and for the forward premium puzzle in Propositions 5.2 and 5.3.

In order to derive the equilibrium we replace $u(W_{t+1}) = -\exp(-\gamma W_{t+1})$ in (6.9) and reexpress problem Γ_t as follows

$$\Gamma_t = \max_{b_t}\left\{-\lambda \log\left(E_t^{\theta'} \exp\left(-\tfrac{1}{\lambda}\exp(-\gamma W_{t+1})\right)\right)\right\} \Leftrightarrow \min_{b_t}\left\{E_t^{\theta'} \exp\left(-\tfrac{1}{\lambda}\exp(-\gamma W_{t+1})\right)\right\}.$$

We show in the appendix that taking the first order condition for b_t, using Stein's Lemma and the exchange rate conjecture (3.7), it follows that there is an interior solution for b_t only if returns satisfy the following condition:

$$E_t^{\theta'}\left(e_{t+1}\right) - e_t = \left(i_t - i_t^f\right) + l_t. \tag{6.10}$$

Condition (6.10) is the well known uncovered interest parity condition plus a time-varying uncertainty premium l_t, which is given by

$$\begin{aligned} l_t &\equiv \frac{Var_t^{\theta'}\left(e_{t+1}\right) E_t^{\theta'} g'\left(J_{t+1}\right)}{E_t^{\theta'} g\left(J_{t+1}\right)}, \quad \text{where } J_{t+1} \equiv \left(i_t - i_t^f\right) - \left(e_{t+1} - e_t\right) \\ g(J_{t+1}) &\equiv -\exp\left(\gamma b_t^s J_{t+1}\right) \exp\left(-\tfrac{1}{\lambda}\exp\left(\gamma b_t^s J_{t+1}\right)\right), \; g'\left(J_{t+1}\right) = \left[1 - \tfrac{1}{\lambda}\exp\left(\gamma b_t J_{t+1}\right)\right] \gamma b_t^s g\left(J_{t+1}\right) \end{aligned} \tag{6.11}$$

Notice that when the risk-aversion coefficient γ is large, $g'\left(J_{t+1}\right)$ is negative and thus the uncertainty premium is large and positive. In contrast, in the risk-neutral case ($\gamma = 0$), the uncertainty premium becomes zero (because $g'\left(J_{t+1}\right) = 0$). This result implies that risk-neutrality combined with a desire for robustness against unstructured uncertainty yields the same equilibrium as risk neutrality under no model uncertainty. Finally, when there is no aversion to model uncertainty (i.e., $1/\lambda$ goes to zero), the uncertainty premium l_t equals $\gamma b_t^s Var_t^{\theta'}(e_{t+1})$, which is the same as in a standard rational expectations equilibrium.[31]

From the uncovered interest parity condition (6.10) we can find the equilibrium.

Proposition 6.6 (Equilibrium). *Under unstructured uncertainty, in a robust linear equilibrium the log exchange rate is*

$$e_t^* = -\left(i_t - i_t^f\right) - \frac{a}{1-a}\hat{x}_t^{\theta'} + \alpha_t^*, \quad \alpha_{t+1}^* = \alpha_t^* + l_t^*,$$

where the estimate of the unobservable trend is given by the standard Kalman filter under the unique baseline probability measure θ'

$$\hat{x}_t^{\theta'} = \left(1 - k_t^{\theta'}\right) a\hat{x}_{t-1}^{\theta'} + k_t^{\theta'}\left(i_t - i_t^f\right), \; k_t^{\theta'} = \frac{a^2\sigma_{t-1}^2 + \sigma_w^2}{a^2\sigma_{t-1}^2 + \sigma_w^2 + \sigma_v^2}, \; \sigma_{t-1}^2 = \frac{\left(a^2\sigma_{t-2}^2 + \sigma_w^2\right)\sigma_v^2}{a^2\sigma_{t-2}^2 + \sigma_w^2 + \sigma_v^2},$$

and l_t^* is given by (6.11) evaluated at $\beta_1 = -\frac{a}{1-a}$, $\beta_2 = -1$ and $b_t = b_t^s$.

Can the forecast-distortion mechanism generate the anomalies? The answer is no. Since e_t^* has the same linear form as (4.6), we know from (5.5) and (5.7) in Propositions 5.2 and 5.3 that the first two terms in e_t^* : (i)do not generate momentum, and (ii)give rise to $\beta^{Fama} = 1$ because the robust gain equals the baseline gain under no misspecification $k_t^{\theta'}$. This implies that the forecast-distortion mechanism is not operational under unstructured uncertainty.

7. Literature Review

Following Fama (1984) the FPP has been documented for many currency pairs and time periods (Chinn (2006), Engel (1996) and Lewis (1995) survey the literature). Several mechanisms have been proposed to account for the FPP. One group of papers concentrates on the risk premium (Alvarez, et. al. (2006), Bekaert (2006), Frankel and Engel

[31] This is because $\lim_{\lambda\to\infty} \frac{E_t^{\theta'} g'(J_{t+1})}{E_t^{\theta'} g(J_{t+1})} = \gamma b_t^2$

(1984), Lusting and Verdelhan (2006), Nelson and Wu (1998)). A second group concentrates on expectational biases (Frankel and Froot (1989), Lewis (1989), Gourinchas and Tornell (2004)). A third group considers a peso-problem (Kaminski (2003), Farhi and Gabaix (2007)). A forth group looks at the microstructure of the trading mechanism (Bacchetta and Van Wincoop (2006), Burnside et. al. (2007), Evans and Lyons (2002), and Sarno, et. al. (2006)). Finally, Backus et. al. (2001) link the FPP to affine models of the term structure. On the empirical front, Frankel and Froot (1989) find, using exchange rate survey data, that expectational biases can better account for the FPP than risk premia. Using interest rate survey data Gourinchas and Tornell (2004) find that among G7 currencies forecasts of interest rate differentials overstate the relative importance of transitory shocks and that the resulting hump-shaped forecast pattern can account for the FPP. Our model has characterized conditions under which such a forecast-distortion mechanism is the outcome of utility optimization. A delayed response also arises in Bacchetta and Van Wincoop (2008) because transactions costs lead agents to make infrequent portfolio decisions.

Burnside et. al. (2008) construct a hedged carry trade portfolio that exploits the FPP and hedges against large adverse exchange rate changes by buying out-of-the-money options. They find that this strategy is profitable, and that neither a large peso-problem nor standard macroeconomic risk factors typically associated with risk premia can account for the main part its returns. These findings are consistent with our model. Although, in our setup robust agents can guard against many classes of misspecification, like a large peso-problem, in the equilibrium that generates the FPP, agents do not guard against such an extreme event. They guard only against misspecification in the link between observations and the underlying return process, and robustness involves only distorting second moments.

As in Hansen and Sargent (2007), we use robust control in formulating and solving the agent's problem. In particular, we focus on the robust filtering problem, and investigate when is it that it delivers the lower gain that is required for the hump-shaped dynamics that underlie the FPP. Unfortunately, under unstructured uncertainty, the robust filtering problem yields an estimator with the same gain as that of the baseline Bayes estimator, and the $\mathtt{H}_\infty$ filtering problem yields a larger gain (Basar and Bernhard

(1995)).[32] These results lead us to consider instead different types of structured uncertainty, and determine conditions under which the robust filtering problem yields a gain lower than the baseline gain. Then we ask when is it that this property is carried over to the equilibrium exchange rate function in a setup where agents solve a joint portfolio-filtering problem.

This paper is linked to several papers that have used robust control to analyze macroeconomic issues. Brock, et. al. (2007), Cogley and Sargent (2005) and Svensson and Williams (2007) have addressed the issue of robust monetary policy design. Another group of papers investigates whether robustness against model misspecification can shed light on the fact that US monetary policy responses tend to be more cautious that those implied by optimal policies. (Dennis (2007), Giannoni (2002), Kasa (2002), Onatski and Stock (2002), Tetlow and von zur Muehlen (2006)). Like in this paper, whether robust policy entails cautiousness or aggressiveness depends on the structure of the uncertainty set. A third group studies asset pricing under aversion to model uncertainty (Bossaerts, et al. (2004), Caggetti, et. al. (2002), Epstein and Wang (1994), Hansen, et. al. (1999), Rigotti and Shannon (2005), and Tornell (2001)). Finally, several papers have analyzed the long-run horizon link between exchange rates and fundamentals as well as their expectations (e.g., Engel and West (2005) and Mark (1995)). Our forecast distortion mechanism works at higher frequencies –a few quarters– and so can be considered as complementary to these papers.

8. Conclusions

This paper brings together the robustness literature and the international finance literature in order to analyze the forward premium puzzle (FPP), a major anomaly in foreign exchange markets. It characterizes conditions under which robustness against model misspecification generates a negative Fama coefficient. Specifically, a desire for robustness leads optimizing agents that hold no misperception to distort the probability distribution of the data-generating process and to make forecasts of interest rate differentials that exhibit delayed overreaction to news, i.e., a humped-shaped pattern. Thus,

[32] The $\mathtt{H}_\infty$ filtering problem minimizes the mean-squared error attenuated by the cumulative energy of the disturbances.

when there is a positive shock to the interest rate differential, the currency does not appreciate enough and therefore needs to catch up subsequently before mean reverting. This delayed overreaction pattern generates a negative Fama coefficient because along the catch-up phase there is a gradual appreciation of the currency alongside a positive interest rate differential. Deriving the forecast distortion that underlies the FPP from an optimizing-robust framework is our contribution to the international finance literature.

Surprisingly, robustness against unstructured uncertainty does not generate the low gain that underlies the hump-shaped dynamics. In fact, under unstructured uncertainty, the H_∞-filter has a greater gain relative to the case where the desire for robustness is absent. We have found that in the standard robust filtering problem, the structure of the uncertainty set is a determining factor of the sensitivity to news of the filter. The search for a hump-shaped pattern has lead us to consider a robust filtering-portfolio problem under several types of structured uncertainty. We have solved it with the aid of Girsanov-like change of measure techniques that translate sets of probability measures into parameter distortions and have allowed us to derive in close-form the forward looking exchange rate and the Fama coefficient. To our knowledge, this constitutes a novel contribution to the robustness literature.

We have found that the forecast-distortion mechanism generates a negative Fama coefficient if there is structured uncertainty about the link between observations and the persistent component of the interest rate differential process, provided there is enough uncertainty aversion and the differential is highly persistent. In contrast, a negative Fama coefficient arises neither under unstructured nor under structured uncertainty about the persistent component of the differential, regardless of whether it is about a parameter or about the shock process. These results are consistent with the finding in the empirical literature that the FPP is less prevalent in high inflation environments than in low inflation ones (e.g., Bansal and Dahlquist (2000)). In the former we should expect that the major source of misspecification is the inflation process, and that nominal interest rate differentials reflect mainly inflation differentials. If the inflation process is persistent and uncertain, we should expect that uncertainty about the persistent component of the nominal interest rate differential is a major source of model uncertainty. Our model would then predict that robust agents will use a higher gain in their forecasts, thereby generating a larger Fama coefficient in high inflation environments than in low inflation ones.

Finally, we would like to mention a few possible extensions. The delayed overreaction that underlies our explanation for the FPP can be used to account for the event-based momentum anomaly found in the finance literature. Since our forecast distortion mechanism generates price changes in the same direction as the price change at the time of the event, and our equilibrium has the same linear form as the standard rational expectations equilibrium price functions, such a link seems relatively straightforward. Another extension would be to explain the existence of the carry trade, under which investors borrow in low yielding currencies and invest in high yielding currencies.

