Ambiguity Aversion: Implications for the Uncovered Interest Rate Parity Puzzle*

Cosmin L. Ilut
Duke University

First version: October 2008
Revised version: February 2010

Abstract

Empirically, high-interest-rate currencies tend to appreciate in the future relative to low-interest-rate currencies instead of depreciating as uncovered interest rate parity (UIP) states. The explanation for the UIP puzzle that I pursue in this paper is that the agents' beliefs are systematically distorted. This perspective receives some support from an extended empirical literature using survey data. I construct a model of exchange rate determination in which ambiguity-averse agents need to solve a filtering problem to form forecasts but face signals about the time-varying hidden state that are of uncertain precision. In the presence of such uncertainty, ambiguity-averse agents take a worst-case evaluation of this precision and respond stronger to bad news than to good news about the payoffs of their investment strategies. Importantly, because of this endogenous systematic underestimation, agents in the next periods will perceive on average positive innovations about the payoffs which will make them re-evaluate upwards the profitability of the strategy. As a result, the model's dynamics imply significant ex-post departures from UIP as equilibrium outcomes. In addition to providing a resolution to the UIP puzzle, the model predicts, consistent with the data, positive profitability, negative skewness and excess kurtosis for payoffs of currency speculation.

Key Words: uncovered interest rate parity, ambiguity aversion, robust filtering.
JEL Classification: D8, E4, F3, G1.

*I would like to thank Gadi Barlevy, Larry Christiano, Eddie Dekel, Martin Eichenbaum, Lars Hansen, Peter Kondor, Nicolas Lehmann-Ziebarth, Jonathan Parker, Giorgio Primiceri, Sergio Rebelo, Tom Sargent, Martin Schneider, Tomasz Strzalecki and seminar participants at the AEA Annual Meeting 2010, MNB-CEPR 8th Macroeconomic Policy Research Workshop (2009), SED 2009, Stanford University (SITE, 2009) and Board of Governors, Chicago Fed, Duke, ECB, New York Fed, NYU, Northwestern, Philadelphia Fed, UC Davis, UC Santa Cruz, Univ. of Virginia for helpful discussions and comments. Correspondence: Department of Economics, Duke University. E-mail: cosmin.ilut@duke.edu

1 Introduction

According to uncovered interest rate parity (UIP), periods when the domestic interest rate is higher than the foreign interest rate should on average be followed by periods of domestic currency depreciation. An implication of UIP is that a regression of realized exchange rate changes on interest rate differentials should produce a coefficient of one. This implication is strongly counterfactual. In practice, UIP regressions (Hansen and Hodrick (1980), Fama (1984)) produce coefficient estimates well below one and sometimes even negative.[1] This anomaly is taken very seriously because the UIP equation is a property of most open economy models. This failure, referred to as the UIP puzzle or the forward premium puzzle[2], implies that traders who borrow in low-interest-rate currencies and lend in high-interest-rate currencies (a strategy known as the "carry trade") make positive profits on average. The standard approach in addressing the UIP puzzle has been to assume rational expectations and time-varying risk premia. This approach has been criticized in two ways: survey evidence has been used to cast doubt on the rational expectations assumption[3] and other empirical research challenges the risk implications of the analysis.[4]

In this paper, I follow a conjecture in the literature that the key to understanding the UIP puzzle lies in departing from the rational expectations assumption.[5] I pursue this conjecture formally, using the assumption that agents are not endowed with complete knowledge of the true data generating process (DGP) and that they confront this uncertainty with ambiguity aversion. I model ambiguity aversion along the lines of the maxmin expected utility (or multiple priors) preferences as in Gilboa and Schmeidler (1989).

I present a model of exchange rate determination which features signal extraction by an ambiguity averse agent that is uncertain about the precision of the signals she receives. The only source of randomness in the environment is the domestic/foreign interest rate differential. I model this as an exogenous stochastic process, which is the sum of unobserved persistent and transitory components. I assume that the agent does not know the variances of

[1] Among recent studies see Chinn and Frankel (2002), Gourinchas and Tornell (2004), Chinn and Meredith (2005), Verdelhan (2008) and Burnside et al. (2008).

[2] Under covered interest rate parity the interest rate differential equals the forward discount. The UIP puzzle can then be restated as the observation that currencies at a forward discount tend to appreciate.

[3] For example, Froot and Frankel (1989), Chinn and Frankel (2002) and Bacchetta et al. (2008) find that most of the predictability of currency excess returns is due to expectational errors.

[4] See Lewis (1995) and Engel (1996) for surveys on this research. See Burnside et al. (2008) for a critical review of recent risk-based explanations. These criticisms are by no means definitive as there is a recent risk-based theoretical literature, including for example Verdelhan (2008), Bansal and Shaliastovich (2007), Alvarez et al. (2008) and Farhi and Gabaix (2008) that argues that the typical empirical exercises are unable by construction to capture the underlying time-variation in risk.

[5] Froot and Thaler (1990), Eichenbaum and Evans (1995), Lyons (2001) and Gourinchas and Tornell (2004) argue that models where agents are slow to respond to news may explain the UIP puzzle.

the innovations in the temporary and persistent components and she allows for the possibility that those variances change over time. In other words, the agent perceives the signals she receives about the hidden persistent state as having uncertain precision or quality.[6] Under ambiguity aversion with maxmin expected utility, the agent simultaneously chooses a belief about the model parameter values and a decision about how many bonds to buy and sell. The bond decision maximizes expected utility subject to the chosen belief and the budget constraint. The belief is chosen so that, conditional on the agent's bond decision, expected utility is minimized subject to a particular constraint. The constraint is that the agent only considers an exogenously-specified finite set of values for the variances. I choose this set so that, in equilibrium, the variance parameters selected by the agent are not implausible in a likelihood ratio sense.

In equilibrium, the agent invests in the higher interest rate bond (investment currency) by borrowing in the lower interest rate bond (funding currency). The larger the estimate of the hidden state of the investment differential, i.e. the differential between the high-interest-rate and the low-interest-rate, the larger is her demand for this strategy is. Conditional on this decision, the agent's expected utility is decreasing in the expected future depreciation of the investment currency. In equilibrium, this depreciation will be stronger when the future demand for the investment currency is lower. Thus, the agent is concerned that the observed investment differential in the future is low which makes the agent worry that the estimate of the hidden state of the investment differential is low. As a result, the initial concern for a depreciation translates into the agent tending to underestimate, compared to the true DGP, the hidden state of this differential. When faced with signals of uncertain precision, ambiguity averse agents act cautiously and underestimate the hidden state by reacting asymmetrically to news: they believe that it is more likely that observed increases in the investment differential have been generated by temporary shocks (low precision of signals) while decreases as reflecting more persistent shocks (high precision of signals).[7] The UIP condition holds ex-ante under these endogenously pessimistic beliefs.

Because the agent underestimates, compared to the true DGP, the persistent component of the investment differential she is, on average, surprised next period by observing a higher investment differential than expected. Under her equilibrium subjective beliefs, these innovations are unexpected good news that increase the estimate of the hidden state. This updating effect creates the possibility that next period the agent finds it optimal to invest

[6]The structure of uncertainty that I investigate, namely signals of uncertain precision, is similar to Epstein and Schneider (2007, 2008). The main difference is that here I consider time-varying hidden states, which generates important dynamics, while their model analyzes a constant hidden parameter.

[7]This asymmetric response to news has been investigated in a static filtering environment by Epstein and Schneider (2008) and Illeditsch (2009).

even more in the investment currency because this higher estimate raises the present value of the future payoffs of investing in the higher interest rate bond. The increased demand will drive up the value of the investment currency contributing to a possible appreciation of the investment currency. Thus, an investment currency could see a subsequent equilibrium appreciation instead of a depreciation as UIP predicts.

The main result of this paper is that such a model of exchange rate determination has the potential to resolve the UIP puzzle. Indeed, for the benchmark calibration, numerical simulations show that in large samples the UIP regression coefficient is negative and statistically significant while in small samples it is mostly negative and statistically not different from zero. The model is based on maximum likelihood estimates of interest rate differentials for developed countries which suggest a high degree of persistence of the hidden state and a large signal to noise ratio for the true DGP. In the benchmark specification, I impose some restrictions on the frequency and magnitudes of the distortions that the agent is considering so that the equilibrium distorted sequence of variances is difficult to distinguish statistically from the true DGP based on a likelihood comparison.[8] Studying other parameterizations, I find that the UIP regression coefficient becomes positive, even though smaller than one, if the true DGP is characterized by a significantly less persistent hidden state or much larger temporary shocks than the benchmark specification.

The gradual incorporation of good news implied by this model can directly account also for the delayed overshooting puzzle. This is an empirically documented impulse response in which following a positive shock to the domestic interest rate the domestic currency experiences a gradual appreciation for several periods instead of an immediate appreciation and then a path of depreciation as UIP implies.[9] For such an experiment, the ambiguity averse agent invests in the domestic currency in equilibrium and thus is worried about its future depreciation. The equilibrium beliefs then imply that the agent tends to overweigh, compared to the true DGP, the possibility that the observed increase in the interest rate reflects the temporary shock. This leads to an underestimation of the hidden state and generates a gradual incorporation of the initial shock into the estimate and the demand of the ambiguity-averse agent. The gradual incorporation of news can generate the gradual appreciation of the high-interest-rate currency.

The intuition for the model's ability to explain the UIP puzzle is related to Gourinchas and Tornell (2004) who show that if, for some unspecified reason, the agent systematically

[8]Eliminating these constraints would qualitatively maintain the same intuition and generate stronger quantitative results at the expense of the agent seeming less interested in the statistical plausibility of her distorted beliefs.

[9]See Eichenbaum and Evans (1995), Grilli and Roubini (1996), Faust and Rogers (2003) and Scholl and Uhlig (2006).

underreacts to signals about the time-varying hidden-state of the interest rate differential this can explain the UIP and the delayed overshooting puzzle. The main difference is that here I investigate a model which addresses the origin and optimality of such beliefs. This model generates endogenous underreaction only to good news, with the agent in fact overreacting to bad news. As an alternative model to generate an endogenous slow response to news, Bacchetta and van Wincoop (2009) use ideas from the rational inattention literature. In their setup, since information is costly to acquire and to process, some investors optimally choose to be inattentive and revise their portfolios infrequently. Their model also implies that agents respond symmetrically to news.

The explanation for the UIP puzzle proposed in this paper relies on placing structure on the type of uncertainty that the agent is concerned about. The agent receives signals of uncertain precision about a time-varying hidden state but otherwise she trusts the other elements of her representation of the DGP. Because of the structured uncertainty, the equilibrium distorted belief is not equivalent to the belief generated by simply increasing the risk aversion and using the rational expectations assumption.

Besides providing an explanation for the UIP puzzle, the theory for exchange rate determination proposed in this paper has several implications for the carry trade. First, directly related to the resolution of the UIP puzzle, the benchmark calibration produces, as in the data, positive average payoffs for the carry trade strategy. Compared to the empirical evidence, the model implied payoffs are smaller and less variable. The model generates positive average payoffs because in equilibrium the subjective probability distribution differs from the objective one by overpredicting bad events and underpredicting good events.

Second, the model implies that carry trade payoffs are characterized by negative skewness and excess kurtosis. This is consistent with the data as recent evidence (Brunnermeier et al. (2008)) suggests that high interest rate currencies tend to appreciate slowly but depreciate suddenly. In my model, an increase in the high-interest-rate compared to the market's expectation produces, relative to rational expectations, a slower appreciation of the investment currency since agents underreact to this type of innovations. However, a decrease in the high interest rate generates a relatively sudden depreciation because agents respond quickly to that type of news.[10] The excess kurtosis is a manifestation of this diminished reaction to good news. The asymmetric response to news is also consistent with the high frequency reaction of exchange rates to fundamentals documented in Andersen et al. (2003). The theory presented in this paper is also consistent with recent empirical findings

[10]Brunnermeier et al. (2008) argue that the data suggests that the realized skewness is related to the rapid unwinding of currency positions, a feature that is replicated by my model. They propose shocks to funding liquidity as a mechanism for this endogeneity.

documented in Jurek (2008) about the conditional time-variation of the payoff's skewness for currency trading.

Third, the model has strong implications for modified carry trade strategies that can improve their ex-post profitability by delivering higher Sharpe ratios. Intuitively, because the equilibrium in the model is characterized by a gradual incorporation of "good news", positive innovations in the investment differential will make the investment currency more likely to appreciate ex-post. In fact, the greater the positive innovations are, the higher the likelihood is of observing ex-post positive payoffs for the strategy. To test these implications, I implement empirically such modified carry trade strategies by following the rule that that the agent invests in the higher interest rate currency when the innovation in the investment differential is above a specific non-negative threshold. I find that in the data such strategies deliver significantly higher Sharpe ratios. Moreover, as predicted by the model, the empirical average ex-post profitability is increasing in the threshold used for conditioning the investment strategy.

The remainder of the paper is organized as follows. Section 2 describes and discusses the model. Section 3 presents a rational expectations version of the model to be contrasted against the ambiguity averse version studied in Section 4. Section 5 presents the model implications for exchange rate determination and discusses alternative specifications. Section 6 concludes. In the Appendix, I provide details on some of the model's equations and statements.

2 Model

2.1 Basic Setup

The basic setup is a typical one good, two-country, dynamic general equilibrium model of exchange rate determination. The focus is to keep the model as simple as possible while retaining the key ingredients needed to highlight the role of ambiguity aversion and signal extraction. For that purpose I will start with a model of risk-neutral, but otherwise ambiguity averse agents.

There are overlapping generations (OLG) of investors who each live two periods, derive utility from end-of-life wealth and are born with zero endowment. There is one good for which purchasing power parity (PPP) holds: $p_t = p_t^* + s_t$, where p_t is the log of price level of the good in the Home country and s_t is the log of the nominal exchange rate defined as the price of the Home currency per unit of foreign currency (FCU). Foreign country variables are indicated with a star. There are one-period nominal bonds in both currencies issued by

the respective governments. Domestic and foreign bonds are in fixed supply in the domestic and foreign currency respectively.

The Home and Foreign nominal interest rates are i_t and i_t^* respectively. The exogenous process is the interest rate differential $r_t = i_t - i_t^*$. The true DGP is the state-space model:

$$\begin{aligned} r_t &= x_t + \sigma_V v_t \\ x_t &= \rho x_{t-1} + \sigma_U u_t \end{aligned} \tag{2.1}$$

The shocks u_t and v_t are both Gaussian white noise. Thus, at time t the observable differential r_t is the sum of a hidden unobservable persistent component, x_t, and a temporary component, $\sigma_V v_t$. The agent entertains the possibility that the true DGP lies in a set of models (i.e. probability distributions over outcomes). The specific assumptions about the subjective beliefs of the agents regarding this process are covered in the next section.

Investors born at time t have are risk-neutral over end-of-life wealth, W_{t+1}, and face a convex cost of capital. In a later section I will address some of the implications of incorporating risk aversion. Their maxmin expected utility at time t is:

$$V_t = \max_{b_t} \min_{\widetilde{P} \in \Lambda} E_t^{\widetilde{P}}[(W_{t+1} - \frac{c}{2} b_t^2) | I_t] \tag{2.2}$$

where I_t is the information available at time t, b_t is the amount of foreign bonds invested and c controls the cost of capital. Agents have a zero endowment and pursue a zero-cost investment strategy: borrowing in one currency and lending in another. Since PPP holds, Foreign and Home investors face the same real returns and therefore will choose the same portfolio.

The set Λ comprises the alternative probability distributions available to the agent. The agent decides which of the the distributions (models) in the set Λ to use in forming their subjective beliefs about the future exchange rate. I postpone the discussion about the optimization over these beliefs to the next sections, noting that the optimal choice for b_t is made under the subjective probability distribution $\widetilde{P}$.

The amount b_t is expressed in domestic currency (USD). To illustrate the investment position suppose that b_t is positive. That means that the agent has borrowed b_t in the domestic currency and obtains $b_t \frac{1}{S_t}$ FCU units, where $S_t = e^{s_t}$. This amount is then invested in foreign bonds and generates $b_t \frac{1}{S_t} \exp(i_t^*)$ of FCU units at time $t+1$. At time $t+1$, the agent has to repay the interest bearing amount of $b_t \exp(i_t)$. Thus, the agent has to exchange back the time $t+1$ proceeds from FCU into USD and obtains $b_t \frac{S_{t+1}}{S_t} \exp(i_t^*)$. The net end-of-life

wealth is then a function of the amount of bonds invested and the excess return:

$$W_{t+1} = b_t \exp(i_t)[\exp(s_{t+1} - s_t + i_t^* - i_t) - 1]$$

An approximation around the steady state of $i_t = i_t^* = 0$ and $s_{t+1} - s_t = 0$ allows the net end-of-life wealth to be rewritten in the more tractable form:

$$W_{t+1} = b_t[s_{t+1} - s_t - (i_t - i_t^*)]$$

To close the model, I specify a Foreign bond market clearing condition similar to Bacchetta and van Wincoop (2009). There is a fixed supply B of Foreign bonds in the Foreign currency. In the steady state, the investor holds no assets since she has a zero endowment. The steady state amount of bonds is held every period by some unspecified traders. They can be interpreted as liquidity traders that have a constant bond demand. The real supply of Foreign bonds is $Be^{-p_t^*} = Be^{s_t}$ where the Home price level is normalized at one. I also normalize the steady state log exchange rate to zero. Thus, the market clearing condition is:

$$b_t = Be^{s_t} - B \tag{2.3}$$

where B is the steady state amount of Foreign bonds. Following Bacchetta and van Wincoop (2009) I also set $B = 0.5$, corresponding to a two-country setup with half of the assets supplied domestically and the other half supplied by the rest of the world. By log-linearizing the RHS of (2.3) around steady state I get the market clearing condition:

$$b_t = .5s_t \tag{2.4}$$

2.2 Model uncertainty

In this paper, the key departure from the standard framework of rational expectations is that I drop the assumption that the shock processes are random variables with known probability distributions. The agent will entertain various possibilities for the data generating process (DGP). She will choose, given the constraints, an optimally distorted distribution for the exogenous process. I will refer to this distribution as the *distorted* model. The objective probability distribution (the true DGP) is assumed to be the constant volatility state-space representation for the exogenous process r_t defined in (2.1). As in the model of multiple priors (or MaxMin Expected Utility) of Gilboa and Schmeidler (1989), the agent

chooses beliefs about the stochastic process that induce the lowest expected utility under that subjective probability distribution. The minimization is constrained by a particular set of possible distortions because otherwise the agent would select infinitely pessimistic probability distributions. Besides beliefs, the agent also selects actions that, under these worst-case scenario beliefs, maximize expected utility.

In the present context the maximizing choice is over the amount of foreign bonds that the agents is deciding to hold, while the minimization is over elements of the set Λ that the agent entertains as possible. The set Λ dictates how I constrain the problem of choosing an optimally distorted model. The type of uncertainty that I investigate is similar to Epstein and Schneider (2007, 2008), except that here I consider time-varying hidden states, while their model analyzes a constant hidden parameter. The agent believes that the standard deviation of the temporary shock is potentially time-varying and is drawn every period from a set Υ. Typical of ambiguity aversion frameworks, the agent's uncertainty manifests in her cautious approach of not placing probabilities on this set. Every period she thinks that any draw can be made out of this set. The agent trusts the remaining elements of the representation in (2.1). Thus the agent uses the following state-space representation:

$$\begin{aligned} r_t &= x_t + \sigma_{V,t} v_t \\ x_t &= \rho x_{t-1} + \sigma_U u_t \end{aligned} \tag{2.5}$$

where v_t and u_t are both Gaussian white noise and $\sigma_{V,t}$ are draws from the set Υ.

The information set is $I_t = \{r_{t-s}, s = 0, ..., t\}$. Using different realizations for the $\sigma_{V,s}$ for various dates $s \leq t$ will imply different posteriors about the hidden state x_t and the future distribution for r_{t+j}, $j > 0$. In equation (2.2) the unknown variable at time t is the realized exchange rate next period. This endogenous variable will depend in equilibrium on the probability distribution for the exogenous interest rate differential. Thus in choosing her pessimistic belief the agent will imagine what could be the worst-case realizations for $\sigma_{V,s}$, $s \leq t$, for the data that she observes. This minimization then reduces to selecting a sequence of

$$\sigma_V^t = \{\sigma_{V,s}, s \leq t : \sigma_{V,s} \in \Upsilon\} \tag{2.6}$$

in the product space $\Upsilon^t : \Upsilon \times \Upsilon ... \Upsilon$. As in Epstein and Schneider (2007), the agent interprets this sequence as a "theory" of how the data were generated.[11] The optimization in (2.2)

[11]Note that the distorted model is not a constant volatility model with a different value for the standard deviation of the shocks than the reference model. Although this possibility is implicitly nested in the setup, the optimal choice will likely be different because sequences with time variation will induce a lower utility for the agent.

then becomes:

$$V_t = \max_{b_t} \min_{\sigma_V^*(r^t) \in \sigma_V^t} E_t^{\widetilde{P}}[(W_{t+1} - \frac{c}{2} b_t^2) | I_t] \tag{2.7}$$

where $\widetilde{P}$ still denotes the subjective probability distribution implied by the known elements of the DGP and the distorted optimal sequence $\sigma_V^*(r^t)$. The latter is a function of time t information which is represented by the history of observables r^t.

For simplicity, I consider the case in which the set Υ contains only three elements: $\sigma_V^L < \sigma_V < \sigma_V^H$. As in Epstein and Schneider (2007), to control how different the distorted model is from the true DGP, I include the value σ_V in the set Υ.[12] I will refer to the sequence $\sigma_V^t = \{\sigma_{V,s} = \sigma_V, s \leq t\}$ as the *reference* model, or reference sequence. The set Υ contains a lower and a higher value than σ_V to allow for the possibility that for some dates s the realization $\sigma_{V,s}$ induces a higher or lower precision of the signal about the hidden state. Given the structure of the model, the worst-case choice is monotonic in the values of the set Υ. Thus, it suffices to consider only the lower and upper bounds of this set.[13]

The proposed structure of uncertainty makes the resulting ambiguity aversion equilibrium observationally non-equivalent to an expected utility model but higher risk aversion. The literature on optimal control has shown that simply invoking robustness reduces in some cases to solving the model under expected utility but with a higher risk aversion.[14] Hansen (2007), Hansen and Sargent (2007), Hansen and Sargent (2008a) and Ju and Miao (2009) show that, in general, a concern for misspecification in the hidden state of the process breaks this link and produces qualitatively different dynamics than simply increasing risk aversion. The setup proposed in this paper can be viewed as an example of such dynamics.

Here the concern for possible misspecification is over the variance of the shocks, which can be thought of as a latent unobserved state, a layer deeper than the hidden state x_t. It is worth noting however that in this model introducing a concern for the uncertainty surrounding r_t and x_t, without any further structure on that uncertainty is still equivalent to simply using expected utility and a higher risk aversion.[15] In Appendix A I present some details for this equivalence in this model. As I show in Section 5.1, higher risk aversion combined with

[12] This does not necessarily imply that σ_V is a priori known. If the agent uses maximum likelihood for a constant volatility model, her point estimate would be asymptotically σ_V.

[13] A more complicated version of the setup could be to have stochastic volatility with known probabilities of the draws as the reference model. The distorted set will then refer to the unwillingness of the agent to trust those probabilities. As above, she will then place time-varying probabilities on these draws. Similar intuition would then apply.

[14] See among others Whittle (1981), Strzalecki (2007) and Barillas et al. (2009).

[15] In the general setup of Hansen and Sargent (2008a) the structured uncertainty proposed here is over "models", controlled by the alternative sequences of variances σ_V^t, each implying a different evolution for the hidden state x_t. When uncertainty is directly over x_t, (the T_2 operator in Hansen and Sargent (2008a)) or over r_t (the T_1 operator in Hansen and Sargent (2008a)) a multiplier preference as in Hansen and Sargent (2008b) reduces in this model to expected utility and a higher risk aversion.

rational expectations does not provide an explanation for the puzzles in a simple mean-variance setup. I then conclude that the type of uncertainty studied in this paper, of signals with uncertain precision produces, in the simple model analyzed here, dynamics that are qualitatively different from the ones obtained with unstructured uncertainty or expected utility but higher risk aversion.

2.3 Statistical constraint on possible distortions

An important question that arises in this setup is how easy it is to distinguish statistically the optimal distorted sequence from the reference one. The robust control literature approaches this problem by using the multiplier preferences in which the distorted model is effectively constrained by a measure of relative entropy to be in some distance of the reference model. The ambiguity aversion models also constrain the minimization by imposing some cost function on this distance.[16] Without some sort of penalty for choosing an alternative model, the agent would select an infinitely pessimistic belief.

I also impose this constraint to avoid the situation in which the implied distorted sequence results in a very unlikely interpretation of the data compared to the true reference model. To quantify the statistical distance between the two models I use a comparison between the log-likelihood of a sample $\{r^t\}$ computed under the reference sequence, $L^{DGP}(r^t)$, and under the distorted optimal sequence, $L^{Dist}(r^t)$. This distance will be increasing in the number of dates s for which the distorted sequence $\sigma_V^{*t} = \{\sigma_{V,s}, s \leq t : \sigma_{V,s} \in \Upsilon\}$ is different from the true sequence $\sigma_V^t = \{\sigma_{V,s} = \sigma_V, s \leq t\}$.

Taking as given a desired average statistical performance of the distorted model and given the set Υ, the constraint effectively restricts the elements in the sequence σ_V^t to be different from the reference model only for a constant number n of dates. Intuitively, if n is low then the two sequences will produce relatively close likelihoods for the sample r^t even though the sample is increasing with t. For example if $n = 2$, as in the main parameterization, it means that the agent, who is interested in the statistical plausibility of her alternative model,[17] chooses only two dates when she is concerned that the realizations of $\sigma_{V,t}$ differ from σ_V. This approach of setting n low can also be interpreted as the agent viewing the possible alternative realizations in σ_V^t as "rare events" compared to the "normal" times of having $\sigma_{V,s} = \sigma_V$.

[16] See Anderson et al. (2003), Maccheroni et al. (2006) and Hansen and Sargent (2008b) among others.

[17] Clearly, the detection error probability is not directly a measure of the level of the agent's uncertainty aversion but only a tool to assess its statistical plausibility. For a discussion on how to recover in general ambiguity aversion from experiments see Strzalecki (2007). For a GMM estimation of the ambiguity aversion parameter for the multiplier preferences see Benigno (2007) and Kleshchelski and Vincent (2007).

For future reference let the restricted sequence be denoted by $\sigma_V^t(r^t)$:

$$\sigma_V^t(r^t) = \{\sigma_{V,s}\}_{s\leq t},\ \sigma_V^t(r^t) \in \{\Upsilon^n \times \sigma_V \times ... \times \sigma_V\} \tag{2.8}$$

The notation $\sigma_V^t(r^t) \in \{\Upsilon^n \times \sigma_V \times ... \times \sigma_V\}$ reflects the fact that $\sigma_V^t(r^t)$ belongs in the product space $\{\underbrace{\Upsilon \times ... \times \Upsilon}_{n \text{ times}} \times \underbrace{\sigma_V \times ... \times \sigma_V}_{t-n \text{ times}}\}$

Part of the optimization over the distorted sequence can be thought of selecting an order out of possible permutations. Let $P(t,n)$ denote the number of possible permutations where t is the number of elements available for selection and n is the number of elements to be selected. This order controls the dates at which the agent is entertaining values of the realized standard deviation that are different than σ_V. After selecting this order the rest of the sequence consists of elements equal to σ_V. As $P(t,n) = t!/(t-n)!$, this number of possible permutations increases significantly with the sample size. The solution described in Section 4.1 shows that the effective number is in fact smaller when the decision rule over this choice takes into account optimality considerations over the element of Υ to be chosen at each date. When the agent is considering distorting a past date she will choose low precision of the signal if that period's innovation is good news for her investment and high precision if it is bad news. Moreover, as discussed later, when the true $\sigma_V = 0$ it can be shown that the agent finds it optimally to distort only the last n periods and have the rest of the sequence consist of elements equal to σ_V.

2.4 Equilibrium concept

I consider an equilibrium concept analogous to a fully revealing rational expectations equilibrium, in which the price reveals all the information available to agents. Let $\{r^t\}$ denote the history of observed interest rate differentials up to time t, $\{r_s\}_{s=0,...t}$. Denote by $\sigma_V^*(r^t)$ the optimal sequence σ_V^t of $\{\sigma_{V,s}, s \leq t : \sigma_{V,s} \in \Upsilon\}$ chosen by the agent at time t based on data $\{r^t\}$ to reflect her belief in an alternative time-varying model. Let $f\ (r^{t+1})$ denote the time-invariant function that controls the conjecture about how next period's exchange rate responds to the history $\{r^{t+1}\}$

$$s_{t+1} = f(r^{t+1})$$

For a reminder, equation (2.7) is the optimization problem faced by the agent that involves both a maximizing choice over bonds and minimizing solution for the distorted model.

Definition 1. *An equilibrium will consist of a conjecture $f(r^{t+1})$, an exchange rate function $s(r^t)$, a bond demand function $b(r^t)$ and an optimal distorted sequence $\sigma_V^*(r^t)$ for $\{r^t\}$,*

$t = 0, 1, ...\infty$ such that agents at time t use the distorted model implied by the sequence of variances $\sigma_V^*(r^t)$ for the state-space defined in (2.5) to form a subjective probability distribution over $r^{t+1} = \{r^t, r_{t+1}\}$ and $f(r^{t+1})$ and satisfy the following equilibrium conditions:

1. Optimality: given $s(r^t)$ and $f(r^{t+1})$, the demand for bonds $b(r^t)$ and the distorted sequence $\sigma_V^*(r^t)$ are the optimal solution for the max min *problem in (2.7).*

2. Market clearing: given $b(r^t)$, $\sigma_V^*(r^t)$ and $f(r^{t+1})$, the exchange rate $s(r^t)$ satisfies the market clearing condition in (2.4).

3. Consistency of beliefs: $s(r^t) = f(r^t)$.

3 The Rational Expectations Model Solution

Before presenting the solution to the model, I first solve the rational expectations version which will serve as a contrast for the ambiguity aversion model. By definition, in the rational expectations case the subjective and the objective probability distributions coincide, i.e. $P = \widetilde{P}$. For ease of notation, I denote $E_t(X) \equiv E_t^P(X)$, where P is the true probability distribution. The DGP is given by the constant volatility state-space described in (2.1). The optimization problem is

$$V_t = \max_{b_t} E_t[(b_t q_{t+1} - \frac{c}{2} b_t^2)]$$

where the log excess return $q_{t+1} = s_{t+1} - s_t - r_t$ and b_t is the amount of foreign bonds demanded expressed in domestic currency. The FOC is

$$b_t = \frac{E_t(q_{t+1})}{c} \tag{3.1}$$

The market clearing condition states that $b_t = 0.5 s_t$. Combining the demand and the supply equation I get the equilibrium condition for the exchange rate:

$$s_t = \frac{E_t(s_{t+1} - r_t)}{1 + 0.5c} \tag{3.2}$$

I call (3.2) the UIP condition in the rational expectations version of the model. Driving c to zero implies the usual risk-neutral version $s_t = E_t(s_{t+1} - r_t)$.

It is easy to see that the solution in (3.1) can accommodate the risk aversion case, where:

$$V_t = \max_{b_t} b_t E_t q_{t+1} - \gamma b_t^2 Var_t q_{t+1}$$

Replacing c with $\gamma Var_t q_{t+1}$ in (3.1) delivers the usual solution for a mean variance utility.

To solve the model, I take the usual approach of a guess and verify method in which the agents are endowed with a guess about the law of motion of the exchange rate. To form expectations agents use the Kalman Filter which, given the Gaussian and linear setup is the optimal filter for the state-space in (2.1).

Let $\widehat{x}_{m,n} \equiv E(x_m|I_n)$ and $\Sigma_{m,n} \equiv E[(x_m - E(x_m|I_n))(x_m - E(x_m|I_n)')]$ denote the estimate and the mean square error of the hidden state for time m given information at time n. As shown in Hamilton (1994), the estimates are updated according to the following recursion:

$$\begin{aligned} \widehat{x}_{t,t} &= \rho\widehat{x}_{t-1,t-1} + K_t(y_t - \rho\widehat{x}_{t-1,t-1}) && (3.3)\\ K_t &= (\rho^2\Sigma_{t-1,t-1} + \sigma_U^2)[\rho^2\Sigma_{t-1,t-1} + \sigma_U^2 + \sigma_V^2]^{-1} && (3.4)\\ \Sigma_{t,t} &= (1 - K_t)(\rho^2\Sigma_{t-1,t-1} + \sigma_U^2) && (3.5) \end{aligned}$$

where K_t is the Kalman gain.

Based on these estimates let the guess about the exchange rate be

$$s_t = a_1\widehat{x}_{t,t} + a_2 r_t \tag{3.6}$$

For simplicity, I assume convergence on the Kalman gain and the variance matrix $\Sigma_{t,t}$. Thus, I have $\Sigma_{t,t} \equiv \Sigma$ and $K_t^{RE} = K$ for all t. Then, since $E_t r_{t+1} = \rho\widehat{x}_{t,t}$ the solution is

$$\begin{aligned} a_2 &= -\frac{1}{1 + 0.5c} && (3.7)\\ a_1 &= -\frac{1}{1 + 0.5c}\frac{\rho}{1 + 0.5c - \rho} && (3.8) \end{aligned}$$

For the case in which $c = 0$, $a_1 = -\frac{\rho}{1-\rho}$, $a_2 = -1$. Notice that when ρ is close to one it implies that a_1 is very large. This is consistent with the "asset" view of the exchange rate. The exchange rate s_t is the negative of the present discounted sum of the interest rate differential. Since the interest rate differential is highly persistent, a_1 will be a large negative number. It shows that s_t reacts strongly to the estimate of the hidden state $\widehat{x}_{t,t}$ because this estimate is the best forecast for future interest rates.

The UIP regression is

$$s_{t+1} - s_t = \beta r_t + \varepsilon_{t+1}$$

In this rational expectations model the dependent variable is

$$s_{t+1} - s_t = a_1(\rho - 1)\widehat{x}_{t,t} + a_1 K(r_{t+1} - \rho\widehat{x}_{t,t}) + a_2(\rho\widehat{x}_{t,t} + \xi_t + \sigma_U u_{t+1} + \sigma_V v_{t+1} - r_t) \tag{3.9}$$

where $\xi_t = x_t - \widehat{x}_{t,t}$ with $\xi_t \sim N(0, \Sigma)$ and independent of time t information. Then taking expectations of (3.9) and using that $E(\varepsilon_{t+1}|I_t) = 0$, I get

$$E_t(s_{t+1}) - s_t = -\widehat{x}_{t,t}\frac{.5\rho c}{(1+.5c)((1+.5c-\rho)} + \frac{1}{1+.5c}r_t$$

Since $cov(\widehat{x}_{t,t}, r_t) = Kvar(r_t)$, the UIP coefficient is

$$\widehat{\beta} = -\frac{.5\rho cK}{(1+.5c)(1+.5c-\rho)} + \frac{1}{1+.5c} \tag{3.10}$$

In the case of $c = 0$, $\widehat{\beta} = 1$ as the UIP predicts. When $c = \gamma Varq_{t+1}$, $\widehat{\beta} < 1$ due to the existence of a rational expectations risk premium in that model.

4 The Ambiguity Aversion Model Solution

In presenting the solution I use the constraints on the sequence $\sigma_V(r^t)$ described in Section 2.3, which follow from the requirement that the distorted sequence is statistically plausible. There I argue that this implies that the agent is statistically forced to be concerned only about a constant number n of dates that differ from σ_V. It is important to emphasize that I impose the restriction on the distorted sequence to differ from the reference model only for few dates purely for reasons related to statistical plausibility. The same intuition applies if the agent is not constrained by this consideration. I will return to this point, and I also note that under some conditions whether n is constant or equal to the sample size t is irrelevant.

Note that for a given deterministic sequence $\sigma_V^*(r^t) = \{\sigma_{V,s}, s = 0, ...t\}$ selected in (2.7) the usual recursive Kalman Filter applies. Thus, *after* this sequence has been optimally chosen by the agent at date t, the recursive filter uses the data from 0 to t to form estimates of the hidden state and their MSE. As shown in Hamilton (1992) the estimates are updated according to the recursion in (3.3), (3.5). The difference with the constant volatility case is that the Kalman gain now incorporates the time-varying volatilities $\sigma_{V,t}^2$:

$$K_t = (\rho^2\Sigma_{t-1,t-1} + \sigma_U^2)[\rho^2\Sigma_{t-1,t-1} + \sigma_U^2 + \sigma_{V,t}^2]^{-1} \tag{4.1}$$

The above notation is not fully satisfactory because it does not keep track of the dependence of the solution $\sigma_V^*(r^t)$ on the time t that is obtained. To correct this I make use of the following notation: $\sigma_{V,(t),s}$ is the value for the standard deviation of the observation shock that was believed at time t to happen at time s. The subscript t in parentheses refers to the

period in which the minimization takes place and the subscript $s \leq t$ refers to the period in the observed sample $0, .., t$ at which the draw for the standard deviation was believed to be equal to $\sigma_{V,(t),s}$.

Such a notation is necessary to underline that the belief is an action taken at date t and thus a function of date t information. There is the possibility that the belief about the realization of the variance at date s is different at dates $t-1$ and t. This can be interpreted as an update, although not Bayesian in nature.

To keep track of this notation I denote by: $I_i^t = \{r_s, H, F, \sigma_U, \sigma_{V,(t),s}, s = 0, ..., i\}$ for $i \leq t$ the information set that the filtering problem has at time i by treating the sequence $\{\sigma_{V,(t),s}\}_{s=0,...,i}$ as known. Note that this sequence is optimally selected at date t. This notation highlights that the filtering problem is backward-looking based on a deterministic sequence $\{\sigma_{V,(t),s}\}_{s=0,...,i}$. Then for $i \leq t$:

$$\begin{aligned} \widehat{x}_{i,i}^t &= E(x_i|I_i^t) = \widehat{x}_{i-1,i-1}^t + K_i^t(r_i - \widehat{x}_{i-1,i-1}^t) \\ \Sigma_{i,i}^t &= E[(x_i^t - \widehat{x}_{i,i}^t)^2|I_i^t] \qquad (4.2) \\ K_i^t &= (\rho^2\Sigma_{i-1,i-1}^t + \sigma_U^2)[\rho^2\Sigma_{i-1,i-1}^t + \sigma_U^2 + \sigma_{V,(t),i-1}^2]^{-1} \qquad (4.3) \end{aligned}$$

Thus $\widehat{x}_{i,i}^t$ is the estimate of the time i hidden state based on the sample $\{r_s\}_{s=0,...,i}$ and K_i^t is the time i Kalman gain by treating the sequence $\{\sigma_{V,(t),s}\}_{s=0,...,i}$ as known.

4.1 The optimal distorted expectations

In order to solve the max min problem in (2.7) I endow the agent with a guess about the relationship between the future exchange rate and the estimates for the exogenous process. As in the definition of equilibrium, let that guess be the function $f(.)$:

$$s(r^{t+1}) = s_{t+1} = f(r^{t+1}) = f(r^t, r_{t+1}) \qquad (4.4)$$

where the agent observes r_t and uses the distorted law of motion for the interest rate differential as in (2.5) to back out the hidden state x_t that controls the future evolution of r_{t+1}.

For the minimization in (2.7) the agent needs to understand how her expected utility depends on the $\sigma_V(r^t)$ out of the possible set given by (2.8). Because the minimizing choice is over sequences formed from the given exogenous set Υ the optimal solution will be a corner solution. To obtain the minimizing sequence in (2.7) the agent has to forecast only the monotonicity direction, and not the exact derivatives, in which different sequences $\sigma_V(r^t)$

influence the only unknown at time t, the expected exchange rate s_{t+1}. To forecast this direction the agents uses the guess in (4.4).