Appendix

Proof of Lemma 4.1. We prove part 1 in three steps. First, we show that if $y_t|x_t$ follows a normal distribution $N\left(x_t, \sigma_v^2\right)$ under the baseline model θ', then under any model $\theta \in \Theta^v$ we have that $y_t|x_t \overset{\theta}{\sim} N\left(x_t, \tilde{\sigma}_v^2\right)$. Second, we show that $x_t|x_{t-1}$ and x_{t-1} have the same distribution under θ as under θ'. We start by computing the probability distribution of $y_t|x_t$ under probability measure θ.

$$\begin{aligned} P^{\theta}\left(y_t < z|x_t\right) &= \int_{\{y_t<z\}} d\theta = \int_{\{y_t<z\}} \frac{d\theta}{d\theta'} d\theta' = \int_{\{y_t<z\}} \exp\left(-\frac{1}{2}\left(\frac{1}{\tilde{\sigma}_v^2} - \frac{1}{\sigma_v^2}\right)\left(y_t - x_t\right)^2\right) \cdot \sqrt{\frac{\sigma_v^2}{\tilde{\sigma}_v^2}} d\theta' \\ &= \int_{-\infty}^{z} \exp\left(-\frac{1}{2}\left(\frac{1}{\tilde{\sigma}_v^2} - \frac{1}{\sigma_v^2}\right)\left(y - x_t\right)^2\right) \cdot \sqrt{\frac{\sigma_v^2}{\tilde{\sigma}_v^2}} \frac{1}{\sqrt{2\pi}\sigma_v} \exp\left[-\frac{\left(y - x_t\right)^2}{2\sigma_v^2}\right] dy \\ &= \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}\tilde{\sigma}_v} \exp\left[-\frac{\left(y - x_t\right)^2}{2\tilde{\sigma}_v^2}\right] dy, \end{aligned}$$

The last equality shows that, conditional on x_t, y_t follows $N\left(x_t, \tilde{\sigma}_v^2\right)$ under measure θ. Next, we compute the distribution function of $x_t|x_{t-1}$ under θ

$$\begin{aligned} P^{\theta}\left(x_t < z|x_{t-1}\right) &= E^{\theta} 1_{\{x_t<z|x_{t-1}\}} = E^{\theta'} 1_{\{x_t<z|x_{t-1}\}} \frac{d\theta}{d\theta'} \\ &= E^{\theta'} 1_{\{x_t<z|x_{t-1}\}} \exp\left(-\frac{1}{2}\left(\frac{1}{\tilde{\sigma}_v^2} - \frac{1}{\sigma_v^2}\right) v_t^2\right) \cdot \sqrt{\frac{\sigma_v^2}{\tilde{\sigma}_v^2}} \\ &= E^{\theta'} 1_{\{x_t<z|x_{t-1}\}} E^{\theta'} \exp\left(-\frac{1}{2}\left(\frac{1}{\tilde{\sigma}_v^2} - \frac{1}{\sigma_v^2}\right) v_t^2\right) \sqrt{\frac{\sigma_v^2}{\tilde{\sigma}_v^2}} \end{aligned}$$

The last equality follows because random variables x_t and v_t are independent under θ'.

Notice that the second expectation in the last equality is equal to one because

$$E^{\theta'} \exp\left(-\frac{1}{2}\left(\frac{1}{\tilde{\sigma}_v^2}-\frac{1}{\sigma_v^2}\right)v_t^2\right)\sqrt{\frac{\sigma_v^2}{\tilde{\sigma}_v^2}} = \int \exp\left(-\frac{1}{2}\left(\frac{1}{\tilde{\sigma}_v^2}-\frac{1}{\sigma_v^2}\right)v_t^2\right)\sqrt{\frac{\sigma_v^2}{\tilde{\sigma}_v^2}}d\theta'$$

$$\begin{aligned} &= \int \exp\left(-\frac{1}{2}\left(\frac{1}{\tilde{\sigma}_v^2}-\frac{1}{\sigma_v^2}\right)v_t^2\right)\sqrt{\frac{\sigma_v^2}{\tilde{\sigma}_v^2}}\frac{1}{\sqrt{2\pi\sigma_v^2}}\exp\left(-\frac{1}{2}\frac{v_t^2}{\sigma_v^2}\right)dv_t \\ &= \int \frac{1}{\sqrt{2\pi\tilde{\sigma}_v^2}}\exp\left(-\frac{1}{2}\frac{v_t^2}{\tilde{\sigma}_v^2}\right)dv_t = 1 \end{aligned}$$

Therefore, we have that $P^{\theta}(x_t < z|x_{t-1}) = E^{\theta}1_{\{x_t<z|x_{t-1}\}} = E^{\theta'}1_{\{x_t<z|x_{t-1}\}} = P^{\theta'}(x_t < z|x_{t-1})$. This shows that $x_t|x_{t-1}$ has the same distribution under θ as under θ'. Lastly, we use a similar argument to show that x_{t-1} has the same distribution under θ as under θ'.

$$\begin{aligned} P^{\theta}(x_{t-1}<z) &= E^{\theta}1_{\{x_{t-1}<z\}} = E^{\theta'}1_{\{x_{t-1}<z\}}\frac{d\theta}{d\theta'} = E^{\theta'}1_{\{x_t<z\}}\exp\left(-\frac{1}{2}\left(\frac{1}{\tilde{\sigma}_v^2}-\frac{1}{\sigma_v^2}\right)v_t^2\right)\cdot\sqrt{\frac{\sigma_v^2}{\tilde{\sigma}_v^2}} \\ &= E^{\theta'}1_{\{x_t<z\}}E^{\theta'}\exp\left(-\frac{1}{2}\left(\frac{1}{\tilde{\sigma}_v^2}-\frac{1}{\sigma_v^2}\right)v_t^2\right)\sqrt{\frac{\sigma_v^2}{\tilde{\sigma}_v^2}} = E^{\theta'}1_{\{x_t<z\}} \end{aligned}$$

To prove part 2 we derive the relative entropy.

$$\begin{aligned} R(\theta||\theta') &= E^{\theta}\log\left(\frac{d\theta}{d\theta'}\right) = E^{\theta}\left(-\frac{1}{2}\left(\frac{1}{\tilde{\sigma}_w^2}-\frac{1}{\sigma_v^2}\right)(y_t-x_t)^2+\frac{1}{2}\left(\log\left(\sigma_v^2\right)-\log\left(\tilde{\sigma}_v^2\right)\right)\right) \\ &= -\frac{1}{2}\left(\frac{1}{\tilde{\sigma}_v^2}-\frac{1}{\sigma_v^2}\right)\cdot E^{\theta}\left(v_t^2\right)+\frac{1}{2}\left(\log\left(\sigma_v^2\right)-\log\left(\tilde{\sigma}_v^2\right)\right) = \frac{1}{2}\left(\frac{\tilde{\sigma}_v^2}{\sigma_v^2}-\log\left(\frac{\tilde{\sigma}_v^2}{\sigma_v^2}\right)-1\right) \end{aligned}$$

Proof of Lemma 4.2. Notice that for a given probability measure θ_t (or given the distorted variance $\tilde{\sigma}_{v,t}^2$), the investor constructs her estimates of x_t and x_{t+1} using Bayes law, given prior estimate $\hat{x}_{t-1}^{\theta_{t-1}}$ and prior variance σ_{t-1}^2. These estimates are

$$\hat{x}_t^{\theta_t} = E_t^{\theta_t}(x_t) = \left(1-k_t^{\theta_t}\right)a\hat{x}_{t-1}^{\theta_{t-1}}+k_t^{\theta_t}y_t, \quad k_t^{\theta_t} = \frac{a^2\sigma_{t-1}^2+\sigma_w^2}{a^2\sigma_{t-1}^2+\sigma_w^2+\tilde{\sigma}_{v,t}^2}$$

$$x_t|I_t \overset{\theta_t}{\sim} N\left(\hat{x}_t^{\theta_t}, \frac{\left(a^2\sigma_{t-1}^2+\sigma_w^2\right)\tilde{\sigma}_{v,t}^2}{a^2\sigma_{t-1}^2+\sigma_w^2+\tilde{\sigma}_{v,t}^2}\right) \quad x_{t+1}|I_t \overset{\theta_t}{\sim} N\left(a\hat{x}_t^{\theta_t}, a^2\frac{\left(a^2\sigma_{t-1}^2+\sigma_w^2\right)\tilde{\sigma}_{v,t}^2}{a^2\sigma_{t-1}^2+\sigma_w^2+\tilde{\sigma}_{v,t}^2}+\sigma_w^2\right)$$

Expectations are computed conditional on the information set I_t. Under probability measure θ_t the excess rate of return (denoted by J_{t+1}) is

$$
\begin{aligned}
J_{t+1} &\equiv (i_t - i_t^f) - (e_{t+1} - e_t) = -\alpha_{t+1} - \beta_1 \hat{x}_{t+1}^{\theta_{t+1}} - \beta_2 \left(i_{t+1} - i_{t+1}^f \right) + e_t + (i_t - i_t^f) \\
&= -\alpha_{t+1} - \left(1 - k_{t+1}^{\theta_{t+1}} \right) \beta_1 a \hat{x}_t^{\theta_t} - \left(k_{t+1}^{\theta_{t+1}} \beta_1 + \beta_2 \right) \left(i_{t+1} - i_{t+1}^f \right) + e_t + (i_t - i_t^f)
\end{aligned}
$$

To derive J_{t+1} note that the t-agent uses her conjecture (3.7) to forecast next period's exchange rate: $e_{t+1}^{conj} = \alpha_{t+1} + \beta_1 \hat{x}_{t+1}^{\theta_{t+1}} + \beta_2 \left(i_{t+1} - i_{t+1}^f \right)$. Furthermore, the t-agent knows that the agent at time $t+1$ will (i) use the same method to distort the probability measure θ_{t+1} as the one used by the $t-$agent, and (ii) will forecast x_{t+1} using Bayes law under this θ_{t+1}, with a given prior mean $\hat{x}_t^{\theta_t}$ and variance σ_t^2. The variance σ_t^2 follows the Kalman filter equation under the baseline measure θ': $\sigma_t^2 = \frac{\left(a^2 \sigma_{t-1}^2 + \sigma_w^2\right) \sigma_v^2}{a^2 \sigma_{t-1}^2 + \sigma_w^2 + \sigma_v^2}$, where $\hat{x}_{t+1}^{\theta_{t+1}}$ is the estimate of hidden state by agent $t+1$. It follows that in J_{t+1} we can replace $\hat{x}_{t+1}^{\theta_{t+1}}$ by $\hat{x}_{t+1}^{\theta_{t+1}} = \left(1 - k_{t+1}^{\theta_{t+1}} \right) a \hat{x}_t^{\theta_t} + k_{t+1}^{\theta_{t+1}} \left(i_{t+1} - i_{t+1}^f \right)$. Under probability measure $\theta_t \in \Theta^v$, J_{t+1} is normally distributed with mean and variance

$$
\begin{aligned}
E_t^{\theta_t}(J_{t+1}) &= (i_t - i_t^f) - \left(E_t^{\theta_t} e_{t+1} - e_t \right) = -\alpha_{t+1} - a(\beta_1 + \beta_2) \hat{x}_t^{\theta_t} + e_t + \left(i_t - i_t^f \right) \\
V_t^{\theta_t}(J_{t+1}) &= \left(k_{t+1}^{\theta_{t+1}} \beta_1 + \beta_2 \right)^2 \left(a^2 \frac{\left(a^2 \sigma_{t-1}^2 + \sigma_w^2 \right) \tilde{\sigma}_{v,t}^2}{a^2 \sigma_{t-1}^2 + \sigma_w^2 + \tilde{\sigma}_{v,t}^2} + \sigma_w^2 + \tilde{\sigma}_{v,t}^2 \right)
\end{aligned} \tag{8.1}
$$