A particular guess in (4.4) is linear in the estimate of the hidden state and in the observed differential:

$$s_{t+1} = a_1 \widehat{x}^{t+1}_{t+1,t+1} + a_2 r_{t+1} \tag{4.5}$$

I will return to investigating the equilibrium properties of this guess, but for the agent's decision it only matters the monotonicity of the average value of s_{t+1} with respect to the average value of r_{t+1}. Suppose the equilibrium guess satisfies this property so we have:

Conjecture 1. *In equilibrium* $\frac{E^{\tilde{P}}_t s_{t+1}}{E^{\tilde{P}}_t r_{t+1}} < 0$.

The guess in (4.5) includes

$$\widehat{x}^{t+1}_{t+1,t+1} = (1 - K^{t+1}_{t+1})\rho \widehat{x}^{t+1}_{t,t} + K^{t+1}_{t+1} r_{t+1}$$

where $K^{t+1}_{t+1}, \widehat{x}^{t+1}_{t,t}$ are the Kalman filter objects described above. Since $K^{t+1}_{t+1} \geq 0$ then

$$\frac{\partial s_{t+1}}{\partial r_{t+1}} = a_1 K^{t+1}_{t+1} + a_2$$

For the guess in (4.5) the same intuition about the parameters a_1 and a_1 holds as in the RE case. A positive realization for r_{t+1} will translate into an appreciation of the domestic currency because the domestic interest rate is higher than the foreign one. Similarly, a higher positive estimate of the hidden state means a higher present value of investing in the domestic currency which in equilibrium will lead to an appreciation of the domestic currency. This intuition highlights that in equilibrium $a_1 K^{t+1}_{t+1} + a_2$ will be a negative number as Conjecture 1 supposes.

The expected interest rate differential is given by the hidden state estimate

$$E^{\tilde{P}}_t (r_{t+1}) = \rho \widehat{x}^t_{t,t} = \rho \widehat{x}^t_{t-1,t-1} + K^t_t (r_t - \rho \widehat{x}^t_{t-1,t-1})$$

This means that $E^{\tilde{P}}_t (r_{t+1})$ is increasing in the innovation $r_t - H' F \widehat{x}^t_{t-1,t-1}$. In turn, the estimate $\widehat{x}^t_{t,t}$ is updated by incorporating this innovation using the gain K^t_t. As clear from (4.1) the Kalman gain is decreasing in the variance of the temporary shock $\sigma^2_{V,(t),t}$. Intuitively, a larger variance of the temporary shock implies less information for updating the estimate of the hidden persistent state. Combining these two monotonicity results, the estimate $\widehat{x}^t_{t,t}$ is increasing in the gain if the innovation is positive. On the other hand, if $(r_t - F \widehat{x}^t_{t-1,t-1}) < 0$ then $\widehat{x}^t_{t,t}$ is decreased by having a larger gain K^t_t.

By construction, expected excess return $E_t^{\widetilde{P}} W_{t+1}$ is monotonic in $E_t^{\widetilde{P}}(s_{t+1})$ since $W_{t+1} = b_t[s_{t+1} - s_t - r_t]$. The sign is given by the position taken in foreign bonds b_t. If the agent decides in equilibrium to invest in domestic bonds and take advantage of a higher domestic rate by borrowing from abroad, i.e. $b_t < 0$, then a higher value for $E_t^{\widetilde{P}}(s_{t+1})$ will hurt her.

Proposition 1. *Expected excess return, $E_t^{\widetilde{P}} W_{t+1}$ is monotonic in $\sigma^2_{V,(t),t}$. The monotonicity is given by the sign of $[b_t(r_t - H'F\widehat{x}^t_{t-1,t-1})]$.*

Proof. By using Conjecture 1, and combining the signs of the partial derivatives involved in $\frac{\partial E_t^{\widetilde{P}} W_{t+1}}{\partial \sigma^2_{V,(t),t}}$. For details, see Appendix B.

The impact of $\sigma^2_{V,(t),t}$ on the expected excess payoffs is given by the following intuitive mechanism. Suppose the agent invests in domestic bonds (i.e. $b_t < 0$). She is then worried about a higher future depreciation of the domestic currency (i.e. that $E_t^{\widetilde{P}}(s_{t+1})$ is higher). A higher future depreciation in equilibrium occurs when the future interest rate differential is lower. A lower average future differential is generated if the current hidden state of the differential is lower. The current hidden state is not observable so it needs to be estimated. Thus the agent is concerned that the estimate is lower, i.e that $\widehat{x}^t_{t,t}$ is lower, but still positive to justify the starting assumption that she invests in the domestic bonds. The variance $\sigma^2_{V,(t),t}$ negatively affects the gain K_t^t which controls the weight put on current innovations to update the estimate of the hidden state. To reflect a concern for a lower estimate $\widehat{x}^t_{t,t}$ the agent chooses to act as if the variance $\sigma^2_{V,(t),t}$ is larger (low gain) when the innovation $(r_t - F\widehat{x}^t_{t-1,t-1})$ is positive and as if $\sigma^2_{V,(t),t}$ is smaller (high gain) if the innovation is negative.

In (2.7) the minimization of $E_t^{\widetilde{P}} W_{t+1}$ is over the sequence $\sigma^*_V(r^t)$. Proposition 1 refers to the monotonicity with respect of a time s element in the sequence, taking as given the time $0, .., s-1$ elements of that sequence. The implication of Proposition 1 is that the decision rule for choosing a distorted $\sigma_{V,(t),s}$ is:

$$\begin{aligned} \sigma_{V,(t),s} &= \sigma_V^H \text{ if } b_t(r_s - F\widehat{x}^t_{s-1,s-1}) < 0 \\ \sigma_{V,(t),s} &= \sigma_V^L \text{ if } b_t(r_s - F\widehat{x}^t_{s-1,s-1}) > 0 \end{aligned} \tag{4.6}$$

One way to interpret this decision is that agents react asymmetrically to news. If the agent decides to invest in the domestic currency, then increases (decreases) in the domestic differential are good (bad) news and from the perspective of the agent that wants to take advantage of such higher rates. An ambiguity averse agent facing information of ambiguous quality will then tend to underweigh good news by treating them as reflecting temporary shocks and overweigh the bad news by fearing that they reflect the persistent shocks.

In Section 2.3, I introduced the restriction that the agent only considers n dates to be different from the reference model. The agent's ultimate concern is that the estimate of the

hidden state is low but positive when $b_t < 0$ and low in absolute value but negative when $b_t > 0$. For example when $b_t < 0$ many such sequences need to be compared to produce the minimum $\widehat{x}^t_{t,t}$. The minimization problem is exactly choosing the sequence out of the feasible ones that produces such a minimum. Out of these possible sequences, (4.6) shows on what type of sequences the agent restricts attention because they negatively affect her utility.

When the true DGP is characterized by no temporary shocks, then the minimization over the feasible sequences is more straightforward.

Remark 1. *Suppose $\sigma_V = 0$. Then the minimization over the sequence defined in (2.8) reduces to minimization over the sequence*

$$\sigma^t_V = \{\sigma_{V,s}, for\ s = t-n+1, ...t, \sigma_{V,s} \in \Upsilon\ and\ \sigma_{V,l} = 0, l < t-n+1\}$$

Proof. The estimate of the hidden state $\widehat{x}^t_{t,t}$ is a weighted average of all previously observed differentials $r_s, s \leq t$, with weights that are a function of the time-varying standard deviation $\sigma_{V,s} \in \Upsilon$. If at any point s with $\sigma_{V,s} = 0$, then all previous differentials r_{s-j}, $j = 1, ..s$, have weights equal to zero for $\widehat{x}^t_{t,t}$. See Appendix B for details.

Remark 1 implies that if $\sigma_V = 0$ the sequences the agent compares are the ones that have elements different from the constant sequences of zeros only in the last n periods. Remark 1 simplifies the problem of finding the optimal sequence to minimize the estimate $\widehat{x}^t_{t,t}$ by only analyzing the last n observed differentials. For those differentials the decision rule in (4.6) gives the direction of the minimizing element in the sequence with $t-n+1 \leq s \leq t$.

In the definition of the equilibrium I denoted the solution to this minimization problem by $\sigma^*_V(r^t)$. Note that this decision rule is taking b_t as given. The next section discusses the optimal condition for b_t.

4.2 The optimal bond position

The bond position b_t has two important features: the magnitude and the sign. In typical exercises with ambiguous signals, such as Epstein and Schneider (2007), Epstein and Schneider (2008), Illeditsch (2009) and reviewed in Epstein and Schneider (2010), the sign is known as the agents holds in equilibrium one particular asset of interest. Here, however, agents are switching positions of where to invest by doing the carry trade, i.e. b_t switches between positive and negative according to the equilibrium conditions. This is easy to see from the market clearing solution where $b_t = 0.5 s_t$ with s_t fluctuating endogenously between an appreciated and depreciated level compared to steady state, which is zero.

The max min problem in (2.7) is:

$$V_t = \max_{b_t} \min_{\sigma_V^*(r^t) \in \sigma_V^t} E_t^{\widetilde{P}}[b_t(s_{t+1} - s_t - r_t) - \frac{c}{2} b_t^2)] \tag{4.7}$$

A property of this maxmin type of optimization is that it can have kinked instead of interior solutions. The possible sequences σ_V^t imply different probability distributions to evaluate the expected excess return $q_{t+1} = (s_{t+1} - s_t - r_t)$. Let J denote the number of distinct expected values of the excess return q_{t+1} implied by the feasible sequences σ_V^t. Let $\{\mu_j\}_{j=1,...J}$ denote these distinct expected values. The ambiguity aversion problem can then be interpreted as the agent being concern about which model j is the true model. Then the problem in can be restated as:

$$V_t = \max_{b_t} \min_{\mu_j} (b_t \mu_j - \frac{c}{2} b_t^2)$$

Letting $g_j(b_t) = (b_t \mu_j - \frac{c}{2} b_t^2)$ and $U(b_t) = \min_{\mu_j} g_j(b_t)$ the problem becomes:

$$V_t = \max_{b_t} U(b_t)$$

To solve for the optimal b_t, I first investigate the kink solution. To do this, take the functions $g_j(b_t)$ and look at their intersection points. In this case the intersection points are $b_t = 0$ which is the possible kink solution. To analyze whether that is the global solution I compute the gradients of these functions at $b_t = 0$:

$$\frac{\partial g_j(b_t)}{\partial b_t}_{|b_t=0} = \mu_j$$

If the product of $\{\mu_j\}_{j=1,...J}$ is negative, then the global solution for b_t is $b_t = 0$. In that case there is no unique solution for the minimization problem in $\min_{\mu_j} g_j(b_t)$ since $g_j(b_t) = 0$. If it is positive, i.e. all μ_j have the same sign, then the solution allows the interchangeability of the max and the min operator:

$$V_t = \min_{\mu_j} \max_{b_t} g_j(b_t)$$

For each model j we have the solution: $b_t = \frac{\mu_j}{c}$. So the minimization problem becomes:

$$V_t = \min_{\mu_j} \frac{\mu_j^2}{2c} \tag{4.8}$$

$$\mu_j^* = \arg\min_{\mu_j} \frac{\mu_j^2}{2c} = \min_{\mu_j}(|\mu_j|) \tag{4.9}$$

Equation (4.8) shows that the solution is then to choose $\mu_j^* = \min_{\mu_j}(|\mu_j|)$ and thus $b_t^* = \frac{\mu_j^*}{c}$. I call this the interior solution.

Notice that the kinked solution means that in this case there is no participation in the market, contrary to the rational expectations model where there will always be participation, a result well known in the ambiguity aversion literature.

4.3 Equilibrium

The equilibrium concept was defined in section 2.4. One element of the equilibrium is the optimal bond position which, as shown in section 4.2, is a function of the expected payoffs $\{\mu_j\}_{j=1,...J}$ implied by the sequences of variances σ_V^t. The optimal position can be the kinked solution $b_t = 0$, depending on the sign of the possible expected payoffs, which in turn are endogenous objects. Thus, to compute the equilibrium I take the following approach.

Suppose the economy at time t is characterized by all the implied possible expected payoffs being positive, i.e. $0 < \mu_1 < ... < \mu_J$ with μ_1 denoting the minimum payoff. In this case we know from (4.9) that μ_1 should be the equilibrium payoff and that b_t is the equilibrium interior solution

$$b_t^* = \frac{\mu_1}{c} > 0$$

Using the sign of b_t^* the optimal sequence $\sigma_V^*(r^t)$ can be computed along the lines of section 4.1 and decision rule (4.6). This generates $\widehat{x}_{t,t}^t = E_t^{\widetilde{P}}(r_{t+1})$. Using the conjectured law of motion $s_{t+1} = f(r^t, r_{t+1})$ the expected $E_t^{\widetilde{P}}(s_{t+1})$ is formed. Given that b_t^* is interior and using the market clearing condition $b_t = 0.5 s_t$ results in

$$s_t = \frac{E_t^{\widetilde{P}}(s_{t+1}) - r_t}{1 + 0.5c} \tag{4.10}$$

The objects $b_t^*, \sigma_V^*(r^t)$, s_t are an equilibrium if $sign(s_t) = sign(b_t)$. If not, the conjectured starting point is not correct and there are at least some sequences of variances that imply negative expected payoffs.

In that case, suppose the the economy at time t is characterized by all the implied possible expected payoffs being negative, i.e. $0 > \mu_J > ... > \mu_1$ with μ_1 now denoting the minimum payoff in absolute value. In this case we know from (4.9) that μ_1 should be the equilibrium payoff and again that

$$b_t^* = \frac{\mu_1}{c} < 0$$

A similar equilibrium logic as above suggests using the sign of b_t^* to compute the minimizing sequence $\sigma_V^*(r^t)$. Then use the market clearing condition and the optimality of b_t^* as an

interior solution to compute s_t. The objects $b_t^*, \sigma_V^*(r^t)$, s_t are an equilibrium if again $sign(s_t) = sign(b_t)$. If not, then this second conjectured starting point for finding the equilibrium is also not correct.

In that case it must be that there are both positive and negative expected payoffs implied by the possible sequences σ_V^t. Then we have from 4.2 and the market clearing solution that the equilibrium solution is no participation, $b_t^* = 0$, and the distorted sequence $\sigma_V^*(r^t)$ is not uniquely determined, as any of the possible sequences σ_V^t are an equilibrium $\sigma_V^*(r^t)$.

The last element defining the equilibrium is whether, based on the conjectured law of motion $f(r^{t+1})$, the equilibrium exchange rate is consistent with $s_t = f(r^t)$. In order to address this consistency I impose the following restriction on the agent's conditional probability model, where n is the constant number of dates at which the agent entertains that the alternative sequence of variances can differ from the reference sequence:

Assumption 1: $P(\sigma_{V,(t+1),t+1} = \sigma_V^H | I_t) + P(\sigma_{V,(t+1),t+1} = \sigma_V^L | I_t) = \frac{n}{t}$.

Assumption 1 has the implication that $\lim_{t\to\infty} E^{\tilde{P}}(\sigma_{V,(t+1),t+1}|I_t) = \sigma_V$.

This assumption states that when the agent at time t is using her conjecture about how exchange rate at time $t+1$ responds to the observables at time $t+1$, the expectation about what agents at time $t+1$ will perceive as the time $t+1$ standard deviation of the temporary shock is consistent with her view about the historical data. The expected value $E^{\tilde{P}}[\sigma_{V,(t+1),t+1}|I_t]$ equals:

$$\sigma_V^H P(\sigma_{V,(t+1),t+1} = \sigma_V^H | I_t) + \sigma_V^L P(\sigma_{V,(t+1),t+1} = \sigma_V^L | I_t) + \sigma_V P(\sigma_{V,(t+1),t+1} = \sigma_V | I_t)$$

By the construction of n, the agent at time t believes that $\sigma_{V,s} \neq \sigma_V$ for $\{s = 0, ..., t\}$ only for n out of t times in the sample of observable data she has. Thus, in forming $E_t^{\tilde{P}}\sigma_{V,t+1}$,the agent uses the same frequency approach to believe that Prob$(\sigma_{V,(t+1),t+1} = \sigma_V^H|I_t)$+Prob$(\sigma_{V,(t+1),t+1} = \sigma_V^L|I_t) = n/t$.

The fact that for t large the expected variance is equal to the constant reference σ_V has important implications for the conjectured law of motion. Take the conjectured law of motion as in (4.5):

$$s_{t+1} = a_1 \widehat{x}_{t+1,t+1}^{t+1} + a_2 r_{t+1}.$$

Then, we have $E_t^{\tilde{P}} s_{t+1} = E_t^{\tilde{P}}[(1-K)a_1\rho\widehat{x}_{t,t}^{t+1} + (a_1K + a_2)r_{t+1}]$, where K is the Kalman gain associated with the sequence of constant variances σ_V. Under the time t probability distribution: $E_t^{\tilde{P}} r_{t+1} = \rho\widehat{x}_{t,t}^t$ so

$$E_t^{\tilde{P}} s_{t+1} = (1-K)a_1\rho E_t^{\tilde{P}}(\widehat{x}_{t,t}^{t+1}) + (a_1K + a_2)\rho\widehat{x}_{t,t}^t. \tag{4.11}$$

The time $t+1$ estimate of the time t hidden state, $\widehat{x}_{t,t}^{t+1}$, is not in general equal to $\widehat{x}_{t,t}^{t}$ because the former is based on a sequence of variances $\sigma_{V,(t+1),s}$ for $s = 1, ...t$ which can differ from $\sigma_{V,(t),s}$, i.e. the worst-case scenario for the time $t+1$ agent can differ from the worst-case scenario of the time t agent. However, under the special case $\sigma_V = 0$, this difference is irrelevant since $K = 1$. This special case is helpful in getting analytical results about the consistent law of motion.[18] Indeed, in a spirit close to that of Remark 1, I get:

Remark 2. *If $\sigma_V = 0$, then under the conjectured (4.5), $E_t^{\widetilde{P}} s_{t+1} = (a_1 + a_2)\rho\widehat{x}_{t,t}^{t}$.*

Proof. Since $\sigma_V = 0$, then $K = 1$ in (4.11).

Proposition 2. *Under Assumption 1 and if $\sigma_V = 0$ the equilibrium law of motion for the exchange rate is:*

$$s_t = a_1\widehat{x}_{t,t}^{t} + a_2 r_t \tag{4.12}$$

where a_1, a_2 are the same coefficients as in equations (3.7) and (3.8) respectively, that characterize the rational expectations case.

Proof. Use Remark 2, the FOC (4.10) and the same guess-and-verify method as in the rational expectations case.

The key limitation of Assumption 1 is that it imposes a model for forming expectations about time s_{t+1} in which the time $t+1$ signal, r_{t+1}, is in the limit not ambiguous. Thus, different than Epstein and Schneider (2008) for example, there is no ambiguity premium coming from the expectation of the arrival of future ambiguous news. Such an ambiguity premium would be generated by the expectation of the asymmetric response to news in the next period.

Assumption 1 can be justified on several grounds. First, according to this assumption $E^{\widetilde{P}}[\sigma_{V,(t+1),t+1}|I_t]$ is consistent with $E^{\widetilde{P}}[\sigma_{V,t+1}|I_t]$, which is the expected realized standard deviation for time $t+1$. A second reason is related to the ability of solving the model. If there is an ambiguity premium generated by the expected arrival of ambiguous news that premium is reflected in the exchange rate at time t and the conjectured law of motion in (4.5) would have to take it into account. But since that premium is not discounted, i.e. there is no discounting of the expected premium in the UIP equation $s_t = E_t^{\widetilde{P}} s_{t+1} - r_t$, then the model would have no solution in which that premium would be different from zero. Remark 3 addresses this point and Appendix B.1 presents some details on this argument.

[18] If $\sigma_V > 0$ then $K < 1$ and finding the coefficients a_1 and a_2 requires numerical solutions to minimize the distance between the perceived law of motion (4.5) and the actual law of motion that take into account the distorted expectations and Assumption 1. There, Conjecture 1 is verified by checking the signs of the two coefficients a_1 and a_2. Such a numerical procedure is detailed in Appendix B.2. Proposition 2 also verifies Conjecture 1 through the implied signs of the analytical a_1 and a_2.

Remark 3. *If the conjectured law of motion in (4.5) is enriched to* $s_{t+1} = a_1 \widehat{x}^{t+1}_{t+1,t+1} + a_2 r_{t+1} + \delta$ *and* $E^{\widetilde{P}}_t a_1 \widehat{x}^{t+1}_{t+1,t+1} = a_1 \rho \widehat{x}^t_{t,t} + \delta$ *then there is no* $\delta \neq 0$ *that implies consistency of beliefs.*

Proof: Using the UIP equation $s_t = E^{\widetilde{P}}_t s_{t+1} - r_t$ and the proposed conjecture implies that $s_t = a_1 E^{\widetilde{P}}_t \widehat{x}^{t+1}_{t+1,t+1} + a_2 E^{\widetilde{P}}_t r_{t+1} + \delta - r_t = (a_1 + a_2)\rho \widehat{x}^t_{t,t} - r_t + 2\delta$. For consistency to hold it must be that $a_2 = -1$, $a_1 = -\frac{\rho}{1-\rho}$ and $\delta = 0$.

A third argument to defend Assumption 1 is related to the fact that the ex-post positive payoffs at $t+1$ implied by the model can be entirely attributed by the agent as a compensation for the uncertainty faced at time t. To analyze this argument, let $E^P_t s_{t+1}$ denote the average equilibrium ex-post s_{t+1} and $E^{\widetilde{P}}_t(s_{t+1})$ denote the expected s_{t+1} implied by the distorted beliefs $\widetilde{P}$, the conjectured law of motion in (4.5) and Assumption 1. Let $\overline{P}$ denote an alternative distribution implicitly defined by $\overline{\sigma}_V$ and the other known elements of the law of motion for r_{t+1} implied by (2.5) such that the set Υ now contains four elements $\{\sigma^L_V < \sigma_V < \overline{\sigma}_V < \sigma^H_V\}$.[19] Then the following can be stated:

Remark 4. *If* $[E^P_t s_{t+1} - E^{\widetilde{P}}_t(s_{t+1})] sign(b_t) \equiv d_t > 0$, *then there exists a* $\overline{\sigma}_V$ *such that* d_t *can be generated by the true DGP being* $\overline{P}$ *and* s_{t+1} *responding symmetrically to news as implied by Assumption 1.*

Proof: See Appendix B.1.

The intuition behind Remark 3 is that when the agent investing at time t observes on average ex-post positive payoffs at time $t+1$, these payoffs are naturally interpreted as a compensation for the uncertainty about $\sigma_V(r^t)$ faced at time t. The true DGP is characterized by the constant σ_V, however, by construction, there is incomplete knowledge about this feature of the economy. Take for example $b_t < 0$. After observing $E^P_t s_{t+1} < E^{\widetilde{P}}_t(s_{t+1})$ and using the conjecture that s_{t+1} is controlled by the law of motion in (4.5) and Assumption 1 the agent would conclude that s_{t+1} has depreciated less ex-post because her expected $r_{t+1}, E^{\widetilde{P}}_t r_{t+1}$, was lower than the possible true DGP $\overline{P}$, i.e. $E^{\widetilde{P}}_t r_{t+1} < E^{\overline{P}}_t r_{t+1}$. This is consistent with the fact that the agent at time t was acting upon a sequence of variances with draws from $\{\sigma^L_V < \sigma_V < \overline{\sigma}_V < \sigma^H_V\}$ that was minimizing expected payoffs. Appendix B.1 details this observation.

4.4 Risk aversion and ambiguous signals

The model so far has been derived under a setup in which the agent is ambiguity averse but otherwise risk-neutral. Introducing risk-aversion complicates the analysis significantly. The

[19]Note that $\widetilde{P}$ remains identically as when $\Upsilon = \{\sigma^L_V < \sigma_V < \sigma^H_V\}$ because the corner solutions are the same.

reason is that with risk-aversion the minimizing choice over the variance of temporary shocks will be influenced by two effects. The first effect analyzed so far is the effect through the expected payoffs: the variance affects the Kalman gain, which in turn affects the estimate of the hidden state and this estimate controls the expected payoffs. However, with risk aversion, the variance has an added direct effect on utility. To see the two channels, consider a mean variance utility:

$$V_t = \max_{b_t} \min_{\sigma_V^*(r^t) \in \sigma_V^t} E_t^{\widetilde{P}}[b_t(s_{t+1} - s_t - r_t)] - \frac{1}{2} b_t^2 Var_t^{\widetilde{P}} s_{t+1}$$

where the minimization is over the same sequence of variances as in (2.8). Suppose the equilibrium is characterized by a similar law of motion as the guess in (4.5) with the coefficients a_1, a_2 potentially different in this case. Then:

$$Var_t^{\widetilde{P}} s_{t+1} = (a_1 K + a_2)^2 Var_t^{\widetilde{P}} r_{t+1}.$$

By the Kalman filtering formulas, as in (3.4), (3.5), and Assumption 1, the conditional variance of r_{t+1} is:

$$Var_t^{\widetilde{P}} r_{t+1} = \rho^2 \Sigma_{t,t}^t + \sigma_U^2 \tag{4.13}$$

It is then easy to establish that:

Proposition 3. *The variance of excess payoff, $b_t^2 Var_t^{\widetilde{P}} s_{t+1}$, is increasing in $\sigma^2_{V,(t),t}$.*

Proof. The variance $b_t^2 Var_t^{\widetilde{P}} s_{t+1}$ is increasing in $Var_t^{\widetilde{P}} r_{t+1}$. By (4.13) the latter is increasing in $\sigma^2_{V,(t),t}$ through the effect on $\Sigma_{t,t}^t$. For details see Appendix B.

Intuitively, a larger variance of the temporary shocks translates directly into a higher variance of the estimates $\Sigma_{t,t}^t$. By choosing higher values of σ_V in the sequence $\sigma_V(r^t)$, she will increase the expected variance of the differential $Var_t^{\widetilde{P}} r_{t+1}$ because $\frac{\partial \Sigma_{t,t}^t}{\partial \sigma^2_{V,(t),t}} > 0$.

The overall effect of $\sigma^2_{V,(t),t}$ on the utility V_t is then coming through two channels. One is the positive relationship between $\sigma^2_{V,(t),t}$ and the variance of the payoffs as in Proposition 3. The other effect is through expected payoffs and is given by Proposition 1 The total partial derivative is then

$$\frac{\partial V_t}{\partial \sigma^2_{V,(t),t}} = \frac{\partial V_t}{\partial E_t^{\widetilde{P}} s_{t+1}} \frac{\partial E_t^{\widetilde{P}} s_{t+1}}{\partial E_t^{\widetilde{P}} s_{rt+1}} \frac{\partial E_t^{\widetilde{P}} r_{t+1}}{\partial \sigma_{V,(t),t}} + \frac{\partial V_t}{\partial Var_t^{\widetilde{P}} r_{t+1}} \frac{\partial Var_t^{\widetilde{P}} r_{t+1}}{\partial \sigma_{V,(t),t}}$$

Using Propositions 1 and 3, the sign of this derivative is:

$$sign\left[\frac{\partial V_t}{\partial \sigma^2_{V,(t),t}}\right] = sign(b_t)sign(r_t - F\widehat{x}^t_{t-1,t-1}) - sign\left[\frac{\partial Var^{\widetilde{P}}_t r_{t+1}}{\partial \sigma^2_{V,(t),t}}\right]$$

From Proposition 3 the $sign(\partial Var^{\widetilde{P}}_t r_{t+1}/\partial \sigma^2_{V,(t),t})$ is positive. Thus, if the sign of $[b_t(r_t - F\widehat{x}^t_{t-1,t-1})]$ is negative, then the two effects align because a higher variance $\sigma^2_{V,(t),t}$ will imply lower expected excess payoffs. However, if the sign is positive, then the two directions are competing. Then it remains a quantitative question to determine which effect is stronger. To analyze this situation, I show in Appendix B.3 that in this setup the probability that the effect through the expected payoffs dominates the one through the variance is almost equal to one. I conclude that in this model the effect of $\sigma_V(r^t)$ on utility goes almost entirely through its effect on expected payoff.

Introducing risk aversion raises another important issue related to the assumed structure of uncertainty. In the risk-neutral case, whether uncertainty is assumed about the realizations of the variances of the temporary shock or the persistent shock is completely innocuous. The reason is that with risk-neutrality, by construction, the driving force in the agent's evaluation is the expected payoff. Expected payoffs are affected by the estimate for the hidden state which in turn depends on the time-varying signal to noise ratios. This implies that it is not the specific equation, either the observation or state equation, in which uncertainty is assumed that matters but the relative strength of the information contained in them. Such an irrelevance of the structure of uncertainty will qualitatively not hold with risk aversion. However, quantitatively, in a setup with risk aversion where expected payoffs drive most of the portfolio decision such issues tend to become mute.

This argument also highlights the fact that it is very important to include the expected return channel in any evaluation of the effect on utility of variances of unobserved shocks. If that effect is absent, the robust estimator features different qualitative properties. For example, there are models that deal with a robust estimator that have considered a setup with commitment to previous distortions in which the agent wants to minimize the estimation mean square error. Li and Tornell (2008) study such a problem and show that if the agent is concerned only about the uncertainty of the temporary shock, then she will act as if the variance of the temporary shock is higher. That generates a steady state robust Kalman gain that is lower than the one implied by the reference model. Such a concern for misspecification generates the type of underreaction to news that has been proposed in the literature as a mechanism to explain the UIP puzzle. However, as Li and Tornell (2008) point out, when uncertainty about the persistent shock is added the concern for misspecification leads to a

higher robust Kalman gain than in the reference model.[20] That implies an overreaction to news and a UIP regression coefficient that is higher than 1, thus moving away from explaining the puzzle.

4.5 Parameterization

In the benchmark case, the reference model is an AR(1) state space representation as in (2.1) where $\sigma_V = 0$. Remarks 1 and 2 show that when $\sigma_V = 0$ there are analytical solutions to the equilibrium of the ambiguity aversion version of the model, a feature that helps understand the mechanics of the model. A value of $\sigma_V = 0$ is also consistent with ML estimates of (2.1) reported in Table 10.

Table 1: Benchmark specification

σ_V	σ_V^H	σ_V^L	σ_U^{DGP}	ρ	c
0	0.00025	0	0.0005	0.98	0

These values imply that the steady state Kalman weight on the innovation used to update the estimate of the current state is 1, 1 and 0.17 for the true DGP, the low variance and the high variance case. Although the lower gain might seem very different than the true DGP, note that the model does not imply that these gains are used for every period. It is only for a few dates in a large sample that such distorted gains are employed. Also note that the cost of capital c is set to zero in the benchmark specification to have UIP hold exactly under the distorted expectations. When discussing the effect of risk aversion that value will be changed to reflect the extent to which risk aversion can affect the results in this model.[21] According to Proposition 2 the resulting equilibrium coefficients a_1 and a_2 can then be determined analytically and they are equal to $-\frac{\rho}{1-\rho} = -49$ and -1 respectively.

As discussed in Section 2.3, in order for the equilibrium distorted sequences of variances to be difficult to distinguish statistically from the reference sequence I restrict the elements in the alternative sequences considered by the agent to be different from the reference model only for a constant number n of dates. To quantify the statistical distance between the two

[20] In this case, as discussed in Basar and Bernhard (1995) and Hansen and Sargent (2008, Ch.17), the robust filter flattens the decomposition of variances across frequencies by accepting higher variances at higher frequencies in exchange for lower variances at lower frequencies.

[21] Note that when $c = 0$ utility is linear in the expected return so if equilibrium $b_t \neq 0$ then $E_t^{\tilde{P}} q_{t+1} = 0$, implying UIP holds ex-ante under the distorted equilibrium $\widetilde{P}$ and b_t is determined by the market clearing condition.

Table 2: Likelihood comparison: $L^{Dist}(r^t) - L^{DGP}(r^t)$

	$Mean$	$St.dev.$	$\%positive$
(1)	-1.4	1.23	0.05
	(0.06)	(0.03)	(0.01)
(2)	-204	2.8	0
	(3.2)	$-$	$-$

models, I use a comparison between the log-likelihood of a sample $\{r^t\}$ computed under the reference sequence of constant variances ($L^{DGP}(r^t)$) and under the equilibrium distorted sequence ($L^{Dist}(r^t)$).

Table 2 reports some statistics for the likelihood comparison $L^{Dist}(r^t) - L^{DGP}(r^t)$, computed for a sample of $T = 300$, for various cases. The table reports the mean and the standard deviation of this difference and the percent of times for which this difference is positive. The standard deviation of these statistics is then computed across $N = 1000$ simulations. Row (1) reports the results for the benchmark parameterization in which $n = 2$. It shows that the average difference between $L^{Dist}(r^t) - L^{DGP}(r^t)$ is around -1.4. Row (2) considers the situations in which there is no restriction on the number of periods for which the agent distorts the sequence so that for any period characterized by good news a low precision of signals is used and a high precision of signals for bad news. In this case $n = t$ and the difference $L^{DGP}(r^t) - L^{Dist}(r^t)$ is increasing with the sample size. For $T = 300$, the average difference across simulations is around 200 log points. Thus, such a distorted sequence would result in an extremely unlikely interpretation of the data. For this reason, I restrict n to be a small number.

5 Results

In this section I present the main implications that the distorted expectations model has for exchange rate puzzles. Figure 1 illustrates for one simulation of the model the exchange rate path under the distorted expectations version (the solid line path) and the rational expectations (RE) version (the - - path). There, for comparison, the same exogenous driving process is used but the estimates are computed under the different model solutions. When the agent is investing in the foreign bond, i.e. $b_t > 0$ and $s_t > 0$, she is concerned that the negative estimate of the hidden state is in fact less negative than what the reference model would imply. Her investment position will reflect her pessimistic assessment of the future distribution and she will invest less compared to the RE model. Thus, $s_t < s_t^{RE}$ when $s_t > 0$

and $s_t > s_t^{RE}$ when $s_t < 0$ where s_t^{RE} denotes the exchange rate under RE.

Figure 1: Model generated exchange rate path

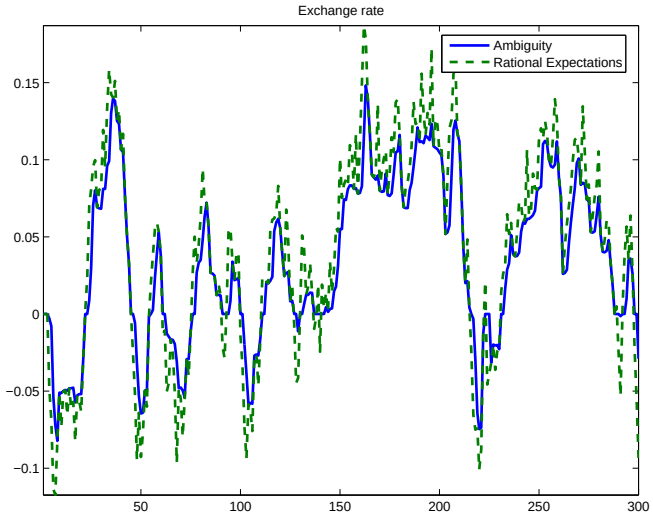


Table 3: Correlations exchange rates

(1)	(2)	(3)	(4)
0.97	0.52	0.18	0.77
(0.01)	(0.04)	(0.05)	(0.02)

Figure 1 shows that the exchange rate under ambiguity reacts asymmetrically to news. When for example $b_t < 0$ the domestic currency appreciates more gradually while depreciating suddenly compared to its RE version. This is consistent with the description that for the high interest rate (investment) currency "exchange rates go up by the stairs and down by the elevator" (see Brunnermeier et al. (2008)). To investigate this asymmetric response further, Table 3 computes correlations between the exchange rate under ambiguity and under RE. Columns (1) and (2) indicate that the unconditional correlation in levels is very high while in first differences is respectively lower. Columns (3) and (4) show that conditional on states in which the investment currency tends to appreciate the correlation in first differences is much weaker than in states in which the investment currency is depreciating.

In terms of other conditional relationships I find that a higher domestic interest rate differential: 1) does *not* predict on average a larger domestic currency depreciation since the UIP regression coefficient is on average negative, not significantly different from zero in small samples, but significantly negative in large samples; 2) predicts a positive excess return from

the carry trade strategy; 3) predicts a negatively skewed excess payoff for the carry trade strategy; 4) is followed by a gradual appreciation of the domestic currency.

5.1 The UIP puzzle and positive carry trade payoffs

In the distorted expectations model UIP holds ex-ante and the expected excess payoffs under the worst-case scenario distribution $\widetilde{P}$ are zero:

$$s_{t+1} - s_t = r_t + \widetilde{\varepsilon}_{t+1} \tag{5.1}$$

where $E_t^{\widetilde{P}}(\widetilde{\varepsilon}_{t+1}|I_t) = 0$. The UIP regression is:

$$s_{t+1} - s_t = \beta r_t + \varepsilon_{t+1} \tag{5.2}$$

where P is the probability distribution implied by the true DGP and $E_t^P(\varepsilon_{t+1}|I_t) = 0$.

Let $\eta_t \equiv E_t^{\widetilde{P}}(s_{t+1}|I_t) - E_t^P(s_{t+1}|I_t)$ denote the expectational error of predicting s_{t+1}. Then combining (5.1) and (5.2) the UIP regression becomes:

$$s_{t+1} - s_t = r_t - \eta_t + \varepsilon_{t+1}.$$

The estimated UIP regression coefficient is then:

$$\widehat{\beta} = 1 - \frac{cov(\eta_t, r_t)}{var(r_t)}. \tag{5.3}$$

Intuitively, if when r_t is increasing the expected s_{t+1} under the market's expectations is increasing compared to the expected s_{t+1} under the true DGP, then the covariance in (5.3) is positive and the estimated $\widehat{\beta}$ can be negative.

Before presenting the results from the ambiguity aversion model, it is worth investigating as a benchmark the case of rational expectations but with risk aversion. As explained in Section 3, the solution for exchange rate under RE with a mean-variance utility is:

$$s_t = \frac{E_t(s_{t+1} - r_t)}{1 + 0.5c}$$

where $c = \gamma Var(s_{t+1})$. From equation (3.10) we have that the lower bound on $\widehat{\beta}$ is obtained by setting $\sigma_V = 0$, so:

$$\widehat{\beta} \geq \widehat{\beta}^L = \frac{1 - \rho}{1 + 0.5c - \rho}.$$

To investigate the magnitude of $\widehat{\beta}^L$ under risk aversion, I report below some simple

calculations based on the main parameterization for the interest rate differential from Table 1. The equilibrium coefficients a_1 and a_2 are given by formulas (3.7) and (3.8). Table 4 reports the model implied exchange rate volatility and $\widehat{\beta}^L$ for various levels of risk aversion γ. The conclusion that emerges from Table 4 is that with a low level of risk aversion the reaction of the exchange rate to the interest rate (equal in this case to a_1+a_2) is large and can generate significant variability in the exchange rate.[22] With a low risk aversion, the model implied $\widehat{\beta}$ is smaller than one, but very close to it. Although the model implied $\widehat{\beta}$ decreases with γ, even with a huge degree of absolute risk aversion the UIP regression coefficient is still positive and large. For example when $\gamma = 500$, the model implied $\widehat{\beta}^L$ is around 0.41.

Table 4: Rational expectations model

	$\gamma = 2$	$\gamma = 10$	$\gamma = 50$	$\gamma = 500$
$std_t(s_{t+1})$	0.0243	0.0222	0.0179	0.0105
$\widehat{\beta}^L$	0.97	0.89	0.71	0.41

This discussion highlights why a model of unstructured uncertainty discussed in Section 2.2 and analyzed in Appendix A, does not fare well in this setup. That type of model is equivalent to a rational expectations framework but with higher risk aversion. Driving the coefficient to zero from above requires appealing to enormous levels of risk aversion. Moreover, this high risk aversion would imply a minuscule response of the exchange rate to the interest rate to the point that the former is flat.