Next, note that problem (4.2) is equivalent to (3.8) because for any normally distributed random variable z, $E^\theta(\exp[-\gamma z]) = \exp\left(-\gamma E^\theta(z) + \frac{\gamma^2}{2} var^\theta(z) \right)$. We solve problem (4.2) by considering the investor as a Stackelberg leader, that takes into account the strategy of nature: $\tilde{\sigma}_{v,t}^2 = s(b_t, e_t)$. Nature then selects $\tilde{\sigma}_{v,t}^2$ conditioning on the agent's choice of b_t. Notice that $\tilde{\sigma}_{v,t}^2$ affects the investor's payoff through its effect on $E_t^{\theta_t}(J_{t+1})$ and $Var_t^{\theta_t}(J_{t+1})$. The first order conditions (FOCs) are

$$
\frac{\partial \Gamma}{\partial \tilde{\sigma}_{v,t}^2} = -\frac{(\gamma b_t)^2}{2} \frac{\partial Var_t^{\theta_t}(J_{t+1})}{\partial \tilde{\sigma}_{v,t}^2} \frac{\exp\left(\frac{(\gamma b_t)^2}{2} Var_t^{\theta_t}(J_{t+1}) \right)}{\exp(\gamma b_t E_t^{\theta_t}(J_{t+1}))} + \frac{\lambda}{2} \left(\frac{1}{\sigma_v^2} - \frac{1}{\tilde{\sigma}_{v,t}^2} \right) = 0 \tag{8.2}
$$

$$
\begin{aligned}
\frac{\partial \Gamma}{\partial b_t} &= -\left(-\gamma E_t^{\theta_t}\left(J_{t+1}\right)+\gamma^2 b_t Var_t^{\theta_t}\left(J_{t+1}\right)\right)\frac{\exp\left(\frac{(\gamma b_t)^2}{2}Var_t^{\theta_t}\left(J_{t+1}\right)\right)}{\exp\left(\gamma b_t E_t^{\theta_t}\left(J_{t+1}\right)\right)}+\frac{\partial \Gamma}{\partial \tilde{\sigma}_{v,t}^2}\frac{d\tilde{\sigma}_{v,t}^2}{db_t} \\
&= \left[\gamma\left[-\left(\alpha_{t+1}+a\left(\beta_1+\beta_2\right)\hat{x}_t^{\theta_t}+e_t+\left(i_t-i_t^f\right)\right)\right]-\gamma^2 b_t Var_t^{\theta_t}\left(J_{t+1}\right)\right]\cdot \\
&\quad \exp\left(-\gamma b_t E_t^{\theta_t}\left(J_{t+1}\right)+\frac{\left(\gamma b_t\right)^2}{2}Var_t^{\theta_t}\left(J_{t+1}\right)\right)=0 \qquad (8.3)
\end{aligned}
$$

The second equality in (8.3) follows from the envelope theorem: $\frac{\partial \Gamma}{\partial \tilde{\sigma}_{v,t}^2}=0$. The second order condition (SOC) for the investor's problem is $0>\frac{\partial^2\Gamma}{\partial b_t^2}=\Gamma_{b_t b_t}(b_t,\tilde{\sigma}_{v,t}^2(b_t))+\Gamma_{b_t\sigma_{v,t}^2}(b_t,\tilde{\sigma}_{v,t}^2(b_t))\frac{d\tilde{\sigma}_{v,t}^2(b_t)}{db_t}+\Gamma_{\tilde{\sigma}_{v,t}^2\tilde{\sigma}_{v,t}^2}(b_t,\tilde{\sigma}_{v,t}^2(b_t))\cdot\left(\frac{d\tilde{\sigma}_{v,t}^2(b_t)}{db_t}\right)^2+\Gamma_{\sigma_{v,t}^2}(b_t,\tilde{\sigma}_{v,t}^2(b_t))\cdot\frac{d^2\tilde{\sigma}_{v,t}^2(b_t)}{db_t^2}$. Notice that the total derivative of nature's first order condition $\Gamma_{\tilde{\sigma}_{v,t}^2}(b,\tilde{\sigma}_{v,t}^2(b))=0$ is $\Gamma_{\tilde{\sigma}_{v,t}^2 b}db+\Gamma_{\tilde{\sigma}_{v,t}^2}d\tilde{\sigma}_{v,t}^2=0$. Thus, $\frac{d\tilde{\sigma}_{v,t}^2(b_t)}{db_t}=-\frac{\Gamma_{\tilde{\sigma}_{v,t}^2 b_t}}{\Gamma_{\tilde{\sigma}_{v,t}^2\tilde{\sigma}_{v,t}^2}}$. Combining this equation with $\Gamma_{b_t\tilde{\sigma}_{v,t}^2}=\Gamma_{\tilde{\sigma}_{v,t}^2 b_t}$, the investor's second order condition becomes

$$
\begin{aligned}
\frac{\partial^2\Gamma}{\partial b_t^2} &= \Gamma_{b_t b_t}(b_t,\tilde{\sigma}_{v,t}^2(b_t))+\Gamma_{b_t\tilde{\sigma}_{v,t}^2}(b_t,\tilde{\sigma}_{v,t}^2(b_t))\cdot\left(-\frac{\Gamma_{\tilde{\sigma}_{v,t}^2 b_t}(b_t,\tilde{\sigma}_{v,t}^2(b_t))}{\Gamma_{\tilde{\sigma}_{v,t}^2\tilde{\sigma}_{v,t}^2}(b_t,\tilde{\sigma}_{v,t}^2(b_t))}\right)+ \\
&\quad \Gamma_{\tilde{\sigma}_{v,t}^2\tilde{\sigma}_{v,t}^2}(b_t,\tilde{\sigma}_{v,t}^2(b_t))\cdot\left(-\frac{\Gamma_{\tilde{\sigma}_{v,t}^2 b}(b_t,\tilde{\sigma}_{v,t}^2(b_t))}{\Gamma_{\tilde{\sigma}_{v,t}^2\tilde{\sigma}_{v,t}^2}(b_t,\tilde{\sigma}_{v,t}^2(b_t))}\right)^2=\Gamma_{b_t b_t}(b_t,\tilde{\sigma}_{v,t}^2(b_t))\leq 0.
\end{aligned}
$$

This condition is unambiguously satisfied because $0\geq\Gamma_{b_t b_t}=$ $-\left[\left(-\gamma E_t^{\theta_t}\left(J_{t+1}\right)+\left(\gamma^2 b_t\right)Var_t^{\theta}\left(J_{t+1}\right)\right)^2+\gamma^2 Var_t^{\theta}\left(J_{t+1}\right)\right]\cdot\frac{\exp\left(\frac{1}{2}(\gamma b_t)^2 Var_t^{\theta_t}(J_{t+1})\right)}{\exp\left((\gamma b_t)E_t^{\theta_t}(J_{t+1})\right)}$. The second-order condition of nature's problem is

$$
\begin{aligned}
\frac{\partial^2\Gamma}{\partial\left(\tilde{\sigma}_{v,t}^2\right)^2} &= -\frac{\left(\gamma b_t\right)^2}{2}\left[\frac{1}{2}\left(\gamma b_t\right)^2\left(\frac{\partial Var_t^{\theta_t}\left(J_{t+1}\right)}{\partial\tilde{\sigma}_{v,t}^2}\right)^2+\frac{\partial^2 Var_t^{\theta_t}\left(J_{t+1}\right)}{\partial\left(\tilde{\sigma}_{v,t}^2\right)^2}\right] \\
&\quad \cdot\exp\left(-\gamma b_t E_t^{\theta_t}\left(J_{t+1}\right)+\frac{\left(\gamma b_t\right)^2}{2}Var_t^{\theta_t}\left(J_{t+1}\right)\right)+\frac{\lambda}{2}\frac{1}{\left(\tilde{\sigma}_{v,t}^2\right)^2}\geq 0
\end{aligned}
$$

It holds holds if and only if $\lambda\geq\lambda_t^*$, where λ_t^* is defined by

$$
\lambda_t^*\equiv\left(\tilde{\sigma}_{v,t}^2\gamma b_t\right)^2\cdot\left[\frac{\left(\gamma b_t\right)^2}{2}\left(\frac{\partial Var_t^{\theta_t}\left(J_{t+1}\right)}{\partial\tilde{\sigma}_{v,t}^2}\right)^2+\frac{\partial^2 Var_t^{\theta_t}\left(J_{t+1}\right)}{\partial\left(\tilde{\sigma}_{v,t}^2\right)^2}\right]\cdot\frac{\exp\left(\frac{(\gamma b_t)^2}{2}Var_t^{\theta_t}\left(J_{t+1}\right)\right)}{\exp\left(\gamma b_t E_t^{\theta_t}\left(J_{t+1}\right)\right)}
$$

The sign of λ_t^* equals the sign of the bracketed term. To derive this sign note that

$\frac{\partial^2 Var_t^{\theta_t}(J_{t+1})}{\partial\left(\tilde{\sigma}_{v,t}^2\right)^2} = -2a^2\left[k_{t+1}^{\theta_{t+1}}\beta_1+\beta_2\right]^2 \frac{\left(a^2\sigma_{t-1}^2+\sigma_w^2\right)^2}{\left(a^2\sigma_{t-1}^2+\sigma_w^2+\tilde{\sigma}_{v,t}^2\right)^3} \leq 0$. Since $\frac{1}{2}\left(\gamma b_t\right)^2\left(\frac{\partial Var_t^{\theta_t}(J_{t+1})}{\partial\tilde{\sigma}_{v,t}^2}\right)^2$ ≥ 0, this term has an ambiguous sign. Factoring $\left[k_{t+1}\beta_1+\beta_2\right]^2$ out, we have that the sign of the bracketed term in λ_t^* equals the sign of

$$\frac{\left(\gamma b_t\right)^2}{2}\left[k_{t+1}^{\theta_{t+1}}\beta_1+\beta_2\right]^2\left(a^2\frac{\left(a^2\sigma_{t-1}^2+\sigma_w^2\right)^2}{\left(a^2\sigma_{t-1}^2+\sigma_w^2+\tilde{\sigma}_{v,t}^2\right)^2}+1\right)^2-2a^2\frac{\left(a^2\sigma_{t-1}^2+\sigma_w^2\right)^2}{\left(a^2\sigma_{t-1}^2+\sigma_w^2+\tilde{\sigma}_{v,t}^2\right)^3}.$$

It then follows that λ_t^* is positive if $\left(\gamma b_t\right)^2$ is large enough, and negative (and not binding) if $\left(\gamma b_t\right)^2$ is small. Therefore, the lower bound for λ is

$$\lambda_t^v = \max\{\lambda_t^*, 0\}. \tag{8.4}$$

Now we have all the elements to characterize the solution to problem (4.2). The first order condition for b_t implies that

$$\begin{aligned} 0 &= -\left[\alpha_{t+1}+a\left(\beta_1+\beta_2\right)\hat{x}_t^{\theta_t}+e_t+\left(i_t-i_t^f\right)\right]-\gamma b_t Var_t^{\theta_t}\left(J_{t+1}\right). \text{ Thus,} \\ b_t(e_t,\tilde{\sigma}_{v,t}^{2\theta}) &= \frac{-\left[\alpha_{t+1}+a\left(\beta_1+\beta_2\right)\hat{x}_t^{\theta_t}-e_t-\left(i_t-i_t^f\right)\right]}{\gamma\cdot Var_t^{\theta_t}\left(J_{t+1}\right)} \end{aligned}$$

Equation (4.4) for b_t follows from this condition. The FOC for $\tilde{\sigma}_{v,t}^2$ implies that

$$\left(\frac{1}{\sigma_v^2}-\frac{1}{\tilde{\sigma}_{v,t}^2}\right)=\frac{\left(\gamma b_t\right)^2}{\lambda}\frac{\partial Var_t^{\theta_t}\left(J_{t+1}\right)}{\partial\tilde{\sigma}_{v,t}^2}\exp\left(-\gamma b_t E_t^{\theta_t}\left(J_{t+1}\right)+\frac{1}{2}\left(\gamma b_t\right)^2 Var_t^{\theta_t}\left(J_{t+1}\right)\right)\geq 0 \tag{8.5}$$

Since $Var^{\theta}\left(J_{t+1}|I_t\right)=\left[\left(k_{t+1}^{\theta_{t+1}}\beta_1+\beta_2\right)\right]^2\left(a^2\frac{\left(a^2\sigma_{t-1}^2+\sigma_w^2\right)\tilde{\sigma}_{v,t}^2}{a^2\sigma_{t-1}^2+\sigma_w^2+\tilde{\sigma}_{v,t}^2}+\sigma_w^2+\tilde{\sigma}_{v,t}^2\right)$, we have that

$$\frac{\partial Var_t^{\theta_t}\left(J_{t+1}\right)}{\partial\tilde{\sigma}_{v,t}^2}=\left[\left(k_{t+1}^{\theta_{t+1}}\beta_1+\beta_2\right)\right]^2\left(a^2\frac{\left(a^2\sigma_{t-1}^2+\sigma_w^2\right)^2}{\left(a^2\sigma_{t-1}^2+\sigma_w^2+\tilde{\sigma}_{v,t}^2\right)^2}+1\right)\geq 0,$$

which implies $\tilde{\sigma}_{v,t}^2 \geq \sigma_v^2$. Hence, we obtain the inequality $k_t^{\theta_t} = \frac{a^2\sigma_{t-1}^2+\sigma_w^2}{a^2\sigma_{t-1}^2+\sigma_w^2+\tilde{\sigma}_{v,t}^2} \leq k_t^{\theta'} = \frac{a^2\sigma_{t-1}^2+\sigma_w^2}{a^2\sigma_{t-1}^2+\sigma_w^2+\sigma_v^2}$. Equation (8.5) also implies that if either λ goes to infinity or $\gamma=0$, then the observation variance is not distorted: $\tilde{\sigma}_{v,t}^2=\sigma_v^2$.