The focus of this paper is to show that a model with ambiguous precision of signals about a time-varying hidden state can provide an explanation for a negative $\widehat{\beta}$. In this model, agents at time t take a worst-case evaluation of the precision of the signals by underestimating, compared to the true DGP, the hidden state of the investment currency. That underestimation is increasing in the size of the hidden state since the Kalman gain multiplies the perceived innovation in the hidden state. Thus, the higher r_t is, the larger is the difference between the expected depreciation of the investment currency under $\widetilde{P}$ compared to P.

Table 5 presents the estimated $\widehat{\beta}$ for the model implied UIP regression in (5.2) in 1000 repeated samples of $T = 300$ and $T = 3000$. Column (1) is the benchmark specification and shows that, for small T, the model is generating a negative average $\widehat{\beta}$, even if not significant statistically. As the sample size is increased and the standard errors reduce, the average and

[22]This latter point has been made by Engel and West (2004) and Engel et al. (2007) who also show that when ρ is close to 1 these models are characterized by a low forecasting power for the interest rate differential in predicting the exchange rate change.

Table 5: Model implied $\widehat{\beta}$ for the UIP regression

		(1)		(2)		(3)		(4)	
		$\widehat{\beta}$	$t_{\widehat{\beta}}$	$\widehat{\beta}$	$t_{\widehat{\beta}}$	$\widehat{\beta}$	$t_{\widehat{\beta}}$	$\widehat{\beta}$	$t_{\widehat{\beta}}$
T=300	Mean	-0.42	-0.81	0.42	1.19	0.38	3.5	-0.62	-1.3
	Median	-0.37	-0.73	0.4	1.16	0.38	3.7	-0.6	-1.2
	St.dev.	0.22	0.38	0.08	0.3	0.05	0.4	0.18	0.51
T=3000	Mean	-0.2	-1.7	0.54	1.75	0.48	1.58	-0.34	-2.2
	Median	-0.19	-1.6	0.52	1.72	0.47	1.55	-0.31	-2.14
	St.dev.	0.03	0.17	0.11	0.32	0.09	0.2	0.05	0.21

median estimate become significant. This highlights that the results of the model are not limited to small samples and in fact are stronger in large samples. The type of ambiguity modeled in this paper is active even as the sample size increases.

Figure 2: Model implied UIP coefficients

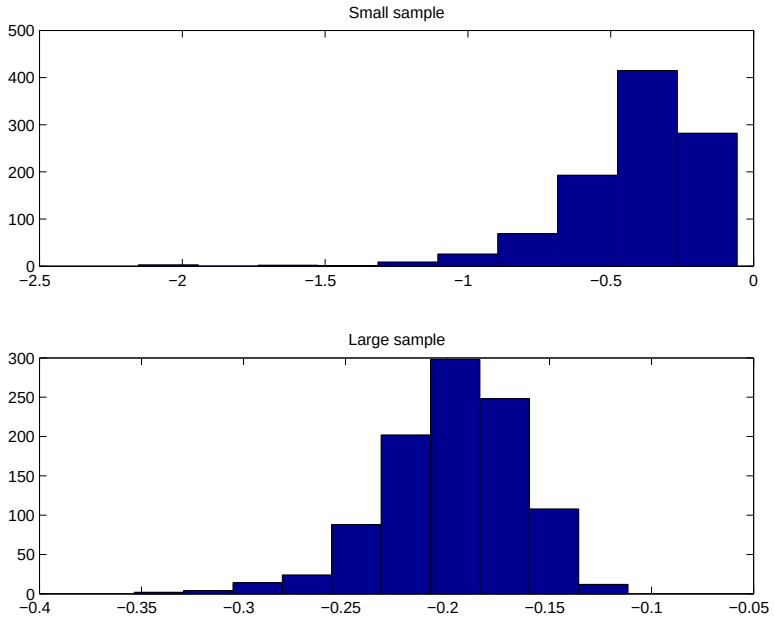


Figure 2 is the histogram of the UIP coefficients across many repeated samples. The top panel plots the histogram of the estimated UIP coefficients on $N = 1000$ samples of $T = 300$. It shows that the vast majority of the estimates are negative. For the large sample of $T = 3000$ the bottom panel of Figure 3 indicates the distribution of the estimates and shows that there are no positive values obtained.

Column (2) of Table 5 considers a case in which the noise-to-signal ratio in the true DGP

is much higher than in the benchmark model. In that case, the estimated UIP coefficient is positive. For that parameterization the steady state Kalman gain for the true DGP, high distorted precision and low distorted precision are 0.3, 1 and 0 respectively. The Kalman gain used by the agent in updating the estimated hidden state is time-varying to reflect the optimal response of the agents to "good" and "bad" news. This optimal time-variation implies two opposing forces on $\widehat{\beta}$: the underreaction makes the coefficient become negative and the overreaction pushes it to be larger than one. The combined effect depends on how far apart the distorted gains are from the one implied by the true DGP. In the benchmark case, as reported in Table 2, the true noise-to-signal ratio is small. In that case the overreaction channel is dominated because the Kalman gain implied by σ_V^L is closer to the reference model. Intuitively, if σ_V is close to zero any $\sigma_V^L < \sigma_V$ will make a small difference on the implied gain. However, if $\sigma_V^H > \sigma_V$ the distorted gain can be considerably smaller. In the variant reported in Column (2), the gain under the reference model is closer to the one implied by σ_V^H, so the underreaction effect is less active. That is the reason why the estimated UIP coefficient is positive.

Column (3) of Table 5 shows that when there is significantly less persistence in the state evolution the model cannot account for the UIP puzzle. For that experiment, the autocorrelation of the hidden state is 0.7. The reaction of the exchange rate to the estimate of the hidden state is strongly affected by this persistence because the present value of future payoffs to a bond is smaller following the same increase in the interest rate. For the same interest rate differential this makes agents demand less of the bond and the investment currency value goes up by less. As Engel and West (2004) argue, for a high persistence of the fundamentals the exchange rate is very sensitive to changes in the present value of future payoffs. They argue that this explains why exchange rates are hard to predict even in a model where they are completely determined by fundamentals. A large sensitivity of the currency's value to the hidden state allows small distortions to the estimate to produce large deviations in the exchange rate evolution. With significantly less persistence, the model has a difficult time in explaining the UIP puzzle.

These exercises show that the two main features of the true DGP that are required for the theory to succeed are relatively small temporary shocks and large persistence of the hidden state. The benchmark parameterization is characterized by these conditions because the data strongly suggests such a calibration. It is important to note though that Columns (2) and (3) of Table 5 show that the benchmark calibration is in fact rather robust to changes as significant variations have to be made to it to reverse the results.

Column (4) of Table 5 presents the results for the parameterization of the model in which there are no restrictions on the number of periods in which the agent can distort the reference

sequence, i.e. $n = t$. For such a case, the model implies a negative and larger in magnitude Fama regression coefficient both in short and large samples. Thus, relaxing the benchmark restrictions of $n < t$ improves the model's ability to generate the empirical puzzles at the cost of implying very unlikely distorted sequences compared to the reference model. As Column (3) of Table 2 reports, this case generates sequences that become increasingly less likely as the sample size grows.

5.2 Positive mean, negative skewness and excess kurtosis of ex-post carry trade payoffs

Section 5.1 described the model's implications for the UIP puzzle. The UIP condition was holding ex-ante under the equilibrium distorted beliefs so the underlying equilibrium speculation strategy has zero expected payoffs under these beliefs. The ex-post failure of the UIP condition has significant consequences for the ex-post profitability of the investment strategy. As argued before, in equilibrium, the investors of this model engage in the carry-trade strategy, which means borrowing in the low interest rate currency and investing in the high interest rate currency. Indeed, the agents' end-of-life wealth is given by $b_t(s_{t+1} - s_t - r_t)$. By the market clearing condition $b_t < 0$ when $s_t < 0$.[23] The payoff on a dollar bet for the carry trade strategy is:

$$\begin{aligned} z_{t+1} &= r_t - (s_{t+1} - s_t) \text{ if } b_t < 0 \\ z_{t+1} &= (s_{t+1} - s_t) - r_t \text{ if } b_t > 0 \end{aligned} \tag{5.4}$$

Using the definition of the carry trade in (5.4), Table 6 describes the model implied non-annualized monthly carry trade payoffs. These payoffs are characterized by a positive mean, negative skewness and excess kurtosis. The excess payoffs have a positive mean because a high interest rate differential predicts on average a zero currency depreciation or even a slight appreciation. The average mean payoff reported in Table 6 is 0.0016.

To compare the model-implied statistics with the data, I compute the payoffs to the carry trade for 16 developed countries for the period 1976-2008. The main statistical properties of those payoffs are reported in Table 11 of Appendix C. There I report that the computed average mean payoff for the carry trade strategy is 0.0041. This does not take into account transactions costs. Burnside et al. (2008) (henceforth BEKR) analyze a more extensive

[23] In simulations, the coefficient of correlation between s_t and r_t is around -0.97. The correlation is not -1 because the exchange rate depends mostly on the sign of the hidden state estimate and this can be different from the sign of the interest rate. Moreover, due to the "inaction" effect, in some cases agents choose not to invest at all if r_t is too close to zero. In these few case, when computing the carry trade payoffs I use the fact that such a payoff is zero.

Table 6: Model implied statistics for the carry trade payoffs

	Mean	Standard Deviation	Sharpe Ratio	Skewness	Kurtosis
$T = 300$	0.0016	0.0161	0.1	-0.135	5.83
	(0.0008)	(0.001)	(0.048)	(0.22)	(0.7)
$T = 3000$	0.0016	0.0172	0.1	-0.16	5.9
	(0.00002)	(0.00003)	(0.014)	(0.08)	(0.23)

data set and find that the average payoff to the carry trade without transactions costs across individual country pairs for the period 1976-2007 ranges from 0.0026 when the base currency is the GBP to 0.0042 when the base currency is the USD. With transaction costs they report a range of 0.0015 to 0.0025. Table 6 also shows that the average standard deviation of the model implied carry trade payoffs is around 0.016. For the data analyzed in Table 11 the average standard deviation is 0.031.

Thus, compared to empirical evidence on the carry trade payoffs, the present model delivers mean payoffs that are around half of those computed without transaction costs and at the lower bound of the empirical payoffs with transaction costs. The model implied average standard deviation of these payoffs is also around half of its empirical counterpart. With both the mean and the standard deviation both lower, the model-implied carry trade payoffs are characterized by a Sharpe ratio of 0.1, very close to the reported empirical value of 0.133.

Figure 3 plots the histogram of the realized carry trade payoffs obtained for $N = 1000$ samples of size $T = 300$. The mean of these payoffs is positive and, compared to a normal distribution with the same mean and variance, the payoffs are negatively skewed and have significant excess kurtosis.

Besides the positive mean of the carry trade payoffs, a very interesting and important feature of the empirical payoffs is the negative skewness. To study this property, table 6 indicates that the model-implied payoffs to the carry trade are on average negatively skewed. The degree of skewness is slightly lower than the one found in the data for the carry trade payoffs analyzed in Table 11. The model implies a negative skewness of -0.135 while the average for the countries in Table 11 is -0.26.

To investigate further the properties of the realized skewness of excess payoffs I construct two tests. The first, more cross-sectional in nature, is similar to that of Brunnermeier et al. (2008). It involves checking whether periods (countries in Brunnermeier et al. (2008)) characterized by a higher domestic currency also experience a negative skewness in the excess

Figure 3: Model generated carry trade payoffs

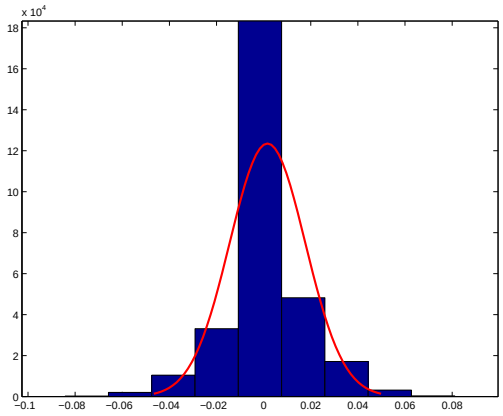


Table 7: Model implied statistics for the "crash risk" regressions

$T = 300$		$T = 3000$	
$\widehat{\beta}_2$	$t_{\widehat{\beta}_2}$	$\widehat{\beta}_2$	$t_{\widehat{\beta}_2}$
-2.2	-6.8	-1.93	-6.32
(1.04)	(2.33)	(0.81)	(1.65)

payoffs. To that end, I simulate the model for $T = 300$ and for each t collect r_t and the realized excess payoffs $ex_{t+1} = r_t - (s_{t+1} - s_t)$. I sort the excess payoffs ex_{t+1} according to the sign of r_t. Denote by ex^+_{t+1} the payoffs when $r_t > 0$ and by ex^-_{t+1} when $r_t < 0$. Consistent with the predictability of excess payoffs, the average of ex^+_{t+1} is positive and the average of ex^-_{t+1} is negative. Importantly, I find that the skewness of ex^+_{t+1} is negative and that of ex^-_{t+1} is positive.

The second test, for a time varying dimension, is to simulate the model and at each date t to take $N = 20000$ draws from the DGP process for the date $t + 1$ realizations of r_{t+1}. Using these draws, I solve the model at time $t + 1$ and then collect the equilibrium implied s_{t+1}. Based on these, I compute the realized excess payoffs $ex_{t+1} = r_t - (s_{t+1} - s_t)$ and their skewness denoted by $Skew_{t+1}$. The same conclusions hold: a higher r_t predicts a lower $Skew_{t+1}$. Positive r_t are associated with negative $Skew_{t+1}$. Table 7 reports the results of the regression

$$Skew_{t+1} = \beta_2 r_t + \xi_{2,t+1}.$$

The model generates a negative significant $\widehat{\beta}_2$ as found empirically by Jurek (2008) who investigates both the cross-section and the time variation and finds that in both dimensions

the effect is very significant.

Consistent with the data, these results imply that investing in a higher interest rate currency produces an average positive excess return which is negatively skewed. The thicker left tail occurs because of the larger reaction to negative innovations and the smaller reaction to positive shocks in the high-interest-rate. Thus "crash risk" in this model is endogenous and it happens when negative shocks hit an otherwise positive estimate of the hidden state.

5.3 Delayed overshooting

The forward premium puzzle refers to the unconditional empirical failure of UIP. This does not necessarily imply that a conditional version of UIP fails too. Following a positive shock to the interest rate the UIP condition states that the domestic currency should appreciate on impact and then follow a depreciation path. Such a mechanism can be investigated empirically by identifying the response of the exchange rate to a monetary policy shock.

Several studies have analyzed this conditional UIP using different identification restrictions. Eichenbaum and Evans (1995), Grilli and Roubini (1996) use short-run restrictions to identify the effect of structural monetary policy shock on the exchange rate. They find significant evidence of delayed overshooting: following a contractionary monetary policy shock the domestic interest rate increases and there is a prolonged period of a domestic currency appreciation. The peak of the impact occurs after one to three years as opposed to happening immediately as predicted by the Dornbusch (1976) overshooting model. Faust and Rogers (2003) find that these results are sensitive to the recursive identification assumptions and that the peak of the exchange rate response is imprecisely estimated. Scholl and Uhlig (2006) use sign restrictions and also find evidence for delayed overshooting. For the country pairs they analyze, the estimated peak occurs within a year or two. Although they differ in their estimates of the length of the delayed overshooting effect, all of these identifying approaches reach a robust conclusion: following a monetary policy shock there are significant deviations from the UIP.

As discussed in Section 1, the model presented in this paper attempts to explain the delayed overshooting puzzle through a mechanism similar to Gourinchas and Tornell (2004). There they posit that if agents, for some reason, systematically underreact to news this behavior can explain the conditional and unconditional UIP puzzles. The difference is that here I investigate a model which addresses the origin and optimality of such beliefs. As explained in previous sections, I find that agents underreact to good news and overreact to bad news. However, for the impulse response function that I analyze and that is typical to the empirical identification, it is only the underreaction effect that shows up, which makes

the intuition of the delayed overshooting similar to the one in Gourinchas and Tornell (2004).

To generate the impulse response of the exchange rate to a shock to the interest rate differential, assume that the economy starts in steady state. Thus, $r_{t-1} = 0$, $b_{t-1} = 0$ and $E^{\tilde{P}}_{t-1} s_t = E^P_{t-1} s_t = 0$. At time t there is a positive shock to the interest rate differential. Take again the true DGP as in (2.1):

$$\begin{aligned} r_t &= x_t + \sigma_V v_t \\ x_t &= \rho x_{t-1} + \sigma_U u_t \end{aligned} \tag{5.5}$$

The values for σ_V, σ_U and ρ are the same as in Section 4.5.

To investigate the average response to an increase in r_t this experiment needs to impose that the observed positive shock to r_t is generated by a combination of a shock to the persistent and the temporary component that corresponds to their true DGP likelihood of occurrence. Suppose r_t increases by α. When $\sigma_V = 0$ this means that the increase in r_t was generated by a shock to u_t, i.e. $\sigma_U u_t = \alpha$. When $\sigma_V > 0$ then the experiment imposes that:

$$x_t = \rho x_{t-1} + E(\sigma_U u_t | \sigma_U u_t + \sigma_V v_t = \alpha) = \rho x_{t-1} + \alpha \frac{\sigma_U^2}{\sigma_V^2 + \sigma_U^2}$$

The next periods shocks are set equal to zero. The RE solution is then: $s_t^{RE} = a_1 \widehat{x}_{t,t}^{RE} + a_2 r_t$ and the distorted expectations solution is $s_t = a_1 \widehat{x}_{t,t}^{t} + a_2 r_t$, where $\widehat{x}_{t,t}^{RE}$ is the estimate of the hidden state x_t under RE and $\widehat{x}_{t,t}^{t}$ is the estimate under the ambiguity aversion model. Under $\sigma_V = 0$, $\widehat{x}_{t,t}^{RE} = \alpha$ while $\widehat{x}_{t,t}^{t} \neq \alpha$.

Figure 4 plots the dynamic response of the exchange rate to the observed increase in r_t. The RE version features the Dornbusch (1976) overshooting result in which the peak of the impact is at time t. Consider the decision at time t. The agent sees the increase in the interest rate differential but she is worried about a significant depreciation at time $t+1$. In equilibrium, she is concerned that this rise in r_t is caused by a temporary rise. She then believes that the true $\sigma_{V,t} = \sigma_V^H$ and acts on this belief by investing much less in the domestic bond than she would under rational expectations. Because we start from the steady state, her estimate at time t is then:

$$\begin{aligned} \widehat{x}_{t,t}^{t} &= K_t^t \alpha \\ K_t^t &= \frac{\rho^2 \Sigma + \sigma_U^2}{\rho^2 \Sigma + \sigma_U^2 + \sigma_V^{2,H}} < 1 \\ \Sigma_{t,t} &= (1 - K_t^t)\sigma_U^2 \end{aligned}$$

where $\Sigma = 0$ since it is the steady state variance of the estimate of the hidden state, which if $\sigma_V = 0$, is perfectly observed in steady state. By underestimating the true hidden state she observes a higher than expected r_{t+1}. Her updated estimate at time $t+1$ is:

$$\begin{aligned}
\widehat{x}^{t+1}_{t+1,t+1} &= \rho\widehat{x}^{t+1}_{t,t} + K^{t+1}_{t+1}(r_{t+1} - \rho\widehat{x}^{t+1}_{t,t}) \\
K^{t+1}_{t+1} &= \frac{\rho^2\Sigma_{t,t} + \sigma^2_U}{\rho^2\Sigma_{t,t} + \sigma^2_U + \sigma^{2,H}_V} \\
\widehat{x}^{t+1}_{t,t} &= \alpha\left[\frac{\rho^2\Sigma + \sigma^2_U}{\rho^2\Sigma + \sigma^2_U + \sigma^{2,H}_V}\right] = \widehat{x}^{t}_{t,t}.
\end{aligned}$$

The reason why here the agent at time $t+1$ chooses to act as if $\sigma_{V,t} = \sigma_{V,t+1} = \sigma^H_V$ is that at time $t+1$ she will invest in the domestic bond. This means that she will treat the observed innovations as reflecting more likely a temporary shock. In this setting, $\widehat{x}^{t+1}_{t,t} = \widehat{x}^{t}_{t,t}$ because the worst-case scenario estimate of the time t hidden state is the same from the perspective of the agent at time t and time $t+1$.

Given our experiment, there are no shocks following period t, so $r_{t+1} = \rho x_t = \rho\alpha$. Thus,

$$\widehat{x}^{t+1}_{t+1,t+1} = \rho\alpha K^t_t + \rho\alpha K^{t+1}_{t+1}(1 - K^t_t)$$

and under the benchmark parameterization $K^t_t = 0.04$ and $K^{t+1}_{t+1} = 0.07$. It then follows that

$$\widehat{x}^{t+1}_{t+1,t+1} > \widehat{x}^{t}_{t,t}$$

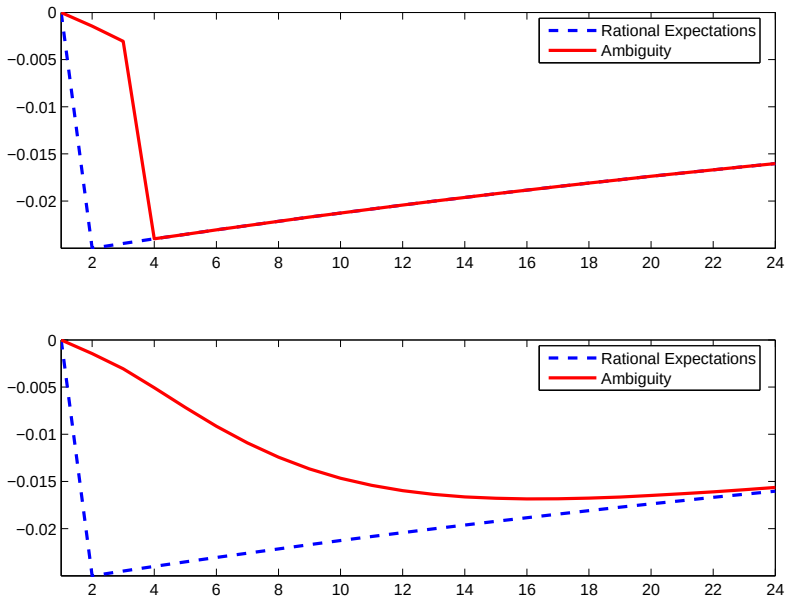
At time t there was an appreciation caused by the positive α, but, more importantly, because the estimate of the hidden state at $t+1$ is higher than at time t, a further appreciation occurs at $t+1$:

$$0 > s_t > s_{t+1}$$

At time $t+2$, if the agent can distort only 2 periods as in the benchmark specification then, as Remark 1 stated, $\sigma_{V,(t+2),t} = \sigma_V = 0$ and $\sigma_{V,(t+2),t+1} = \sigma_{V,(t+2),t+2} = \sigma^H_V$. If there would be no restriction on n, so that the agent can distort any period then $\sigma_{V,(t+2),t} = \sigma_{V,(t+2),t+1} = \sigma_{V,(t+2),t+2} = \sigma^H_V$. A similar argument applies for future periods. Eventually, the estimate of the hidden state converges to the RE case.

The top plot in Figure 4 shows the evolution of s_t in the benchmark specification in which the agent distorts the sequence $\sigma_V(r^t)$ only for $n = 2$ periods. There the shock occurs in period 2. The rational expectations model predicts overshooting so the peak of the response is at period 2. With ambiguity aversion and $n = 2$, the peak response occurs 2 periods later than in the rational expectations model. There is a gradual appreciation until period

Figure 4: Delayed overshooting



4, followed by a depreciation after the estimates of the hidden state converge under the distorted and rational expectations. The bottom plot is one in which there is no restriction on n, i.e. $n = t$. The agent can distort any period of the sample she observes, and in this case she does so by choosing a low precision of the signal for every period. In this case the appreciation is much more gradual and the peak is 14 periods later than the time of the shock. This plot also reproduces the intuition for the persistent delayed overshooting in Gourinchas and Tornell (2004). However, as I argued in the discussion of the statistical plausibility of the distorted sequences, such a specification implies a very unlikely interpretation of the observed sample. In fact, as the sample size increases such distorted sequences generate likelihoods that become increasingly lower than the ones under the reference model.

The conclusion from Figure 4 is that the model can generate qualitatively the delayed overshooting implication. The benchmark specification implies a quick peak and a short-lived deviation from UIP since the agent is limited in distorting the time-varying precision of signals by statistical plausibility considerations. When these considerations are absent, the model delivers significantly longer delayed overshooting.

5.4 A modified carry trade strategy

A further empirical implication of the model is that there are conditioning variables that can improve upon the ex-post profitability of the carry trade. Recall that the standard carry trade

was to borrow in the low interest rate currency and lend in the high interest rate currency. The conditioning variable in that strategy was the sign of the interest rate differential. The model predicted ex-post positive payoffs to the strategy because of the underestimation of the hidden state controlling the investment differential, i.e. the differential between the high and the low interest rate. Part of the agents' optimal response to news was to underreact to increases in the investment differential. Thus, if one conditions on the size of the innovation in the investment differential, the model predicts significant differences in the ex-post profitability of the speculation strategy.

As evident from the optimal decision rule in (4.6) or from the experiment of the delayed overshooting, the higher the innovation is at time t in the investment differential, the smaller $\widehat{x}^t_{t,t}|$ is compared to the estimate under RE, $|\widehat{x}^{RE}_{t,t}|$, the larger is $|E^{\widetilde{P}}_t s_{t+1} - E^P_t s_{t+1}|$ and thus the larger is the ex-post profitability of the carry trade. Thus, a modified carry trade strategy suggested by the model is the following:

$$\begin{aligned} z_{t+1} &= r_t - (s_{t+1} - s_t) \text{ if } r_t > E^{\widetilde{P}}_{t-1} r_t > 0 \\ z_{t+1} &= (s_{t+1} - s_t) - r_t \text{ if } r_t < E^{\widetilde{P}}_{t-1} r_t < 0 \end{aligned} \tag{5.6}$$

The strategy differs from the standard one in (5.4) by conditioning the direction of the speculation not only on the sign of r_t but also on the size of the innovation in the investment differential. When taking this implication to the data, I will make use of the innovation under the true DGP, $E^P_{t-1} r_t$, since $E^{\widetilde{P}}_{t-1} r_t$ is not readily observed. Since in the model $sign[(E^P_{t-1} r_t - E^{\widetilde{P}}_{t-1} r_t) r_t] > 0$ the modified carry trade strategy can be made more stringent by conditioning on $(r_t - E^P_{t-1} r_t)$:

$$z_{t+1} = sign(r_t)[r_t - (s_{t+1} - s_t)] \text{ if } r_t(r_t - E^P_{t-1} r_t) > 0. \tag{5.7}$$

In the benchmark parameterization $E^P_{t-1} r_t = \rho \widehat{x}^{RE}_{t-1,t-1}$ which can be computed by using the reference sequence of constant variances σ_V. When $\sigma_V = 0$ this implies $E^P_{t-1} r_t = \rho r_{t-1}$.

Table 8 presents the model implied payoffs to the modified carry trade payoffs described in (5.7).[24] It shows that the ex-post profitability of this modified carry trade is much larger in the model.

Table 9 analyzes a similar strategy in the data. For that, I compute $E^P_{t-1} r_t$ by finding the AR(p) representation that fits best, in terms of the BIC criterion, the interest rate differential

[24]It should be noted that the payoffs are computed only over the time periods when this strategy is "active". As the strategy in (5.7) implies, there are periods in which neither of the conditions deciding the investment direction is satisfied. The same way periods of inaction were excluded from the computation of the payoffs to the standard carry trade strategy reported in Table 6, the periods of inaction implied by this strategy are not included. The payoffs are reported as $N = 1000$ simulations with a time length of $T = 300$.

Table 8: Model implied statistics for the modified carry trade payoffs

Mean	Standard Deviation	Sharpe Ratio	Skewness	Kurtosis
0.007	0.0162	0.43	−0.135	5.83
(0.0007)	(0.001)	(0.04)	(0.22)	(0.7)

characterizing each country pair analyzed. Table 9 reports the average statistics across the country pairs analyzed. Row (1) reports results for the carry trade strategy described in (5.7). It shows that such a modified strategy outperforms the standard carry strategy by generating a higher mean return. In Appendix C, Table 12 provides detailed statistics for each country pair for this strategy. For 15 out of the 16 countries analyzed, the modified carry trade delivers higher payoffs.

As noted, the reason why the modified strategy outperforms the standard strategy in the model is that it conditions on positive innovations in the investment currency which are incorporated slowly in the estimates of the hidden state. By this logic, the model also implies that when the modified strategy conditions on innovations that are not only positive but increasingly larger than a positive number, the modified strategy performs ex-post increasingly better because the investment currency is increasingly likely to appreciate ex-post. If I define the payoffs to the modified strategy as following the rule in (5.8), then the model implies that the average ex-post payoffs are increasing in the threshold μ :

$$z_{t+1} = \left\{ \begin{array}{l} [r_t - (s_{t+1} - s_t)] \text{ if } r_t > 0, (r_t - E^P_{t-1} r_t) > \mu > 0 \\ [(s_{t+1} - s_t) - r_t] \text{ if } r_t < 0, (r_t - E^P_{t-1} r_t) < -\mu < 0 \end{array} \right\} \tag{5.8}$$

Table 9: Standard and modified carry trade payoffs, empirical averages

Averages	Standard carry trade payoffs			Modified carry trade payoffs		
	Mean	Standard Deviation	Sharpe Ratio	Mean	Standard Deviation	Sharpe Ratio
(1)	0.0041	0.031	0.133	0.0052	0.031	0.171
(2)	0.0041	0.031	0.133	0.0073	0.034	0.214
(3)	0.0041	0.031	0.133	0.0109	0.036	0.295

To evaluate this implication in the data, I construct payoffs based on (5.8) where I vary μ. Row (1) in Table 9 implicitly set $\mu = 0$. Row (2) in Table 9 reports such results for $\mu = 0.5\sigma(r_t - E^P_{t-1} r_t)$, where $\sigma(r_t - E^P_{t-1} r_t)$ denotes the sample standard deviation of the

innovations $r_t - E^P_{t-1} r_t$. In row (3) of Table 9, μ is increased to $\sigma(r_t - E^P_{t-1} r_t)$. These two thresholds produce average monthly Sharpe ratio that are around 60% and 120% respectively larger than for the standard carry trade. The conclusion from these empirical exercises is that the modified carry trade strategies can generate average payoffs that are much larger than the standard carry trade and that this ex-post profitability is increasing with the threshold μ. Such results provide further support for the model's implications.

6 Conclusions

This paper contributes to the theoretical literature that attempts to explain the observed deviations from UIP through systematic expectational errors. Such an approach is motivated by the empirical literature based on survey data for the foreign exchange market that finds significant evidence against the rational expectations assumption and the empirical research that challenges the time-varying risk assumption.

I present a model of exchange rate determination which features signal extraction by an ambiguity averse agent that is uncertain about the precision of the signals she receives. When deciding on the optimal investment position, the agent is estimating the time-varying hidden state of the exogenous observed interest rate differential. In equilibrium, the agent invests in the higher interest rate currency (investment currency) by borrowing in the lower interest rate currency (funding currency). The agent entertains the possibility that the data could have been generated by various sequences of time-varying signal to noise ratios. Faced with uncertainty agents choose to act on pessimistic beliefs so that, compared to the true DGP, they underestimate the hidden state of the differential between the interest rate paid by the bonds in the investment and funding currency. Given the assumed structure of uncertainty, agents underestimate the hidden state by reacting in equilibrium asymmetrically to signals about it: they treat positive innovations, which in equilibrium are good news for the investor, as reflecting a temporary shock, but negative innovations, which are bad news in equilibrium, as signaling a persistent shock.

The systematic underestimation implies that agents perceive on average positive innovations when updating the estimate. This creates the possibility of a further increased demand next period for the investment currency and a gradual appreciation of it. Thus the model can provide an explanation for the UIP and delayed overshooting puzzle.

I find through model simulation that the benchmark specification generates an asymptotically negative UIP regression coefficient. In small samples the magnitude of the coefficient is similar but it is less significant statistically. In comparative statistics exercises I find that the coefficient becomes positive, even though smaller than one, if the true DGP is

characterized by a significantly less persistent hidden state and larger temporary shocks. The benchmark specification also imposes constraints on the set of possible distortions that the agent contemplates implying that the equilibrium subjective probability distribution is close statistically to the objective ones. If these constraints are relaxed, the same qualitative results hold but quantitatively they become stronger.

The model provides a unified explanation for the main stylized facts of the excess currency payoffs: predictability, negative skewness and excess kurtosis. Predictability is directly related to the ex-post failure of UIP: investing in the investment currency by borrowing in the lower funding currency delivers positive payoffs. The benchmark calibration implies positive but smaller and less variable excess payoffs than in the data. The negative skewness is caused by the asymmetric response to news. On one hand, when the interest rate of the investment currency decreases compared to the market's expectation agents respond strongly to this negative news and the investment currency depreciates more than in the rational expectations model. On the other hand, when there is a positive innovation in this interest rate agents underreact to this information and the currency appreciates slower. Excess kurtosis is a manifestation of the fact that the equilibrium interaction between the subjective and objective probability distribution implies small excess payoffs more often.

The model predicts that modified carry trade strategies can be implemented that are characterized by significantly higher ex-post profitability than the standard carry trade. The conditioning variable for such more profitable strategies is the innovation in the investment differential, i.e. the differential between the higher interest rate and the lower interest rate. When this innovation is positive the model implies a gradual incorporation of its informational content thus predicting a higher likelihood of observing ex-post an appreciation of the investment currency. In fact, the larger the positive innovation is, the higher is such a probability. Implementing such strategies in the data, I find that, as the model suggests, they do generate significantly higher Sharpe ratios than the standard carry trade.

The theory proposed in this paper can be applied to other settings that involve forecastability of excess returns. Bacchetta et al. (2008) use survey data to conclude that most of the predictability of excess returns in bond, stock and foreign exchange market is caused by predictability of expectational errors. Interestingly, in the stock market similar momentum strategies and impulse responses as the delayed overshooting puzzle have been documented (as for example in Hong and Stein (1999)) in which stock prices tend to respond slowly to new public releases. The model analyzed in this paper proposes dynamic filtering of signals with uncertain precision as a mechanism to generate predictable expectational errors and excess returns.

References

Alvarez, F., A. Atkeson, and P. J. Kehoe (2008): "Time-varying Risk, Interest Rates, and Exchange Rates in General Equilibrium," Staff Report 371, Federal Reserve Bank of Minneapolis.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and C. Vega (2003): "Micro Effects of Macro Announcements: Real-Time Price Discovery in Foreign Exchange," *American Economic Review*, 93, 38–62.

Anderson, E. W., L. P. Hansen, and T. J. Sargent (2003): "A Quartet of Semigroups for Model Specification, Robustness, Prices of Risk, and Model Detection," *Journal of the European Economic Association*, 1, 68–123.

Bacchetta, P., E. Mertens, and E. van Wincoop (2008): "Predictability in Financial Markets: What Do Survey Expectations Tell Us?" *Journal of International Money and Finance*, forthcoming.

Bacchetta, P. and E. van Wincoop (2009): "Infrequent Portfolio Decisions: A Solution to the Forward Discount Puzzle," *American Economic Review*, forthcoming.

Bansal, R. and I. Shaliastovich (2007): "Risk and Return in Bond, Currency and Equity Markets," Mimeo.

Barillas, F., L. Hansen, and T. Sargent (2009): "Doubts or Variability?" *Journal of Economic Theory*, 144, 2388–2418.

Basar, T. and P. Bernhard (1995): H_∞ *Optimal Control and Related Minimax Design Problems*, Birkhuser, Boston.

Benigno, P. (2007): "Portfolio Choices with Near Rational Agents: A Solution for Some International-Finance Puzzles," NBER Working Papers 13173.

Brunnermeier, M., S. Nagel, and L. Pedersen (2008): "Carry Trades and Currency Crashes," *NBER Macroeconomics Annual*, forthcoming.

Burnside, A. C., M. S. Eichenbaum, I. Kleshchelski, and S. Rebelo (2008): "Do Peso Problems Explain the Returns to the Carry Trade?" NBER WP 14504.

Chinn, M. D. and J. A. Frankel (2002): "Survey Data on Exchange Rate Expectations: More Currencies, More Horizons, More Tests," in *Monetary Policy, Capital Flows and Financial Market Development in the Era of Financial Globalization: Essays in Honor of Max Fry*, ed. by W. Allen and D. Dickinson, London, Routledge, 145–167.

Chinn, M. D. and G. Meredith (2005): "Testing Uncovered Interest Parity at Short and Long Horizons during the Post-Bretton Woods Era," NBER Working Papers 11077.

Dornbusch, R. (1976): "Expectations and Exchange Rate Dynamics," *Journal of Political Economy*, 84, 1161–76.

Eichenbaum, M. and C. L. Evans (1995): "Some Empirical Evidence on the Effects of Shocks to Monetary Policy on Exchange Rates," *The Quarterly Journal of Economics*, 110, 975–1009.

Engel, C. (1996): "The Forward Discount Anomaly and the Risk Premium: A Survey of Recent Evidence," *Journal of Empirical Finance*, 3, 123–192.

Engel, C., N. C. Mark, and K. D. West (2007): "Exchange Rate Models Are Not as Bad as You Think," NBER Working Papers 13318.

Engel, C. and K. D. West (2004): "Accounting for Exchange-Rate Variability in Present-Value Models When the Discount Factor Is Near 1," *American Economic Review*, 94, 119–125.

Epstein, L. G. and M. Schneider (2007): "Learning Under Ambiguity," *Review of Economic Studies*, 74, 1275–1303.

——— (2008): "Ambiguity, Information Quality, and Asset Pricing," *Journal of Finance*, 63, 197–228.

——— (2010): "Ambiguity and Asset Markets," *Annual Review of Financial Economics*, forthcoming.

Fama, E. F. (1984): "Forward and Spot Exchange Rates," *Journal of Monetary Economics*, 14, 319–338.

Farhi, E. and X. Gabaix (2008): "Rare Disasters and Exchange Rates," NBER Working Papers 13805.

Faust, J. and J. H. Rogers (2003): "Monetary Policy's Role in Exchange Rate Behavior," *Journal of Monetary Economics*, 50, 1403–1424.

Froot, K. A. and J. A. Frankel (1989): "Forward Discount Bias: Is It an Exchange Risk Premium?" *The Quarterly Journal of Economics*, 104, 139–61.

Froot, K. A. and R. H. Thaler (1990): "Anomalies: Foreign Exchange," *Journal of Economic Perspectives*, 4, 179–92.

Gilboa, I. and D. Schmeidler (1989): "Maxmin Expected Utility with Non-unique Prior," *Journal of Mathematical Economics*, 18, 141–153.

Gourinchas, P.-O. and A. Tornell (2004): "Exchange Rate Puzzles and Distorted Beliefs," *Journal of International Economics*, 64, 303–333.

Grilli, V. and N. Roubini (1996): "Liquidity Models in Open Economies: Theory and empirical evidence," *European Economic Review*, 40, 847–859.

Hamilton, J. D. (1994): *Time Series Analysis*, Princeton University Press.

Hansen, L. P. (2007): "Beliefs, Doubts and Learning: Valuing Macroeconomic Risk," *American Economic Review*, 97, 1–30.

Hansen, L. P. and R. J. Hodrick (1980): "Forward Exchange Rates as Optimal

Predictors of Future Spot Rates: An Econometric Analysis," *Journal of Political Economy*, 88, 829–53.