Proof of Proposition 4.3. In equilibrium, the market-clearing condition $b_t^*(e_t^*)=b_t^s$

implies $e_t^* = -\left(i_t - i_t^f\right) + a\left(\beta_1^* + \beta_2^*\right)\hat{x}_t^{\theta^*} + \alpha_{t+1}^* + \gamma b_t^s Var_t^{\theta^*}\left(J_{t+1}\right)$. Equalizing coefficients with conjecture (3.7) we obtain $\beta_1^* = -\frac{a}{1-a}$, $\beta_2^* = -1$, $\alpha_t^* = \alpha_{t+1}^* + \gamma b_t^s Var_t^{\theta^*}\left(J_{t+1}\right)$. Furthermore, all variables in λ_t^v, the lower bound for λ, are evaluated at equilibrium values.

Proof of Lemma 5.1. We use the standard result in the control literature that if time starts in the infinite past, then σ_t^2 has converged to its steady-state value ξ^* for any time $t \in [0, T]$, where $\xi^* = \frac{-\left(\sigma_w^2 + \sigma_v^2 - a^2\sigma_v^2\right) + \sqrt{\left(\sigma_w^2 + \sigma_v^2 - a^2\sigma_v^2\right)^2 + 4a^2\sigma_w^2\sigma_v^2}}{2a^2}$ solves the steady-state equation of $\sigma_t^2 = \frac{\left(a^2\sigma_{t-1}^2 + \sigma_w^2\right)\sigma_v^2}{a^2\sigma_{t-1}^2 + \sigma_w^2 + \sigma_v^2}$. Since the supply of bonds b_t^s is a constant $\bar{b}$ on $[0, T]$, in equilibrium k^{θ^*} and $\tilde{\sigma}_v^{2^*}$ are jointly determined by the following system of equations

$$k^{\theta^*} = \frac{a^2\xi^* + \sigma_w^2}{a^2\xi^* + \sigma_w^2 + \tilde{\sigma}_v^{2^*}}, \qquad \text{and} \tag{8.6}$$

$$\begin{aligned} \lambda\left(\frac{1}{\sigma_v^2} - \frac{1}{\tilde{\sigma}_v^{2^*}}\right) &= \left[\gamma\bar{b}\left(\frac{ak^{\theta^*}}{1-a} + 1\right)k^{\theta^*}\right]^2 \cdot \left(a^2\frac{a^2\xi^* + \sigma_w^2}{a^2\xi^* + \sigma_w^2 + \tilde{\sigma}_v^{2^*}} + 1\right) \\ &\quad \cdot \exp\left(\frac{\left(\gamma\bar{b}\right)^2}{2}\left[\frac{ak^{\theta^*}}{1-a} + 1\right]^2\left(a^2k^{\theta^*}\tilde{\sigma}_v^{2^*} + \sigma_w^2 + \tilde{\sigma}_v^{2^*}\right)\right) \end{aligned} \tag{8.7}$$

Therefore, $Var_t^{\theta^*}\left(J_{t+1}\right)$ is also constant for $t \in [0, T]$: $Var_t^{\theta^*}\left(J_{t+1}\right) = \left[\frac{ak^{\theta^*}}{1-a} + 1\right]^2 \Phi^*$ with $\Phi^* = \left(a^2\frac{\left(a^2\xi^* + \sigma_w^2\right)\tilde{\sigma}_v^{2^*}}{a^2\xi^* + \sigma_w^2 + \tilde{\sigma}_v^{2^*}} + \sigma_w^2 + \tilde{\sigma}_v^{2^*}\right)$. Next, we show that for any fixed T there exists a finite α_t^* so that the exchange rate function does not explode. For $t \in [0, T]$,

$$\begin{aligned} \alpha_t &= \alpha_{t+1} + \frac{1}{2}\gamma b_t^s Var_t^{\theta^*}\left(J_{t+1}\right) = \alpha_{T+1} + \frac{1}{2}\gamma\sum_{i=1}^{T-t}\bar{b}\left[\frac{ak^{\theta^*}}{1-a} + 1\right]^2\Phi^* \\ &= \alpha_{T+1} + \frac{T-t}{2}\gamma\bar{b}\left[\frac{ak^{\theta^*}}{1-a} + 1\right]^2\Phi^* \end{aligned} \tag{8.8}$$

We will prove that α_{T+1} is finite so that α_t does not explode. Notice that $\alpha_{T+1} = \alpha_{T+j} + \frac{1}{2}\gamma\sum_{i=1}^{j} b_{T+i}^s Var_t^{\theta^*}\left(J_{T+i+1}\right)$. Letting j go to infinity, we have

$$\alpha_{T+1} = \alpha_\infty + \frac{1}{2}\gamma\sum_{i=1}^{\infty} b_{T+i}^s Var_t^{\theta^*}\left(J_{T+i+1}\right) \tag{8.9}$$

To find α_∞ note that (8.8) implies $\alpha_\infty = \alpha_\infty + \lim_{t\to\infty}\frac{1}{2}\gamma b_t^s Var_t^{\theta^*}\left(J_{t+1}\right)$, and that the

b_t^s process in (5.1) implies that for $t \in (T, \infty]$

$$\lim_{t\to\infty} \frac{1}{2}\gamma b_t^s Var_t^{\theta^*}(J_{t+1}) = \lim_{t\to\infty} \frac{1}{2}\gamma \bar{b} \exp\left(-(t-T)\right) \phi_t Var_t^{\theta^*}(J_{t+1}) = \lim_{t\to\infty} \frac{1}{2}\gamma \bar{b} \exp\left(-(t-T)\right) = 0.$$

Therefore, α_∞ exists and is equal to a constant $\alpha_\infty = c \in R$. Plugging back into equation (8.9), we have that

$$\alpha_{T+1} = c + \frac{1}{2}\gamma \sum_{i=1}^{\infty} b_{T+i}^s Var_t^{\theta^*}(J_{T+i+1}) = c + \frac{1}{2}\gamma \sum_{i=1}^{\infty} \bar{b} \exp(-i) = c + \tfrac{1}{2}\gamma \tfrac{\bar{b}}{1-\exp(-1)}.$$

This proves that $\alpha_{T+1} = c + \frac{1}{2}\gamma \frac{\bar{b}}{1-\exp(-1)}$ is finite. Hence, for any fixed T and $t \in [0, T]$, α_t^* can be expressed as

$$\begin{aligned} \alpha_t^* &= c + \frac{\gamma}{2}\frac{\bar{b}}{1-\exp(-1)} + \frac{T-t}{2}\gamma \bar{b}\left[\frac{ak^{\theta^*}}{1-a} + 1\right]^2 \Phi^* < \infty, \quad \text{with} \quad (8.10) \\ \Phi^* &= \left(a^2 \frac{(a^2\xi^* + \sigma_w^2)\,\tilde{\sigma}_v^{2^*}}{a^2\xi^* + \sigma_w^2 + \tilde{\sigma}_v^{2^*}} + \sigma_w^2 + \tilde{\sigma}_v^{2^*}\right) \text{ and } c \in R. \end{aligned}$$

Furthermore, the lower bound for λ is

$$\lambda^v = Max\{0, \lambda^*\} \tag{8.11}$$

where λ^* is λ_t^* in equation (8.4) evaluated at steady-state values.

Proof of Proposition 5.2. Let $y_0 = x_0 = \hat{x}_0^{\theta^*} = 0$ and recall that under the data generating process θ', $v_t \overset{\theta'}{\sim} N(0, \sigma_v^2)$ and $w_t \overset{\theta'}{\sim} N(0, \sigma_w^2)$ for $t = 1, 2, ...$ Since in equilibrium the exchange rate and the forecasts are linear in the initial shocks ε and κ, the average impulse response (5.3) can be expressed as

$$\begin{aligned} e_t^{av} &= E^{\theta'}(\varepsilon | y_1 = 1)\cdot e_t(\varepsilon, 0) + E^{\theta'}(\kappa | y_1 = 1)\cdot e_t(0, \kappa) - e_t(0,0) \\ &= q^{\theta'}\cdot e_t(\varepsilon, 0) + [1 - q^{\theta'}]\cdot e_t(0,\kappa) - e_t(0,0) \end{aligned} \tag{8.12}$$

The result $E(\varepsilon | y_1 = 1) = q^{\theta'}$ and $E(\kappa | y_1 = 1) = 1 - q^{\theta'}$ is proven below in Lemma 8.1. Next, we compute the impulse response to a transitory shock and to persistent shock separately. Then we compute the impulse response to an average shock.

IRF to a transitory shock. If we condition on a transitory shock at time $t = 1$, then

$v_1 = \kappa$ and all other shocks are zero. It follows that the data generated are $x_t = 0$ for all $t \geq 1$, $y_1 = \kappa$, and $y_t = 0$, for all $t \geq 2$. Plugging these data in filter (5.2), it follows that the estimate of the unobservable trend at time t is $\hat{x}_t = (a(1-k))^{t-1} k^{\theta^*} \kappa$. Plugging $\hat{x}_t$ and y_t into the exchange rate function, we have that the IRF to a transitory shock is

$$\hat{e}_t(0,\kappa) = \begin{cases} 0 & t = 0 \\ \alpha^* - \left(1 + \frac{ak^{\theta^*}}{1-a}\right)\kappa & t = 1 \\ \alpha^* - \frac{a^t k^{\theta^*}\left(1-k^{\theta^*}\right)^{t-1}}{1-a}\kappa & t \geq 2 \end{cases} \tag{8.13}$$

IRF to a persistent shock. If we condition on a persistent shock at time $t = 1$, then $w_0 = \varepsilon$ and all other shocks are zero. It follows that the data generated are $x_t = a^{t-1}\varepsilon$ and $y_t = a^{t-1}\varepsilon$ for all $t \geq 1$. Plugging these data in filter (5.2), it follows that the estimate of the unobservable trend at time t is $\hat{x}_t = a^{t-1}\left(1 - \left(1-k^{\theta^*}\right)^t\right)\varepsilon$. Plugging $\hat{x}_t$ and y_t into the exchange rate function, we have that the IRF to a persistent shock is

$$\hat{e}_t(\varepsilon, 0) = \alpha_t^* - \frac{a^{t-1}}{1-a}\left(1 - a\left(1-k^{\theta^*}\right)^t\right)\varepsilon, \qquad t \geq 1 \tag{8.14}$$

IRF to an average shock. Substituting (8.13) and (8.14) in (8.12) we have that

$$\begin{aligned} e_1^{av} &= \left[\alpha_1^* - \left(1 + \frac{ak^{\theta^*}}{1-a}\right)\right]\left(1-q^{\theta'}\right) + \left[\alpha^* - \frac{1}{1-a}\left(1 - a\left(1-k^{\theta^*}\right)\right)\right] q^{\theta'} - \alpha_1^* \\ &= -\left(1 + \frac{ak^{\theta^*}}{1-a}\right) + \left[1 + \frac{ak^{\theta^*}}{1-a} - \frac{1}{1-a}\left(1 - a\left(1-k^{\theta^*}\right)\right)\right] q^{\theta'} = -\left(1 + \frac{ak^{\theta^*}}{1-a}\right) \end{aligned}$$

For $t \geq 2$, we have that

$$\begin{aligned} e_t^{av} &= \left[\alpha_2^* - \frac{a^t k^{\theta^*}\left(1-k^{\theta^*}\right)^{t-1}}{1-a}\right]\left(1-q^{\theta'}\right) + \left[\alpha_2^* - \frac{a^{t-1}}{1-a}\left(1 - a\left(1-k^{\theta^*}\right)^t\right)\right] q^{\theta'} - \alpha_2^* \\ &= -\frac{a^t k^{\theta^*}\left(1-k^{\theta^*}\right)^{t-1}}{1-a} - \left[\frac{a^{t-1}}{1-a}\left(1 - a\left(1-k^{\theta^*}\right)^t\right) - \frac{a^t k^{\theta^*}\left(1-k^{\theta^*}\right)^{t-1}}{1-a}\right] q^{\theta'} \\ &= -\frac{a^t k^{\theta^*}\left(1-k^{\theta^*}\right)^{t-1}}{1-a} - \frac{a^{t-1}}{1-a}\left(1 - a\left(1-k^{\theta^*}\right)^{t-1}\right) q^{\theta'} = -\frac{a^{t-1}}{1-a}\left[q^{\theta'} + a\frac{k^{\theta^*} - q^{\theta'}}{\left(1-k^{\theta^*}\right)^{1-t}}\right] \end{aligned}$$

Therefore, the IRF to a y-shock is given by

$$e_t^{av} = \begin{cases} 0 & t=0 \\ -\frac{1-a\left(1-k^{\theta^*}\right)}{1-a} & t=1 \\ -\frac{a^{t-1}}{1-a}\left[q^{\theta'} + a\left(1-k^{\theta^*}\right)^{t-1}\left(k^{\theta^*}-q^{\theta'}\right)\right] & t \geq 2 \end{cases}$$

We obtain (5.5) in the Proposition by taking the first difference of this equation. Part (i) follows directly from (5.5) because $e_{t+1}^{av} - e_t^{av} < 0$ if and only if $k^{\theta^*} < q^{\theta'}$. Part (ii) follows because there exists a unique time τ, such that mean-reversion occurs after τ :

$$e_{t+1}^{av} - e_t^{av} \geq 0, \text{ for } t \geq \tau \;\Leftrightarrow\; [1-k^{\theta^*}]^{\tau-1}[1-a(1-k^{\theta^*})][q^{\theta'}-k^{\theta^*}] \leq q^{\theta'}[\frac{1}{a}-1].$$

τ is the smallest integer greater or equal to τ^*

$$\tau^* = 1 + \left[\log\left(q^{\theta'}[\frac{1}{a}-1]\right) - \log\left([1-a(1-k^{\theta^*})][q^{\theta'}-k^{\theta^*}]\right)\right]\left[\log\left(1-k^{\theta^*}\right)\right]^{-1}$$

Lemma 8.1. *Under the data generating process* $E^{\theta'}(\varepsilon|y_1=1) = \frac{1}{1+\sigma_v^2/\sigma_w^2} \equiv q^{\theta'}$.

Proof. Let $w_0 = \varepsilon$ and $v_1 = \kappa$, with $\varepsilon \sim (0,\sigma_w^2)$ and $\kappa \sim N(0,\sigma_v^2)$, and consider $y_1 = \varepsilon + \kappa$ a realization of a random variable with a distribution $y_1|\varepsilon \overset{\theta'}{\sim} N(\varepsilon,\sigma_v^2)$. If prior distribution is $\varepsilon \overset{\theta'}{\sim} N(0,\sigma_w^2)$, Bayes law implies that the posterior is

$$\varepsilon|y_1 \overset{\theta'}{\sim} N(m,n^2), \text{ with } m = (1-h)m_0 + hy_1,\; n^2 = (1-h)n_0^2,\; h = \frac{n_0^2}{n_0^2+\tilde{\sigma}_v^2}.$$

Since $m_0 = 0$ and $n_0^2 = \sigma_w^2$, it follows that $E^{\theta'}(\varepsilon|_{y_1=1}) = \frac{\sigma_w^2}{\sigma_w^2+\sigma_v^2}y_1 = \frac{1}{1+\sigma_v^2/\sigma_w^2}$.

Derivation of (5.9). Given the exchange rate function $e_{t+1} = \alpha_{t+1} - \frac{a}{1-a}\hat{x}_{t+1}^{\theta^*} - y_{t+1}$ and updating formula $\hat{x}_{t+1}^{\theta^*} = \left(1-k^{\theta^*}\right)a\hat{x}_t^{\theta^*} + k^{\theta^*}y_{t+1}$, it follows that $e_{t+1} = \alpha_{t+1} - \frac{a^2}{1-a}\left(1-k^{\theta^*}\right)\hat{x}_t^{\theta^*} - \left(k^{\theta^*}\frac{a}{1-a}+1\right)y_{t+1}$. Conditioning on information I_t, and taking expectations under the robust measure θ^*, as well as under the baseline θ', we have

$$\begin{aligned} E_t^{\theta^*}(e_{t+1}) &= \alpha_{t+1} - \frac{a^2}{1-a}\left(1-k^{\theta^*}\right)\hat{x}_t^{\theta^*} - \left(k^{\theta^*}\frac{a}{1-a}+1\right)E_t^{\theta^*}(y_{t+1}) \\ E_t^{\theta'}(e_{t+1}) &= \alpha_{t+1} - \frac{a^2}{1-a}\left(1-k^{\theta^*}\right)\hat{x}_t^{\theta*} - \left(k^{\theta^*}\frac{a}{1-a}+1\right)E_t^{\theta'}(y_{t+1}) \end{aligned}$$

Equation (5.9) follows from $E_t^{\theta'}(y_{t+1}) = E_t^{\theta'}(x_{t+1})$ and $E_t^{\theta^*}(y_{t+1}) = E_t^{\theta^*}(x_{t+1})$

$$\begin{aligned}\Lambda_t &= E_t^{\theta^*}(e_{t+1}) - e_t + \zeta_t - E_t^{\theta'}(e_{t+1} - e_t) = \left[E_t^{\theta'}(x_{t+1}) - E_t^{\theta^*}(x_{t+1})\right]\left[1 + \frac{ak^{\theta^*}}{1-a}\right] + \zeta_t \\ &= \left[E_t^{\theta'}(x_{t+1}) - E_t^{\theta^*}(x_{t+1})\right]\left[1 + \frac{ak^{\theta^*}}{1-a}\right] + \zeta_t\end{aligned}$$

Proof of Proposition 5.3. From the robust uncovered interest parity condition (4.5) it follows that $E_t^{\theta^*}(\Delta e_{t+1}) = y_t - \zeta_t$, where $y_t \equiv i_t - i_t^f$. If we define the forecast error as $\upsilon_{t+1} \equiv \Delta e_{t+1} - E_t^{\theta^*}(\Delta e_{t+1})$, it follows that

$$\begin{aligned}cov^{\theta'}(\Delta e_{t+1}, y_t) &= cov^{\theta'}(E_t^{\theta^*}(\Delta e_{t+1}) + \upsilon_{t+1}, y_t) = cov^{\theta'}((y_t - \zeta_t) + \upsilon_{t+1}, y_t) \\ &= var^{\theta'}(y_t) + cov^{\theta'}(\upsilon_{t+1},\ y_t) \qquad (8.15)\end{aligned}$$

The last step follows because the bond's supply process (5.1) implies that the premium ζ_t is deterministic (by Lemma 5.1), so $cov^{\theta'}(\zeta_t,\ y_t) = 0$. Let's develop $cov^{\theta'}(\upsilon_{t+1},\ y_t)$.

$$cov^{\theta'}(\upsilon_{t+1}, y_t) = cov^{\theta'}(\Delta e_{t+1} - E_t^{\theta^*}(\Delta e_{t+1}), y_t) = cov^{\theta'}(\zeta_t - \Lambda_t, y_t) = -cov^{\theta'}(\Lambda_t, y_t) \qquad (8.16)$$

The second equality follows from (5.9): $\Lambda_t = [E_t^{\theta^*}(\Delta e_{t+1}) - E_t^{\theta'}(\Delta e_{t+1})] + \zeta_t$. The third equality follows because the premium ζ_t is deterministic. Replacing (8.16) in (8.15) we get $\beta^{Fama} = 1 - \lim_{t\to\infty} \frac{cov^{\theta'}(\Lambda_t, y_t)}{var^{\theta'}(y_t)}$. Notice that we can reexpress $cov^{\theta'}(\Lambda_t,\ y_t)$ as

$$\begin{aligned}cov^{\theta'}(\Lambda_t, y_t) &= \xi \cdot cov^{\theta'}(E_t^{\theta'}(x_{t+1}) - E_t^{\theta^*}(x_{t+1}), y_t), \quad \xi := \left(k^{\theta^*}\frac{a}{1-a} + 1\right) \\ &= \xi \cdot cov^{\theta'}\left(a\left(\hat{x}_t^{\theta'} - \hat{x}_t^{\theta^*}\right), y_t\right) = a\xi \cdot cov^{\theta'}\left(\hat{x}_t^{\theta'} - \hat{x}_t^{\theta^*}, y_t\right)\end{aligned}$$

Here, we have used the fact that $E_t^{\theta'}(x_{t+1}) = a\hat{x}_t^{\theta'}$ and $E_t^{\theta^*}(x_{t+1}) = a\hat{x}_t^{\theta^*}$ because $E_t^{\theta'}(w_t) = 0$ and $E_t^{\theta^*}(w_t) = 0$. Next, we compute $cov^{\theta'}\left(\hat{x}^{\theta^*}, y_t\right)$.

$$\begin{aligned}cov^{\theta'}\left(\hat{x}_t^{\theta^*}, y_t\right) &= cov^{\theta'}\left(\hat{x}_t^{\theta^*}, x_t\right) + k^{\theta^*}\sigma_v^2 = cov^{\theta'}\left(a\left(1 - k^{\theta^*}\right)\hat{x}^{\theta^*}_{t-1} + ky_t, x_t\right) + k^{\theta^*}\sigma_v^2 \\ &= a^2\left(1 - k^{\theta^*}\right)cov^{\theta'}\left(\hat{x}_{t-1}^{\theta^*}, x_{t-1}\right) + k^{\theta^*}Var^{\theta'}(x_t).\end{aligned}$$

Using the stationarity of $cov^{\theta'}(\hat{x}_t, y_t)$ and the stationary value of $Var^{\theta'}(x) = \frac{\sigma_w^2}{1-a^2}$

$$\begin{aligned}
cov^{\theta'}(\hat{x}, y) &\equiv \lim_{t\to\infty} cov^{\theta'}\left(\hat{x}_t^{\theta^*}, y_t\right) = \frac{k^{\theta^*}}{1-a^2}\frac{1}{1-a^2\left(1-k^{\theta^*}\right)}\sigma_w^2 + k^{\theta^*}\sigma_v^2 \\
cov^{\theta'}\left(\hat{x}^{\theta'}, y\right) &\equiv \lim_{t\to\infty} cov^{\theta'}\left(\hat{x}_t^{\theta'}, y_t\right) = \frac{k^{\theta'}}{1-a^2}\frac{1}{1-a^2\left(1-k^{\theta'}\right)}\sigma_w^2 + k^{\theta'}\sigma_v^2
\end{aligned}$$