Hansen, L. P. and T. Sargent (2008a): "Fragile Beliefs and the Price of Model Uncertainty," Mimeo.

Hansen, L. P. and T. J. Sargent (2007): "Recursive Robust Estimation and Control without Commitment," *Journal of Economic Theory*, 136, 1–27.

——— (2008b): *Robustness*, Princeton University Press.

Hong, H. and J. C. Stein (1999): "A Unified Theory of Underreaction, Momentum Trading and Overreaction in Asset Markets," *Journal of Finance*, 54, 2143–2184.

Illeditsch, P. (2009): "Ambiguity Aversion, Risk Aversion, and Asset Pricing," Mimeo.

Ju, N. and J. Miao (2009): "Ambiguity, learning, and asset returns," Mimeo.

Jurek, J. W. (2008): "Crash-neutral Currency Carry Trades," Mimeo.

Kleshchelski, I. and N. Vincent (2007): "Robust Equilibrium Yield Curves," Mimeo.

Lewis, K. K. (1995): "Puzzles in International Finance Markets," in *Handbook of International Economics*, ed. by G. G. Grossman and K. Rogoff, 1913–1971.

Li, M. and A. Tornell (2008): "Exchange Rates Under Robustness: The Forward Premium Puzzle and Momentum," Mimeo.

Lyons, R. (2001): *The Microstructure Approach to Exchange Rates*, MIT Press:Cambridge, MA.

Maccheroni, F., M. Marinacci, and A. Rustichini (2006): "Ambiguity Aversion, Robustness, and the Variational Representation of Preferences," *Econometrica*, 74, 1447–1498.

Scholl, A. and H. Uhlig (2006): "New Evidence on the Puzzles: Monetary Policy and Exchange Rates," Working Paper 5, Society for Computational Economics.

Strzalecki, T. (2007): "Axiomatic Foundations of Multiplier Preferences," Mimeo.

Verdelhan, A. (2008): "A Habit-Based Explanation of the Exchange Rate Risk Premium," *Journal of Finance*, forthcoming.

Whittle, P. (1981): "Risk-sensitive Linear/Quadratic/Gaussian control," *Advances in Applied Probability*, 13, 764–777.

APPENDIX

A Unstructured uncertainty

The multiplier preference is the modification to the expected utility that involves solving the problem:

$$\max_{b} \min_{\widetilde{P}\in\Phi} E^{\widetilde{P}}[U(c(b;\varepsilon))] + \theta R(\widetilde{P}|P) \tag{A.1}$$

where $U(c(b;\varepsilon))$ is the utility function derived from the consumption plan $c(b;\varepsilon)$, with b being the control and ε the underlying stochastic process. The parameter θ is controlling the level of uncertainty aversion and Φ is a closed and convex set of probability measures and $R(\widetilde{P}|P)$ is the relative entropy of probability measure $\widetilde{P}$ with respect of measure P :

$$R(\widetilde{P}|P) = \left\{ \begin{array}{c} \int_{\Omega} \log(\frac{d\widetilde{P}}{dP}) d\widetilde{P} \text{ if } \widetilde{P} \text{ is absolutely continuous w.r.t } P \\ \infty \text{ otherwise} \end{array} \right\} \tag{A.2}$$

Hansen and Sargent (2008b) refer to the situation in which there is no restriction on the nature of Φ as unstructured uncertainty. For this case, as shown for example in Strzalecki (2007) the problem in (A.1) is equivalent to[25]:

$$\max_{b} E^{P}[-\exp(-\frac{1}{\theta}U(c(b;\varepsilon))]$$

In the present case, where $U(c(b;\varepsilon) = b_t(s_{t+1} - s_t - r_t)$, using the multiplier preferences would be equivalent to maximizing a negative exponential utility under the reference P :

$$\max_{b} E^{P}[-\exp(-\frac{1}{\theta}b_t q_{t+1})]$$

In the case of normality of q_{t+1} this would in turn be equivalent to maximizing a mean-variance utility with an absolute risk aversion of $\frac{1}{\theta}$ so that the results of section 5.1 can be used to show that for the present setup introducing the variance channel generates minimal risk corrections.

In the above formulas, the alternative measure $\widetilde{P}$ was taken with respect to the interest rate differential, thus allowing unstructured uncertainty around the reference process of r_t. If one instead focuses on uncertainty about x_t, then from work such as Hansen and

[25]The equivalence is true in a Savage setting. This result is well known in the decision theory literature and in the literature on large-deviations. For a meaningful distinction between the two preferences Strzalecki (2007) stresses the importance of using the Anscombe-Aumann setting, where objective risk coexists with subjective uncertainty.

Sargent (2007), it is not clear that the qualitative results are the same. However, in the present setup with risk-neutrality it is equivalent to use unstructured uncertainty about r_t or "unstructured" uncertainty on x_t.To illustrate this, consider the problem in which the agent takes the process in (2.1) as an approximated model and she surrounds it with a set of alternative models such as:

$$\begin{aligned} r_{t+1} &= x_{t+1} + \sigma_V v_{t+1} + \epsilon_{t+1} \\ x_t &= \rho x_{t-1} + \sigma_U u_t \end{aligned}$$

The shocks ϵ_t can have non-linear dynamics that feed back on the history of the state variables. Thus r_{t+1}, conditional on x_t, is distributed $N(Fx_t + \epsilon_{t+1}, \sigma_U^2 + \sigma_V^2)$. Under RE, the hidden state x_t is distributed $N(\widehat{x}_{t,t}^{RE}, \Sigma_{t,t})$ where $\widehat{x}_{t,t}^{RE}$ is the estimate under Kalman Filter for (2.1). In this setting, Hansen and Sargent (2007) propose two robustness corrections: one that distorts r_{t+1}, conditional on x_t, through the mean of (ϵ_{t+1}) and another that distorts the distribution over the hidden state. Hansen and Sargent (2007) analyze the case in which the reference model for the hidden state is given by the Kalman Filter applied to the approximating state-space representation (2.1). Hence x_t is distributed as $N(\widehat{x}_{t,t}^{RE} + \varepsilon_t, \Sigma_{t,t})$ and ε_t is the arbitrary unknown conditional mean distortion of the hidden state. These alternative models are constrained to be close to the approximating model by using the conditional relative entropy defined in (A.2) above. After taking into account these distortions the maximization occurs under the transformed conditional distribution

$$r_{t+1} \overset{\widetilde{P}}{\sim} N(\rho\widehat{x}_{t,t}^{RE} + \rho\varepsilon_t + \epsilon_{t+1}, \rho^2\Sigma_{t,t} + \sigma_U^2 + \sigma_V^2). \tag{A.3}$$

Note that under the reference model

$$r_{t+1} \overset{P}{\sim} N(\rho\widehat{x}_{t,t}^{RE}, \rho^2\Sigma_{t,t} + \sigma_U^2 + \sigma_V^2) \tag{A.4}$$

The relative entropy in (A.2) for the implied distributions $\widetilde{P}$ in (A.3) and P in (A.4) is:

$$R(\widetilde{P}|P) = \frac{(\rho\varepsilon_t + \epsilon_{t+1})^2}{2(\rho^2\Sigma_{t,t} + \sigma_U^2 + \sigma_V^2)}$$

Denote the overall distortion $\rho\varepsilon_t + \epsilon_{t+1}$ by ω_{t+1} and use that $Var_t^P(r_{t+1}) = \rho^2\Sigma_{t,t} + \sigma_U^2 + \sigma_V^2$. The multiplier preferences defined in (A.1) imply:

$$\max_{b_t} \min_{\omega_{t+1}} E_t^{\widetilde{P}}[b_t(s_{t+1} - s_t - r_t)] + \theta \frac{\omega_{t+1}^2}{2Var_t^P(r_{t+1})} \tag{A.5}$$

To solve for an equilibrium in this setup use a guess and verify approach and conjecture that the solution for s_{t+1} is $s_t = a_1 \widehat{x}^{RE}_{t,t} + a_2 r_t$ with unknown coefficients a_1 and a_2. Then $E^{\tilde{P}}_t(s_{t+1}) = a_1 E^{\tilde{P}}_t(\widehat{x}^{RE}_{t+1,t+1}) + a_2 E^{\tilde{P}}_t(r_{t+1})$. Using the fact that the estimate at time $t+1$ is formed by the Kalman filter updating formulas, I get $E^{\tilde{P}}_t(s_{t+1}) = (a_1 + a_2)\rho \widehat{x}^{RE}_{t,t} + (a_1 K + a_2)\omega_{t+1}$. Replacing $E^{\tilde{P}}_t(s_{t+1})$ in (A.5) and taking the FOC with respect to ω_{t+1} I obtain

$$\omega_{t+1} = -\frac{Var^P_t(r_{t+1})}{\theta} b_t (a_1 K + a_2)$$

In equilibrium the same market clearing condition holds and $b_t = .5 s_t$. The FOC with respect to bonds requires $s_t = E^{\tilde{P}}_t(s_{t+1}) - r_t$. Substituting the solution for ω_{t+1} and rearranging the risk-neutral UIP condition becomes:

$$s_t = [1 + (a_1 K + a_2)^2 \frac{Var^P(r_{t+1})}{2\theta}]^{-1} [(a_1 K + a_2)\rho \widehat{x}^{RE}_{t,t} - r_t]$$

Using that the conditional variance of the exchange rate is $Var^P(s_{t+1}) = (a_1 K + a_2)^2 Var^P(r_{t+1})$, the following conditions are satisfied when verifying the guess for the conjecture about s_t :

$$a_2 = -[1 + \frac{Var^P(s_{t+1})}{2\theta}]^{-1} \quad \text{(A.6)}$$

$$a_1 = [1 + \frac{Var^P(s_{t+1})}{2\theta}]^{-1} (a_1 + a_2)\rho \quad \text{(A.7)}$$

This is exactly the solution described in (3.1) with $c = \frac{1}{\theta} Var(s_{t+1})$. With the market clearing condition of $b_t = .5 s_t$ the implied UIP regression coefficient has been shown in section 5 to be marginally smaller than 1 because such risk corrections are very small in this setup.

B Distorted expectations model equations

Proof of *Proposition 1:*

The estimate of the hidden state controls the expected interest rate differential. Its law of motion is driven by the Kalman filter as defined in (4.1):

$$E^{\tilde{P}}_t(r_{t+1}) = \rho \widehat{x}^t_{t,t} = \rho \widehat{x}^t_{t-1,t-1} + K^t_t (r_t - \rho \widehat{x}^t_{t-1,t-1}).$$

From (4.1):

$$\frac{\partial K_t^t}{\partial \sigma^2_{V,(t),t}} = -(\rho^2 \Sigma^t_{t-1,t-1} + \sigma^2_U)[(\rho^2 \Sigma^t_{t-1,t-1} + \sigma^2_U) + \sigma^2_{V,(t),t}]^{-2} < 0. \tag{B.1}$$

Thus

$$\frac{\partial \widehat{x}^t_{t,t}}{\partial \sigma_{V,(t),t}} = \frac{\partial \widehat{x}^t_{t,t}}{\partial \partial K_t^t} \frac{\partial K_t^t}{\partial \sigma^2_{V,(t),t}} = \frac{\partial K_t^t}{\partial \sigma^2_{V,(t),t}} (r_t - \rho \widehat{x}^t_{t-1,t-1}).$$

By the definition of end-of-life wealth, we have: $W_{t+1} = b_t[s_{t+1} - s_t - r_t]$.

Combining the various partial derivatives involved, I get the effect of $\sigma_{V,(t),t}$ on $E_t^{\tilde{P}} W_{t+1}$:

$$\begin{aligned} \frac{\partial E_t^{\tilde{P}} W_{t+1}}{\partial \sigma^2_{V,(t),t}} &= \frac{\partial E_t^{\tilde{P}} W_{t+1}}{\partial E_t^{\tilde{P}} s_{t+1}} \frac{\partial E_t^{\tilde{P}}(s_{t+1})}{\partial E_t^{\tilde{P}} r_{t+1}} \frac{\partial \widehat{x}^t_{t,t}}{\partial \sigma_{V,(t),t}} \\ sign\left[\frac{\partial E_t^{\tilde{P}} W_{t+1}}{\partial \sigma^2_{V,(t),t}}\right] &= sign\left[b_t \frac{\partial E_t^{\tilde{P}}(s_{t+1})}{\partial E_t^{\tilde{P}} r_{t+1}} \frac{\partial K_t^t}{\partial \sigma^2_{V,(t),t}} (r_t - F \widehat{x}^t_{t-1,t-1})\right]. \end{aligned} \tag{B.2}$$

Using Conjecture 1 that in equilibrium expected exchange rate is decreasing in the average interest rate differential, i.e. that $\frac{\partial E_t^{\tilde{P}}(s_{t+1})}{\partial E_t^{\tilde{P}} r_{t+1}} < 0$ and the fact that $\frac{\partial K_t^t}{\partial \sigma^2_{V,(t),t}} < 0$, we get a sharper prediction on the sign in (B.2):

$$sign\left[\frac{\partial E_t^{\tilde{P}} W_{t+1}}{\partial \sigma^2_{V,(t),t}}\right] = sign\left[b_t (r_t - F \widehat{x}^t_{t-1,t-1})\right].$$

This establishes *Proposition* 1.

Proof of *Remark 1:*

The recursion of the Kalman filter implies that:

$$\widehat{x}^t_{t,t} = (1 - K_t^t)\rho \widehat{x}^t_{t-1,t-1} + K_t^t r_t = K_t^t r_t + \sum_{j=1}^{t} \rho^j K^t_{t-j} r_{t-j} \left[\prod_{i=0}^{j-1} (1 - K^t_{t-i})\right].$$

Suppose that the first time going backwards in the sequence $t, ..., 0$ when the Kalman gain is equal to one is time m, i.e. $K_m^t = 1$ and $K_s^t < 1$ for all $t \geq s > m$. Then:

$$\begin{aligned} \widehat{x}^t_{t,t} &= K_t^t r_t + \sum_{j=1}^{t-m} \rho^j K^t_{t-j} r_{t-j} \left[\prod_{i=0}^{j-1} (1 - K^t_{t-i})\right] + \sum_{j=t-(m-1)}^{t} \rho^j K^t_{t-j} r_{t-j} \left[\prod_{i=0}^{j-1} (1 - K^t_{t-i})\right] \\ &= K_t^t r_t + \sum_{j=1}^{t-m} \rho^j K^t_{t-j} r_{t-j} \left[\prod_{i=0}^{j-1} (1 - K^t_{t-i})\right] \end{aligned}$$

since $1 - K_m^t = 0$ so that all the differentials r from 0 to $m-1$ receive zero weight in the estimate $\widehat{x}_{t,t}^t$. Intuitively, $K_m^t = 1$ means that the hidden state at time m was exactly equal to the observed differential r_m so there is no more information contained in the previous estimate $\widehat{x}_{m-1,m-1}^t$.

Now take the problem of the agent at time t of selecting the n dates at which to consider realizations of $\sigma_{V,(t),s}^2$ $s \leq t$ different from the constant $\sigma_V = 0$. Take m to be the first date going backwards from time t where the agent is considering that the realization $\sigma_{V,(t),m}^2$ equals $\sigma_V = 0$. Remark 1 states that $m = t - n$. According to the above formula for $\widehat{x}_{t,t}^t$ and using the same definition for m, any observation r_s with $0 < s < m$ receives no weight in the estimate of $\widehat{x}_{t,t}^t$, independent of the realization of $\sigma_{V,(t),s}^2$, $s < m$. Thus the minimization of $\widehat{x}_{t,t}^t$ over the n dates must involve choosing the dates n to be between m and t. This establishes *Remark 1.*

B.1 Discussion of Assumption 1

Discussion on *Remark 3*:

Suppose that the conjectured law of motion for s_{t+1} takes into account the presence of future ambiguous news. That is, take the conjectured law of motion from (4.5) and take expectations at time t :

$$\begin{aligned} E_t^{\widetilde{P}}(s_{t+1}) &= a_1 E_t^{\widetilde{P}}(\widehat{x}_{t+1,t+1}^{t+1}) + a_2 E_t^{\widetilde{P}}(r_{t+1}) \\ &= a_1 E_t^{\widetilde{P}} \rho \widehat{x}_{t,t}^{t+1} + E_t^{\widetilde{P}} K_{t+1}(r_{t+1} - \rho \widehat{x}_{t,t}^{t+1}) + a_2 \rho \widehat{x}_{t,t}^t. \end{aligned}$$

We have that

$$r_{t+1} \overset{\widetilde{P}}{\sim} N(\rho \widehat{x}_{t,t}^t, \rho^2 \Sigma_{t,t}^t + E_t^{\widetilde{P}}(\sigma_{V,t+1}^2) + \sigma_U^2)$$

where $E_t^{\widetilde{P}}(\sigma_{V,t+1}^2) = \sigma_V^2$ if the agent expects that the average realized $\sigma_{V,t+1}^2 = \sigma_V^2$ or $E_t^{\widetilde{P}}(\sigma_{V,t+1}^2) = (\sigma_V^H)^2$ if we take the approach as in Epstein and Schneider (2008) that the agent also minimizes over the future possible realized σ_{t+1}^V. The reason that $\sigma_V^H = \arg\min_{\sigma_{t+1}^V} E_t^{\widetilde{P}}(W_{t+1})$ is that the correction $E_t^{\widetilde{P}} K_{t+1}(r_{t+1} - \rho \widehat{x}_{t,t}^{t+1})$ will be larger in absolute value when $\sigma_{V,t+1} = \sigma_V^H$. Denote $n_{t+1} \equiv r_{t+1} - \rho \widehat{x}_{t,t}^{t+1}$. Then:

$$E_t^{\widetilde{P}} K_{t+1} n_{t+1} = E_t^{\widetilde{P}}[K_{t+1} n_{t+1} | n_{t+1} > 0] P(n_{t+1} > 0) + E_t^{\widetilde{P}}[K_{t+1} n_{t+1} | n_{t+1} < 0] P(n_{t+1} < 0) \tag{B.3}$$

and the RHS of (B.3) can be rewritten to take into account the position b_{t+1} as:

$$\begin{aligned} &E_t^{\widetilde{P}}[K_{t+1}^H n_{t+1}|n_{t+1}>0]P_t[n_{t+1}>0]\Pr_t(b_{t+1}<0)+\\ &+E_t^{\widetilde{P}}[K_{t+1}^H n_{t+1}|n_{t+1}<0]P_t[n_{t+1}<0]\Pr_t(b_{t+1}>0)+\\ &+E_t^{\widetilde{P}}[K_{t+1}^L n_{t+1}|n_{t+1}>0]P_t[n_{t+1}>0]\Pr_t(b_{t+1}>0)+\\ &+E_t^{\widetilde{P}}[K_{t+1}^L n_{t+1}|n_{t+1}<0]P_t[n_{t+1}<0]\Pr_t(b_{t+1}<0) \end{aligned} \tag{B.4}$$

where K_{t+1}^H, K_{t+1}^L is the Kalman gain for time $t+1$ with $\sigma_{V,t+1}^{t+1} = \sigma_V^H$ and $\sigma_{V,t+1}^{t+1} = \sigma_V^L$ respectively. P_t denotes the probability conditional on time t information. This asymmetric response to news is implied by the optimal choice over $\sigma_{V,t+1}^{t+1}$ in (4.6) of underreacting to positive innovations and overreacting to negative news, where "positive" and "negative" are defined with respect to the equilibrium position b_{t+1} as explained in Section 4.1.