Finally, by substituting back in β^{Fama} we obtain

$$\beta^{Fama} = 1 - \lim_{t\to\infty}\frac{a\xi\cdot cov^{\theta'}\left(\hat{x}_t^{\theta'} - \hat{x}_t^{\theta^*}, y_t\right)}{var^{\theta'}(y_t)} = 1 - a\xi\frac{cov^{\theta'}(\hat{x}, y) - cov^{\theta'}\left(\hat{x}^{\theta'}, y\right)}{\lim_{t\to\infty} var^{\theta'}(y_t)}.$$

Therefore, $\beta^{Fama} = 1 - \frac{K_1}{K_2}$, where $K_2 := \frac{\sigma_w^2}{1-a^2} + \sigma_v^2$, and

$$\begin{aligned}
K_1 &= a\left(\frac{ak^{\theta^*}}{1-a} + 1\right)\left[\frac{k^{\theta'}}{1-a^2}\frac{1}{1-a^2\left(1-k^{\theta'}\right)}\sigma_w^2 - \frac{k^{\theta^*}}{1-a^2}\frac{1}{1-a^2\left(1-k^{\theta^*}\right)}\sigma_w^2 + \left(k^{\theta'} - k^{\theta^*}\right)\sigma_v^2\right] \\
&= a\left(\frac{ak^{\theta^*}}{1-a} + 1\right)\left[\frac{1}{1-a^2}\left(\frac{k^{\theta'}}{1-a^2\left(1-k^{\theta'}\right)} - \frac{k^{\theta^*}}{1-a^2\left(1-k^{\theta^*}\right)}\right)\sigma_w^2 + \left(k^{\theta'} - k^{\theta^*}\right)\sigma_v^2\right] \\
&= a\left(\frac{ak^{\theta^*}}{1-a} + 1\right)\left(\frac{1}{\left(1-a^2\left(1-k^{\theta'}\right)\right)\left(1-a^2\left(1-k^{\theta^*}\right)\right)}\sigma_w^2 + \sigma_v^2\right)\left(k^{\theta'} - k^{\theta^*}\right)
\end{aligned}$$

It follows that

$$\begin{aligned}
\frac{K_1}{K_2} &= \frac{a\left(\frac{ak^{\theta^*}}{1-a} + 1\right)\left(\frac{1}{\left(1-a^2\left(1-k^{\theta'}\right)\right)\left(1-a^2\left(1-k^{\theta^*}\right)\right)}\sigma_w^2 + \sigma_v^2\right)\left(k^{\theta'} - k^{\theta^*}\right)}{\sigma_v^2 + \frac{\sigma_w^2}{1-a^2}} \\
&= \frac{a\left(a(1+a)k^{\theta^*} + (1-a^2)\right)\left(\frac{1}{\left(1-a^2\left(1-k^{\theta'}\right)\right)\left(1-a^2\left(1-k^{\theta^*}\right)\right)} + \frac{\sigma_v^2}{\sigma_w^2}\right)\left(k^{\theta'} - k^{\theta^*}\right)}{\frac{\sigma_v^2}{\sigma_w^2}(1-a^2) + 1}
\end{aligned}$$

Part 1 of Proposition 5.3 follows directly from the fact that for any $a \in (0,1)$ the sign of $K_1(k)$ equals the sign of $k^{\theta'} - k^{\theta^*}$. To prove part 2, we need to show that $K_1/K_2 > 1$ for some large a and that $K_1/K_2 < 1$ for some low a.

$$\lim_{a\uparrow 1}\frac{K_1}{K_2} = 2\bar{k}\left(\bar{k}' - \bar{k}\right)\left(\frac{1}{\bar{k}\bar{k}'} + \frac{\sigma_v^2}{\sigma_w^2}\right) = 2\left[1 - \frac{\bar{k}}{\bar{k}'} + \bar{k}\bar{k}'\frac{\sigma_v^2}{\sigma_w^2} - \bar{k}^2\frac{\sigma_v^2}{\sigma_w^2}\right],$$

where $\bar{k}$ and $\bar{k}'$ denote k^{θ^*} and $k^{\theta'}$ evaluated at $a = 1$ and $k^{\theta'} = \frac{a^2\sigma^2+\sigma_w^2}{a^2\sigma^2+\sigma_w^2+\sigma_v^2}$ and $k^{\theta^*} = \frac{a^2\sigma^2+\sigma_w^2}{a^2\sigma^2+\sigma_w^2+\tilde{\sigma}_v^{2*}}$, where $\sigma^2 = \frac{-\left(\sigma_w^2+\sigma_v^2-a^2\sigma_v^2\right)+\sqrt{(\sigma_w^2+\sigma_v^2-a^2\sigma_v^2)^2+4a^2\sigma_w^2\sigma_v^2}}{2a^2}$, and $\tilde{\sigma}_v^{2*}$ solves the first order condition (8.2). It follows from (8.2) that $\lim_{\lambda\to\lambda^v} \tilde{\sigma}_v^{2*} = \infty$, where λ^v is the lowest admissible bound for λ, so that the SOC is satisfied. Since $\lim_{\lambda\to\lambda^v} k^{\theta^*} = 0$ for all a, we have that

$$\lim_{\lambda\to\lambda^v} \lim_{a\uparrow 1} \frac{K_1}{K_2} = 2\left[1 - \frac{0}{\bar{k}'} + 0 \cdot \bar{k}' \frac{\sigma_v^2}{\sigma_w^2} - 0 \cdot \frac{\sigma_v^2}{\sigma_w^2}\right] = 2$$

Since $\tilde{\sigma}_v^{2*}$ and k^{θ^*} are continuous functions of λ, it follows that there is a $\underline{\tilde{\sigma}_v^{2*}}$ such that $\lim_{a\uparrow 1} \frac{K_1}{K_2} > 1$ for any $\tilde{\sigma}_v^{2*} > \underline{\tilde{\sigma}_v^{2*}}$. This proves part 2. To prove part 3 note that $\frac{K_1}{K_2}|_{a=0} = 0$ and observe that $\frac{K_1}{K_2}$ is continuous on $a \in (-1, 1)$. Thus, there exists a neighborhood $r_a = (-\eta, \eta)$ of a around zero, such that $\frac{K_1}{K_2} < \varepsilon << 1$, whenever $a \in r_a$, where ε is an arbitrary small number. In other words, when $a \in (0, \eta)$, we always have $\beta^{Fama} = 1 - \frac{K_1}{K_2} > 0$.

Proof of Lemma 6.1. This proof follows the same steps at those in the proof of Lemma 4.1 and can be found in the extended appendix in our websites.

Proof of Proposition 6.2. The proof is similar to that of Lemma 4.2 and can be found in the extended appendix. The only difference with respect to the case of observational uncertainty is that under an alternative model $\theta \in \Theta^w$ the variance of returns is not given by (8.1), but by

$$Var_t^{\theta_t}\left(J_{t+1}\right) = \left[\left(k_{t+1}^{\theta_{t+1}}\beta_1 + \beta_2\right)\right]^2 \left(a^2 \frac{\left(a^2\sigma_{t-1}^2 + \tilde{\sigma}_{w,t}^2\right)\sigma_v^2}{a^2\sigma_{t-1}^2 + \sigma_v^2 + \tilde{\sigma}_{w,t}^2} + \sigma_v^2 + \tilde{\sigma}_{w,t}^2\right). \tag{8.17}$$

In this case the second order condition for $\tilde{\sigma}_{w,t}^2$ holds if and only if $\lambda \geq \lambda_t^w$, where λ_t^w is defined by

$$\lambda_t^w = \max\left\{0, \lambda_t^{**}\right\}, \quad \text{with} \tag{8.18}$$

$$\lambda_t^{**} \equiv \left(\tilde{\sigma}_{w,t}^2\gamma b_t\right)^2 \cdot \left[\frac{(\gamma b_t)^2}{2}\left(\frac{\partial Var_t^{\theta_t}\left(J_{t+1}\right)}{\partial\tilde{\sigma}_{w,t}^2}\right)^2 + \frac{\partial^2 Var_t^{\theta_t}\left(J_{t+1}\right)}{\partial\left(\tilde{\sigma}_{w,t}^2\right)^2}\right] \cdot \frac{\exp\left(Var_t^{\theta_t}\left(J_{t+1}\right)(\gamma b_t)^2/2\right)}{\exp\left(\gamma b_t E_t^{\theta_t}\left(J_{t+1}\right)\right)}$$

Proof of Lemma 6.3. The proof is similar to that of Lemma 4.1 and can be found in the extended appendix.

Proof of Proposition 6.4. The proof is similar to that of Lemma 4.2 and can be found in the extended appendix. In this case the variance of returns is

$$V_t^{\theta_t}(J_{t+1}) = \left(k_{t+1}^{\theta_{t+1}}\beta_1 + \beta_2\right)^2 \left[(a+\delta_t)^2 \frac{\left((a+\delta_t)^2 \sigma_{t-1}^2 + \sigma_w^2\right)\sigma_v^2}{(a+\delta_t)^2 \sigma_{t-1}^2 + \sigma_w^2 + \sigma_v^2} + \sigma_w^2 + \sigma_v^2\right] \tag{8.19}$$

The second order condition for δ_t holds if and only if $\lambda \geq \lambda_t^a$, where λ_t^a is defined by

$$\lambda_t^a = \max\left\{0, \lambda_t^{***}\right\}, \quad \text{where} \tag{8.20}$$

$$\lambda_t^{***} \equiv \frac{(\gamma b_t)^2}{\left(\hat{x}_t^{\theta_t}\right)^2+1} \cdot \left[\frac{(\gamma b_t)^2}{2}\left(\frac{\partial Var_t^{\theta_t}(J_t)}{\partial \delta_t}\right)^2 + \frac{\partial^2 Var_t^{\theta_t}(J_{t+1})}{\partial \delta_t^2}\right] \cdot \exp\left(-\gamma b_t E_t^{\theta_t}(J_{t+1}) + \frac{(\gamma b_t)^2}{2} Var_t^{\theta_t}(J_{t+1})\right) \tag{8.21}$$

Proof of Lemma 6.5. We use a representation theorem in Dupuis and Ellis (1997) to show that if θ' is the baseline model and $\theta \in \Theta^u$, then the agent's problem simplifies to the RHS in (6.9).

Lemma 8.2 (Variational Representation). *Let $(\Omega, \mathcal{I})$ be a measurable space, and let $\mathcal{P}(\Omega)$ denote the set of all probability measures defined on it. If f is a bounded measurable function mapping from Ω into $\mathbb{R}$, and $\theta \in P(\Omega)$, then; (a) the following variational formula holds:*

$$-\log \int_\Omega e^{-f(\omega)}\theta(d\omega) = \inf_{\gamma \in \mathcal{P}(\Omega)} \left\{R(\gamma||\theta) + \int_\Omega f(\omega)\gamma(d\omega)\right\} \tag{8.22}$$

(b) The infimum in (8.22) is uniquely achieved by a probability measure γ^0 which is absolutely continuous with respect to θ and has the Radon-Nikodym derivative

$$\frac{d\gamma^0}{d\theta} \equiv e^{-f(\omega)} \frac{1}{\int_\Omega e^{-f(\omega)}\theta(d\omega)}$$

Applying the above Lemma to our robust utility function, we have that

$\Gamma_t = \max_{b_t} \inf_{\theta \in P(\Omega)} \left\{E_t^\theta\left[u(W_{t+1}) + \lambda \cdot \Re(\theta||\theta')\right]\right\} = \max_{b_t}\left\{-\lambda \log\left(E_t^{\theta'} \exp\left(-\frac{1}{\lambda}u(W_{t+1})\right)\right)\right\}.$

Proof of (6.10). From the first order condition for b_t we have

$$E_t^{\theta'}\left[-J_{t+1}\exp\left(\gamma b_t(J_{t+1})\right)\exp\left(-\frac{1}{\lambda}\exp\left(\gamma b_t(J_{t+1})\right)\right)\right] = 0$$

where $J_{t+1} \equiv \left(i_t - i_t^f\right) - (e_{t+1} - e_t)$. Next, we use *Stein's Lemma* (see for example, Casella and Berger [14] page 187), which states that for a normally distributed random variable $X \sim N(\mu, \sigma^2)$, and a function $g(x)$ with $E|g'(X)| < \infty$, we have $E[g(X)(X-\mu)] = \sigma^2 E g'(X)$. Starting with the first order condition for b_t, the conjecture implies that J_{t+1} has a conditional normal distribution $N\left(E_t^{\theta'}\left(\left(i_t - i_t^f\right) - (e_{t+1} - e_t)\right), Var_t^{\theta'}(e_{t+1})\right)$. Thus, let

$$g(J_{t+1}) = -\exp(\gamma b_t J_{t+1}) \exp\left(-\frac{1}{\lambda}\exp(\gamma b_t J_{t+1})\right)$$

The derivative of $g(J_{t+1})$ is $g'(J_{t+1}) = \left[1 - \frac{1}{\lambda}\exp(\gamma b_t J_{t+1})\right]\gamma b_t g(J_{t+1})$. It follows that the right hand side of the first order condition can be written as

$$\begin{aligned} E_t^{\theta'}(J_{t+1} g(J_{t+1})) &= E_t^{\theta'}\left(J_{t+1} - E_t^{\theta'}(J_{t+1}) + E_t^{\theta'}(J_{t+1})\right) g(J_{t+1}) \\ &= E_t^{\theta'}\left(J_{t+1} - E_t^{\theta'}(J_{t+1})\right) g(J_{t+1}) + E_t^{\theta'}(J_{t+1}) E_t^{\theta'} g(J_{t+1}) \end{aligned}$$

Applying Stein's Lemma to the right side of the first order condition, we have that $0 = Var_t^{\theta'}(e_{t+1}) E g'(J_{t+1}) + E_t^{\theta'}(J_{t+1}) E_t^{\theta'} g(J_{t+1})$. Therefore, $E_t^{\theta'}(-J_{t+1}) = \frac{Var_t^{\theta'}(e_{t+1}) E g'(J_{t+1})}{E_t^{\theta'} g(J_{t+1})}$. Hence, $E_t^{\theta'}(-J_{t+1}) - l_t = 0$ or $E_t^{\theta'}\left(e_{t+1} - e_t + i_t^f - i_t\right) - l_t = 0$, where $l_t \equiv \frac{Var_t^{\theta'}(e_{t+1}) E_t^{\theta'} g'(J_{t+1})}{E_t^{\theta'} g(J_{t+1})}$, $g(J_{t+1}) = -\exp(\gamma b_t J_{t+1}) \exp\left(-\frac{1}{\lambda}\exp(\gamma b_t J_{t+1})\right)$, and $g'(J_{t+1}) = \left[1 - \frac{1}{\lambda}\exp(\gamma b_t J_{t+1})\right]\gamma b_t g(J_{t+1})$. Thus,
$l_t = \frac{Var_t^{\theta'}(e_{t+1}) E_t^{\theta'}\left[\left(1 - \frac{1}{\lambda}\exp\left(-\gamma b_t\left(e_{t+1} - e_t + i_t^f - i\right)\right)\right)\gamma b_t \exp\left(-\gamma b_t\left(e_{t+1} - e_t + i_t^f - i\right)\right)\exp\left(-\frac{1}{\lambda}\exp\left(-\gamma b_t\left(e_{t+1} - e_t + i_t^f - i\right)\right)\right)\right]}{E_t^{\theta'}\left[\exp\left(-\gamma b_t\left(e_{t+1} - e_t + i_t^f - i_t\right)\right)\exp\left(-\frac{1}{\lambda}\exp\left(-\gamma b_t\left(e_{t+1} - e_t + i_t^f - i_t\right)\right)\right)\right]}$.

Proof of Proposition 6.6. From the conjecture (3.7) $e_t = \alpha_t + \beta_1 \hat{x}_t + \beta_2\left(i_t - i_t^f\right)$

$$\begin{aligned} E_t^{\theta'}(e_{t+1}) &= \alpha_{t+1} + \beta_1 E_t^{\theta'} \hat{x}_{t+1} + \beta_2 E_t^{\theta'}\left(i_t - i_{t+1}^f\right) = \alpha_{t+1} + (\beta_1 + \beta_2) E_t^{\theta'} \hat{x}_{t+1} \\ &= \alpha_{t+1} + (\beta_1 + \beta_2) E_t^{\theta'}\left((1 - k_t) a \hat{x}_t + k_t\left(i_{t+1} - i_{t+1}^f\right)\right) = \alpha_{t+1} + a(\beta_1 + \beta_2)\hat{x}_t \end{aligned}$$

In equilibrium, $b_t = b_t^s = \bar{b}$, and the exchange rate function must satisfy $E_t^{\theta'}\left(e_{t+1} - e_t + i_t^f - i_t\right) - l_t = 0$. Plugging e_t and $E_t^{\theta'}(e_{t+1})$ into the exchange rate function, we have $0 = \alpha_{t+1} - \alpha_t - l_t + [a(\beta_1 + \beta_2) - \beta_1]\hat{x}_t - (\beta_2 + 1)(i_t - i_t^f)$. Since the equation must hold for any $\hat{x}_t$ and $i_t - i_t^f$, it follows that $a(\beta_1 + \beta_2) - \beta_1 = 0$, $\beta_2 + 1 = 0$, $\alpha_{t+1} - \alpha_t + l_t = 0$. Hence, the coefficients of the equilibrium exchange rate function are $\beta_1^* = -\frac{a}{1-a}$, $\beta_2^* = -1$, and $\alpha_{t+1}^* = \alpha_t^* + l_t^*$, where l_t^* is l_t evaluated at $\left(\beta_1^*, \beta_2^*, \bar{b}\right)$.

References

[1] Alvarez, F., A. Atkeson and P. Kehoe, 2007, "Time-Varying Risk, Interest Rates, and Exchange Rates in General Equilibrium," mimeo.

[2] Bacchetta, P. and E. van Wincoop, 2006, "Can Information Heterogeneity Explain the Exchange Rate Determination Puzzle?," *American Economic Review*.

[3] Bacchetta, P. and E. van Wincoop, 2008, "Infrequent Portfolio Decisions: A Solution to the Forward Discount Puzzle," mimeo.

[4] Backus, D., S. Foresi and C. Telmer, 2001, "Term Structure Models and the Forward Premium Anomaly," *Journal of Finance,* 56(1), 279-304.

[5] Bansal R. and M. Dahlquist, 2000, "The Forward Premium Puzzle: Different Tales from Developed and Emerging Economies," *Journal of International Economics*.

[6] Basar, T. and P. Bernhard, 1995, $\mathcal{H}_\infty$- *Optimal Control and Related Minimax Design Problems*, Birkhauser.

[7] Bekaert, G., 1996, "The Time-Variation of Risk and Return in Foreign Exchange Markets: A General Equilibrium Perspective," *Review of Financial Studies.*

[8] Bodie, Zvi, Alex Kane, Alan Marcus, *Investments*, 7th Ed, McGraw-Hill Irwin.

[9] Bossaerts, P., P. Ghirardato, S. Guarneschelli and W. Zame, 2004, "The Effect of Ambiguity on Prices and Choices in Financial Markets," mimeo.

[10] Brock, W., S. Durlauf and K. West, 2007, "Model Uncertainty and Policy Evaluation: Some Theory and Empirics," *Journal of Econometrics*, 136, 2, 629-664.

[11] Burnside, C., M. Eichenbaum, and S. Rebelo, 2007, "Understanding the Forward Premium Puzzle: A Microstructure Approach," mimeo.

[12] Burnside, C., M. Eichenbaum, I. Kleshchelski and S. Rebelo, 2008, "Do Peso Problems explain the Returns to the Carry Trade?" mimeo.

[13] Cagetti, M., L. Hansen, T. Sargent, and N. Williams, 2002, "Robustness and Pricing with Uncertain Growth," *The Review of Financial Studies*, Vol. 15, No. 2.

[14] Casella, G. and R. Berger, *Statistical Inference*.

[15] Chinn, M. 2006, "The (Partial) Rehabilitation of Interest Rate Parity in the Floating Rate Era," *Journal of International Money and Finance* 25, 7-21.

[16] Cogley, T. and T. Sargent, 2005, "The Conquest of U.S. Inflation: Learning and Robustness to Model Uncertainty," *Review of Economic Dynamics.*

[17] Denis, R., 2007, "Model Uncertainty and Monetary Policy," mimeo.

[18] Dornbusch, R., 1976, “Expectations and Exchange Rate Dynamics,” *Journal of Political Economy*, 84, 1161-1176.

[19] Dupuis, P. and R. Ellis, 1997, *A weak convergence approach to the theory of large deviations*, John Wiley & Sons.

[20] Durlauf, S. W. Brock and K. West, 2007, "Model Uncertainty and Policy Evaluation: Some Theory and Empirics," *Journal of Econometrics*, 136, 2, 629-664.

[21] Eichenbaum, M. and C. Evans, 1995, “Some Empirical Evidence on the Effects of Monetary Policy Shocks on Exchange Rates,” *Quarterly Journal of Economics.*

[22] Engel, C., 1996, "The Forward Discount Anomaly and the Risk Premium: A Survey of Recent Evidence," *Journal of Empirical Finance* 3, 123-192.

[23] Engel C., and K. West, 2005, "Exchange Rates and Fundamentals," *Journal of Political Economy* 113, 485-517.

[24] Epstein, L. and M. Schneider, "Recursive Multiple Priors," *Journal of Economic Theory*, 2003, 1-31.

[25] Epstein, L. and T. Wang, I, 1994, "Intertemporal Asset Pricing under Knightian Uncertainty," *Econometrica*, Vol. 62, No. 2, pp. 283-322.

[26] Evans, M.and R. Lyons, 2002, "Order Flow and Exchange Rate Dynamics," *Journal of Political Economy*, 110, 1.

[27] Fama, E., 1984, "Forward and Spot Exchange Rates," *J. of Monetary Economics.*

[28] Farhi, E. and X. Gabaix, 2008,"Rare Disasters and Exchange Rates", mimeo.

[29] Frankel, J., and C. Engel, 1984, "Do Asset-Demand Functions Optimize Over the Mean and Variance of Real Returns? A Six-Currency Test," *Journal of International Economics* 17 309-323.

[30] Frankel, J. and K. Froot, 1989, "Forward Discount Bias: Is it an Exchange Risk Premium?" *The Quarterly Journal of Economics*, 104, 1, 139-161.

[31] Foland, G, 2001, *Real Analysis- Modern Techniques and Their Applications.*

[32] Friend, I. and M. Blume, 1975, "The Demand for Risky Assets," *American Economic Review*, Vol. 65, No. 5, pp. 900-922.

[33] Gilboa, I. and D. Schmeidler, 1989, Maxmin Expected Utility with Non-unique Prior, *Journal of Mathematical Economics*, 18, 141-153.

[34] Giannoni, M., 2002, "Does Model Uncertainty Justify Caution? Robust Optimal Monetary Policy in a Forward-Looking Model," *Macroeconomic Dynamics.*

[35] Gourinchas, P. and A. Tornell, 2004, “Exchange Rate Dynamics and Misperception,” *Journal of International Economics.*

[36] Grossman, S. and R. Shiller, 1981, "The Determinants of the Variability of Stock Market Prices," *The American Economic Review*, Papers and Proceedings.

[37] Hansen, L. and T. Sargent, 2007, *Robustness*, forthcoming, Princeton U. Press.

[38] Hansen, L., T. Sargent, G. Turmuhambetova, and N. Williams, 2006, "Robust Control, min-max Expected Utility and Model Misspecification," *Journal of Economic Theory.*

[39] Hansen, L., T. Sargent, and T. Tallarini, 1999, "Robust Permanent Income and Pricing," *Review of Economic Studies,* 66 , 873-907.

[40] Kaminski, G., 1993, "Is There a Peso Problem? Evidence from the Dollar/Pound Exchange Rate", *American Economic Review*, 83, 450-472.

[41] Kasa, K., 2002, "Model uncertainty, robust policies, and the value of commitment. *Macroeconomic Dynamics* 6, 145–166.

[42] Lewis, K., 1989, "Changing Beliefs and Systematic Rational Forecast Errors with Evidence from Foreign Exchange", *American Economic Review*, 79, 621-636.

[43] Lewis, K., 1995, "Puzzles in International Financial Markets," in Gene M. Grossman and Kenneth Rogoff (eds), *Handbook of International Economics.*

[44] Lustig, H. and Verdelhan, A., 2007, "The Cross-Section of Foreign Currency Risk Premia and US Consumption Growth Risk," *American Economic Review*, 97(1).

[45] Mark, N., 1995, "Exchange Rates and Fundamentals: Evidence on Long-Horizon Predictability," *American Economic Review*, 85, 201-218.

[46] Mark, Nelson and Yangru Wu, 1998, "Rethinking Deviations from Uncovered Interest Parity: The Role of Covariance Risk and Noise," *The Economic Journal.*

[47] Onatski, A. and J. Stock, 2002, Robust monetary policy under model uncertainty in a small model of the U.S. economy, *Macroeconomic Dynamics* 6, 85–110.

[48] Rigotti, L. and C. Shannon, 2005, "Uncertainty and Risk in Financial Markets," *Econometrica*, Vol. 73, No. 1, pp. 203-243.

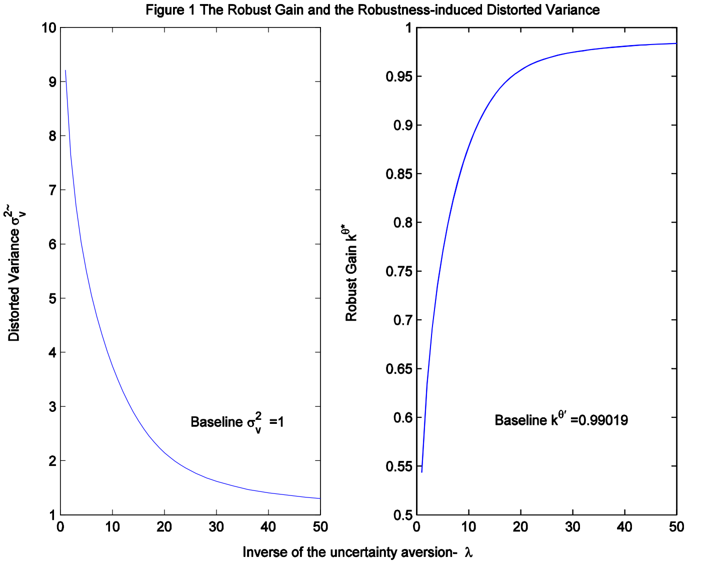
[49] Sarno, L., G. Leon, G. Valente and H. Leon, 2006, "Nonlinearity in Deviations from Uncovered Interest Parity: An Explanation of the Forward Bias Puzzle," *Review of Finance*, 10, 443-482.

[50] Svensson, L. and N. Williams, 2007, "Bayesian and Adaptive Optimal Policy under Model Uncertainty," mimeo.

[51] Tetlow, R. and P. von zur Muehlen, 2004, "Avoiding Nash Inflation: Bayesian and robust responses to model uncertainty" *Review of Economic Dynamics.*

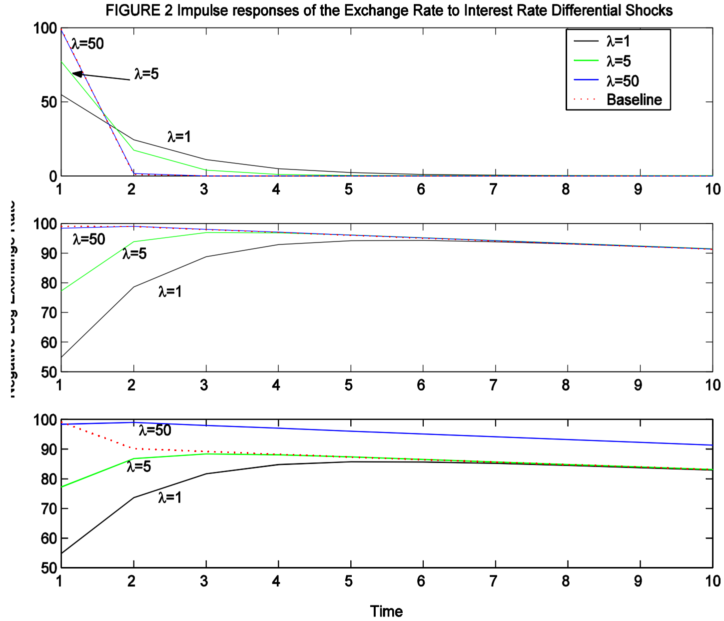
[52] Tornell, A., 2001, "Robust H-infinity Forecasting and Asset Pricing Anomalies," mimeo.

Figure 1 The Robust Gain and the Robustness-induced Distorted Variance



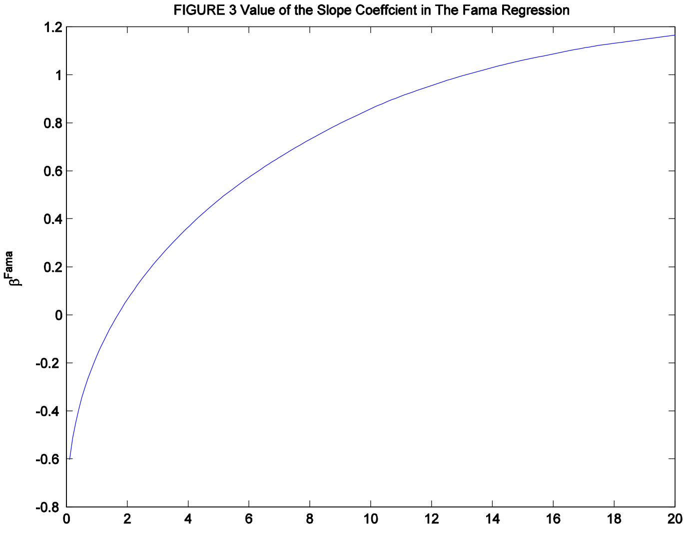
Note. This figure plots $\tilde{\sigma}_v^{2*}$and $k^{\theta*}$ solved from equations 8.6 and 8.7 by setting γ=2.6, a=0.99,σ_v^2 =1 and σ_w^2=100.

FIGURE 2 Impulse responses of the Exchange Rate to Interest Rate Differential Shocks



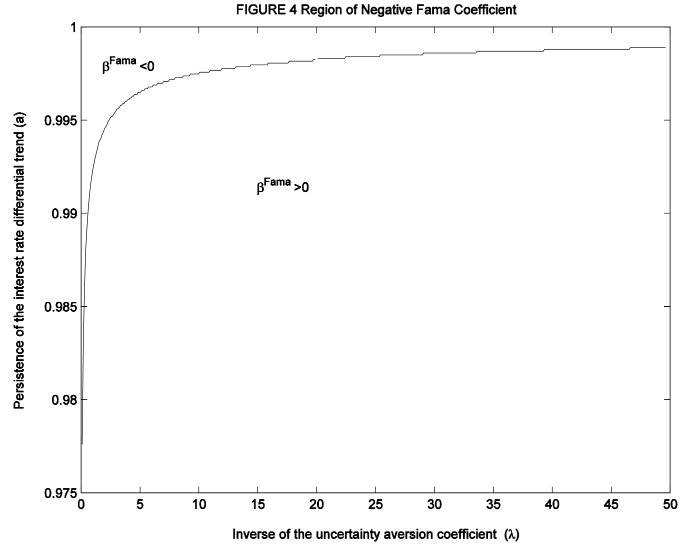
Note. This figure plots the impulse responses to a transitory shock (γ_1=κ), to a persistence shock (γ_1=ε) and to a shock of unknown duration (γ_1=κ+ε=1) by setting γ=2.6, a=0.99, σ_v^2 =1 and σ_w^2=100. Notice that the appreciation shows up as an upward move in the graphs because we plot -log(E).

FIGURE 3 Value of the Slope Coeffcient in The Fama Regression

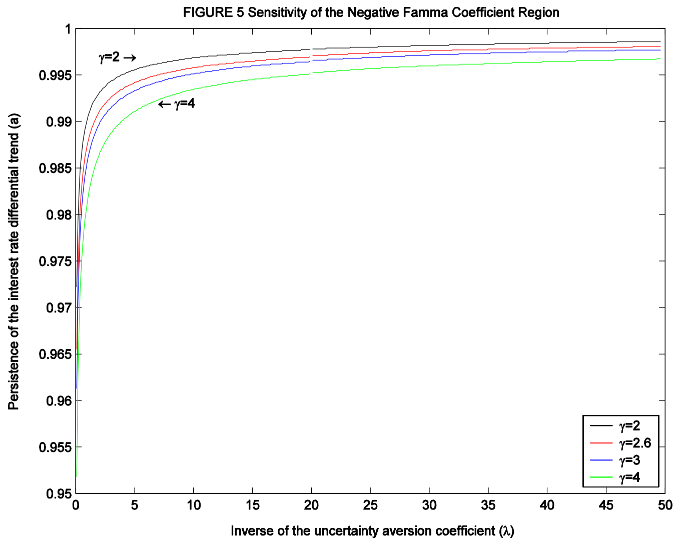


Note. This figure plots the theoretical Fama coefficient in equation (5.7) by setting γ=2.6, a=0.99, σ_v^2 =1 and σ_w^2=100.

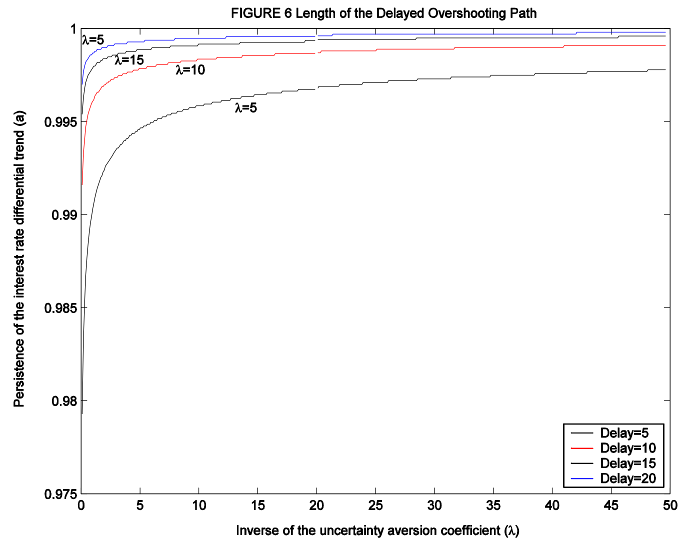
FIGURE 4 Region of Negative Fama Coefficient



Note. This figure plots the sign of theoretical Fama coefficient in equation (5.7) for different pairs (a,λ) by setting γ=2.6, σ_v^2 =1 and σ_w^2=100.



Note. This figure plots the sign of theoretical Fama coefficient (5.7) for different pairs (a,λ) by letting the absolute risk-aversion coefficient (γ) range over the interval [2,4] and setting σ_v^2 =1 and σ_w^2=100.



Note. This figure plots the Length of the Delayed Overshooting Path in (5.6) for different pairs (a,λ) by setting γ=2.6, σ_v^2 =1 and σ_w^2=100.