Take the first moment of a truncated normal distribution where $x \sim N(\mu, \sigma)$. Then, with $\phi(.)$ and $\Phi(.)$ denoting the pdf and cdf of a unit normal:

$$E(x|x \leq a) = \mu - \sigma \frac{\phi\left(\frac{a-\mu}{\sigma}\right)}{\Phi\left(\frac{a-\mu}{\sigma}\right)}.$$

In general we want to compute $E(kx)$ where $k = k^H$ if $x \leq a$ and $k = k^L$ if $x > a$:

$$\begin{aligned} E(kx) &= k^L E(x|x \leq a)\Pr(x \leq a) + k^H E(x|x>a)(1-\Pr(x \leq a)) \\ E(kx) &= \mu\left[\Phi\left(\frac{a-\mu}{\sigma}\right)(k^L - k^H) + k^H\right] - \sigma\phi\left(\frac{a-\mu}{\sigma}\right)[k^L - k^H]. \end{aligned} \tag{B.5}$$

Equation (B.3) features a much more complicated version of (B.5) for several reasons. The object $\rho\widehat{x}_{t,t}^{t+1}$ is generated from the perspective of the time $t+1$ worst-case scenario and in general is not equal to $\rho\widehat{x}_{t,t}^{t}$. Its details are also a function of what the assumptions are on n, the number of past periods for which the agent distorts elements of the sequence of variances $\sigma_V(r^{t+1})$ as defined in formula (2.8). Related to this is the fact that $\Pr_t(b_{t+1} < 0)$ also appears in the evaluation of the expectation.

Such an object also appears in Epstein and Schneider (2008) except that their setup is significantly simplified since the agents always hold the asset in the same direction and the estimation is static. To replicate that situation one could take the approach of setting $n = t$, so the agent distorts all the previous periods and impose that conditional on having an investment direction at time t, the same direction is held next period, i.e. $\Pr(b_{t+1}b_t > 0) = 1$. Also take K_{t+1}^H and K_{t+1}^L to be equal to the static Kalman gain: $\frac{\sigma_U^2}{\sigma_U^2 + \sigma_V^{2,(i)}}$ for $i = H, L$

respectively. Then (B.3) becomes:

$$\begin{aligned} E_t^{\widetilde{P}} K_{t+1} n_{t+1} &= sign(-b_t)\{E_t^{\widetilde{P}}[K^H n_{t+1} | n_{t+1} > 0] \Pr(n_{t+1} > 0) + \\ &\quad + E_t^{\widetilde{P}}[K^L n_{t+1} | n_{t+1} < 0] \Pr[n_{t+1} < 0]\} \end{aligned}$$

and using (B.5) with $n_{t+1} \sim N(0, \sigma^2)$ and $\sigma = [\Sigma + (\sigma_V^H)^2 + \sigma_U^2]^{0.5}$:

$$E_t^{\widetilde{P}} K_{t+1} n_{t+1} = sign(b_t)\sigma \frac{1}{\sqrt{2\pi}}[K^L - K^H]. \tag{B.6}$$

Then the expected s_{t+1} should be:

$$E_t^{\widetilde{P}}(s_{t+1}) = (a_1 + a_2)\rho \widehat{x}_{t,t}^t + a_1 sign(b_t)\sigma \frac{1}{\sqrt{2\pi}}[K^L - K^H].$$

Equation (B.5) shows the type of correction in the expectation of future s_{t+1} that is implied by taking into account the asymmetric response to future news. Take for example the case in which $b_t < 0$ so the agent invests at time t in the domestic currency. Then, because $a_1 < 0$ the expected s_{t+1} is larger by $|a_1|\delta$, with $\delta = \sigma \frac{1}{\sqrt{2\pi}}[K^L - K^H] > 0$. Thus, there is a higher expected depreciation and

$$s_t = E_t^{\widetilde{P}}(s_{t+1}) - r_t = (a_1 + a_2)\rho \widehat{x}_{t,t}^t + |a_1|\delta.$$

If the conjectured law of motion takes this constant into account as:

$$E_t^{\widetilde{P}}(s_{t+1}) = a_1 E_t^{\widetilde{P}}(\widehat{x}_{t+1,t+1}^{t+1}) + a_2 E_t^{\widetilde{P}}(r_{t+1}) + |a_1|\delta$$

then by the equilibrium UIP equation:

$$s_t = E_t^{\widetilde{P}}(s_{t+1}) - r_t = (a_1 + a_2)\rho \widehat{x}_{t,t}^t + 2|a_1|\delta$$

and there is no conjectured law of motion that in equilibrium implies consistency of beliefs and equates s_t with $E_t^{\widetilde{P}}(s_{t+1}) - r_t$ because there is no discounting of the constant δ in the asset pricing equation, in contrast with the model of Epstein and Schneider (2008). This concludes the discussion on *Remark 3*.

Proof of *Remark 4*:

Forming expectations according to the conjectured law of motion in (4.5):

$$E_t^{\widetilde{P}}(s_{t+1}) = a_1 E_t^{\widetilde{P}}(\widehat{x}_{t+1,t+1}^{t+1}) + a_2 E_t^{\widetilde{P}}(r_{t+1}).$$

Under the true DGP, the average realized s_{t+1} is:

$$\begin{aligned} E_t^P s_{t+1} &= a_1 E_t^P(\widehat{x}_{t+1,t+1}^{t+1}) + a_2 E_t^P(r_{t+1}) \\ E_t^P s_{t+1} &= a_1[E_t^P \rho \widehat{x}_{t,t}^{t+1} + E_t^P K_{t+1}(r_{t+1} - \rho \widehat{x}_{t,t}^{t+1})] + a_2 \rho \widehat{x}_{t,t}^{RE}. \end{aligned} \tag{B.7}$$

Take for example $b_t < 0$ so that $\widehat{x}_{t,t}^{t} < \widehat{x}_{t,t}^{RE}$. The average $E_t^P s_{t+1}$ is different from $E_t^{\widetilde{P}}(s_{t+1})$ due to two factors: First, $E_t^{\widetilde{P}}(r_{t+1}) < E_t^P(r_{t+1})$, which activates a channel that allows for a possible ex-post average appreciation, i.e. $E_t^P s_{t+1} < E_t^{\widetilde{P}}(s_{t+1})$. Second, the asymmetric response to signals next period when agents overreact to negative innovations and underreact to positive innovations is implied by $E_t^P K_{t+1}(r_{t+1} - \rho \widehat{x}_{t,t}^{t+1}) < K E_t^P(r_{t+1} - \rho \widehat{x}_{t,t}^{t+1})$, creating a channel that works against the average ex-post appreciation.

As long as the overall effect is that $E_t^P s_{t+1} < E_t^{\widetilde{P}}(s_{t+1})$ the following argument can be made: suppose the agent at time t observes the average $E_t^P s_{t+1}$ and conjectures that this is indeed generated by the law of motion in (4.5). Then,

$$E_t^P(s_{t+1}) = (a_1 + a_2) E_t^{\overline{P}} r_{t+1} < E_t^{\widetilde{P}}(s_{t+1}) = (a_1 + a_2) E_t^{\widetilde{P}} r_{t+1}$$

where $E_t^{\overline{P}}$ is formed under what is perceived by the agent at time t as possibly having generated the true data. However, we have from (B.7) that for the true $E_t^P s_{t+1}$:

$$E_t^P s_{t+1} > (a_1 + a_2) E_t^P r_{t+1}$$

because of the average asymmetric response. Then, because $(a_1 + a_2) < 0$:

$$E_t^{\widetilde{P}} r_{t+1} < E_t^{\overline{P}} r_{t+1} < E_t^P r_{t+1}. \tag{B.8}$$

By construction, the agent does not know the true DGP, so after observing $E_t^P s_{t+1} < E_t^{\widetilde{P}}(s_{t+1})$ and using the conjecture that s_{t+1} is controlled by the law of motion the agent would conclude that s_{t+1} has depreciated less ex-post because her expected $r_{t+1}, E_t^{\widetilde{P}} r_{t+1}$, was lower than the possible true DGP, i.e. $E_t^{\widetilde{P}} r_{t+1} < E_t^{\overline{P}} r_{t+1}$. This is consistent with the fact that the agent at time t was acting upon the conjectured law of motion and under a sequence of variances as in (2.2) that was minimizing expected payoffs:

$$\min_{\widetilde{P} \in \Lambda} E_t^{\widetilde{P}}(r_{t+1}) \tag{B.9}$$

Thus, if the set Λ contains the distributions implied by $\overline{P}$ and P, then the observed $E_t^P s_{t+1} < E_t^{\widetilde{P}}(s_{t+1})$ is consistent with the minimization in (B.9) and the conjectured law of motion as

in (4.5).

The key element in this argument is that the first channel dominates so that $E_t^P s_{t+1} < E_t^{\widetilde{P}}(s_{t+1})$. If, for example, there are no dynamics coming from the underestimation of the hidden state at time t then the asymmetric response to news at time $t+1$ will generate an average higher than expected depreciation, i.e. $E_t^P s_{t+1} > E_t^{\widetilde{P}}(s_{t+1})$, which will imply that $E_t^{\widetilde{P}} r_{t+1} > E_t^{\overline{P}} r_{t+1}$ being inconsistent with the minimization in (B.9). A similar argument applies for the case of $b_t > 0$ where the inequalities above should be reversed. This establishes *Remark 4*.

B.2 A general numerical solution procedure for the ambiguity aversion model

Under the conjectured law of motion and Assumption 1, expected exchange rate is given by (4.11). Proposition 2 stated for the special case of $\sigma_V = 0$ then $s_t = a_1 \widehat{x}_{t,t}^t + a_2 r_t$, where a_1, a_2 are the same analytical coefficients as in equations (3.7) and (3.7) respectively, that characterize the rational expectations case. For the case $\sigma_V > 0$ a more general numerical procedure is required to recover the coefficients a_1, a_2.

The solution to the ambiguity aversion equilibrium can be summarized by the following steps:

1. Start with an initial guess about a_1, a_2.

2. For each t, make a guess about the sign of b_t to use in (4.6).

3. Use (4.6) and call the resulting optimal sequence $\sigma_V^*(r^t)$. Use the Kalman filter based on the sequence $\sigma_V^*(r^t)$ to form an estimate for $\widehat{x}_{t,t}^t$ and $\Sigma_{t,t}^t$.

4. Draw realizations for r_{t+1} from $N(\rho \widehat{x}_{t,t}^t, \rho^2 \Sigma_{t,t}^t + \sigma_U^2)$, where $\Sigma_{t,t}^t$ is defined in (4.2). Form the sample $r^{t+1} = (r^t, r_{t+1})$. For each realization perform Steps 2 and 3 above to obtain the sequence $\sigma_V^*(r^{t+1})$.

5. For each realization in step 4 use $\sigma_V^*(r^{t+1})$ to compute $\widehat{x}_{t+1,t+1}^{t+1}$ and use the conjecture in (4.5) to generate a realized $s_{t+1} = a_1 \widehat{x}_{t+1,t+1}^{t+1} + a_2 r_{t+1}$.

6. The distribution of s_{t+1} in step 5 defines the subjective probability distribution for the agent at time t. Use the FOC (4.10) to solve for s_t^*.

7. If $sign(s_t^*) = sign(b_t)$ the solution is $\sigma_V^*(r^t)$ and s_t^* and an indicator function $i_t = 1$. If $sign(s_t^*) \neq sign(b_t)$, switch the sign of the initial guess in step 2.

8. If there is no convergence on the sign of s_t^* and b_t, the solution is $b_t^* = s_t^* = 0$ and the indicator function $i_t = 0$.

9. Regress s_t^* on $\widetilde{a}_1 \widehat{x}_{t,t}^t$ and $\widetilde{a}_2 r_t$ for all the t when $i_t = 1$. If $\min_i |a_i - \widetilde{a}_i| > \varepsilon$, then reiterate from step 1 with $a_i = \widetilde{a}_i$. If not then stop and the minimizing coefficients are $\widetilde{a}_i$.

Point 8 of this iteration is related to the discussion in 4.2 which describes how the bond solution to the problem in (4.7) can be the kinked solution $b_t = 0$.

B.3 Risk aversion and distorted expectations model equations

Proof of *Proposition 3*:

The variance $Var_t^{\widetilde{P}} W_{t+1} = b_t^2 Var_t^{\widetilde{P}} s_{t+1}$. In turn, using the conjecture (4.5), Assumption 1 and taking as given $\widehat{x}_{t,t}^{t+1}, K_{t+1}^{t+1}$:

$$Var_t^{\widetilde{P}} s_{t+1} = (a_1 K + a_2)^2 [\rho^2 \Sigma_{t,t}^t + \sigma_U^2].$$

Use the formula in (4.5) for the Kalman gain and the recursion $\Sigma_{t,t}^t = \Sigma_{t,t-1}^t (1 - K_t)$. Then,

$$\frac{\partial \Sigma_{t,t}}{\partial \sigma_{V,(t),t}^2} = \frac{\partial \Sigma_{t,t}}{\partial K_t} \frac{\partial K_t^t}{\partial \sigma_{V,(t),t}^2} > 0.$$

Thus,

$$\frac{\partial Var_t^{\widetilde{P}} W_{t+1}}{\partial \sigma_{V,(t),t}^2} > 0.$$

This establishes *Proposition* 3.

To study the effect of $\sigma_{V,(t),t}^2$ on the utility consider the total partial derivative:

$$\frac{\partial V_t}{\partial \sigma_{V,(t),t}^2} = \frac{\partial V_t}{\partial E_t^{\widetilde{P}} s_{t+1}} \frac{\partial E_t^{\widetilde{P}} s_{t+1}}{\partial E_t^{\widetilde{P}} r_{t+1}} \frac{\partial E_t^{\widetilde{P}} r_{t+1}}{\partial \sigma_{V,(t),t}^2} + \frac{\partial V_t}{\partial Var_t^{\widetilde{P}} s_{t+1}} \frac{\partial Var_t^{\widetilde{P}} s_{t+1}}{\partial Var_t^{\widetilde{P}} r_{t+1}} \frac{\partial Var_t^{\widetilde{P}} r_{t+1}}{\partial \sigma_{V,(t),t}^2}.$$

The sign of this derivative is :

$$sign(\frac{\partial V_t}{\partial \sigma_{V,(t),t}^2}) = sign(b_t) sign(r_t - \rho \widehat{x}_{t-1,t-1}^t) - sign(\frac{\partial Var_t^{\widetilde{P}} r_{t+1}}{\partial \sigma_{V,(t),t}^2}).$$

Since $sign(\frac{\partial Var_t^{\widetilde{P}} r_{t+1}}{\partial \sigma_{V,(t),t}^2}) > 0$, if the sign of $sign(b_t) sign(r_t - \rho \widehat{x}_{t-1,t-1}^t)$ is also positive then the sign of $\frac{\partial V_t}{\partial \sigma_{V,(t),t}^2}$ is ambiguous. To study that case, compute

$$\begin{aligned} \frac{\partial V_t}{\partial \sigma_{V,(t),t}^2} &= -(a_1 K + a_2) \rho b_t (r_t - \rho \widehat{x}_{t-1,t-1}^t)(\rho^2 \Sigma_{t-1,t-1}^t + \sigma_U^2)[\rho^2 \Sigma_{t-1,t-1}^t + \sigma_U^2 + \sigma_{V,(t),t}^2]^{-2} \\ &\quad + (1-\gamma)(a_1 K + a_2)^2 b_t^2 \rho^2 [\rho^2 \Sigma_{t-1,t-1}^t + \sigma_U^2]^2 [\rho^2 \Sigma_{t-1,t-1}^t + \sigma_U^2 + \sigma_{V,(t),t}^2]^{-2}. \end{aligned}$$

By the filtering solution $(r_t - \rho \widehat{x}_{t-1,t-1}^t) = (\rho^2 \Sigma_{t-1,t-1}^t + \sigma_U^2 + \sigma_{V,(t),t}^2)^{0.5} \xi_t$, where $\xi_t \sim N(0,1)$.

To investigate the ambiguous case compute $P\left[(\partial V_t/\partial\sigma^2_{V,(t),t}) > 0|(b_t\xi_t > 0)\right]$ which equals

$$P\{b_t\xi_t > (1-\gamma)(a_1K+a_2)\rho b_t^2 \frac{(\rho^2\Sigma^t_{t-1,t-1}+\sigma^2_U)}{\left[\rho^2\Sigma^t_{t-1,t-1}+\sigma^2_U+\sigma^2_{V,(t),t}\right]^{0.5}}|(b_t\xi_t>0)\} \tag{B.10}$$

To get an upper bound on the probability take the case of $K=1, \Sigma^t_{t-1,t-1}=0$ and $\sigma^2_{V,(t),t}=0$. In this case, if $b_t > 0$ then (B.10) becomes:

$$\Pr[\xi_t > (1-\gamma)(a_1+a_2)\rho b_t\sigma_U|\xi_t>0].$$

Adding to the benchmark parameterization $\gamma = 10$ and noting that $b_t = 0.5s_t$ with the model-implied standard deviation of s_t around 0.025 the probability that the expected return channel dominates is close to one. A similar calculation applies for $b_t < 0$.

C Supplementary tables

Table 10: ML estimates of a state space representation with constant volatilities

	Austria†	Belg.†	Canada	France†	Germ.†	Italy†	Japan	Neth.†	Switz.	UK
ρ	0.99	0.96	0.99	0.99	0.98	0.98	0.98	0.99	0.98	0.99
	(0.007)	(0.016)	(0.004)	(0.007)	(0.007)	(0.015)	(0.014)	(0.008)	(0.011)	(0.008)
σ_V	1.75	0.33	0.87	0.001	0.12	0.0012	3.8415	0.0001	3.93	0.0001
	(0.72)	(3.72)	(0.49)	(2.2)	(8.38)	(2.36)	(0.69)	(0.56)	(0.86)	(0.68)
σ_U	5.92	5.75	4.04	5.28	5.92	5.69	5.75	3.67	7.82	3.71
	(0.47)	(0.49)	(0.26)	(0.24)	(0.42)	(0.323)	(0.71)	(0.21)	(0.82)	(0.21)

The state-space representation is described in (2.1). The sample is M1 1981-M12 2007 except for countries † for which data ends in M12:1998. The entries in the columns for the standard deviations σ_V, σ_U are reported as the estimated values $\times 1000$. Standard errors are in parentheses.

Table 11: Empirical UIP regression and carry trade payoffs

	UIP regression		Carry trade payoffs				
Country	$\widehat{\alpha}$	$\widehat{\beta}$	Mean	Standard Deviation	Sharpe Ratio	Skewness	Excess Kurtosis
Austria†	0.003	-1.01	0.0022	0.034	0.064	-0.145	0.785
	(0.002)	(0.73)	(0.0022)	(0.002)	(0.067)	(0.175)	(0.397)
Belgium†	0.000	-0.66	0.0069	0.033	0.208	-0.0003	0.904
	(0.002)	(0.64)	(0.0020)	(0.002)	(0.062)	(0.229)	(0.431)
Canada	0.000	-0.6	0.0019	0.016	0.115	-0.501	1.273
	(0.001)	(0.5)	(0.0009)	(0.001)	(0.055)	(0.157)	(0.46)
Denmark	0.000	-0.63	0.0083	0.030	0.276	-0.128	0.838
	(0.002)	(0.47)	(0.0017)	(0.001)	(0.059)	(0.140)	(0.405)
France†	0.000	0.06	0.0053	0.032	0.168	-0.044	0.437
	(0.003)	(0.71)	(0.0020)	(0.002)	(0.062)	(0.154)	(0.311)
Germany†	0.003	-0.66	0.0012	0.034	0.035	-0.183	0.475
	(0.002)	(0.83)	(0.0022)	(0.002)	(0.065)	(0.127)	(0.313)
Ireland†	0.000	0.38	0.0053	0.032	0.165	-0.021	0.354
	(0.003)	(0.98)	(0.0023)	(0.002)	(0.072)	(0.178)	(0.375)
Italy†	-0.001	0.26	0.0027	0.030	0.091	-0.332	1.116
	(0.003)	(0.4)	(0.0021)	(0.002)	(0.07)	(0.242)	(0.564)
Japan‡	0.01	-2.55	0.0028	0.035	0.08	-0.67	1.706
	(0.003)	(0.69)	(0.0020)	(0.002)	(0.059)	(0.249)	(0.9)
Netherlands†	0.004	-1.68	0.0035	0.034	0.103	-0.122	0.625
	(0.002)	(0.81)	(0.0023)	(0.002)	(0.068)	(0.209)	(0.391)
Norway	0.000	-0.51	0.0052	0.029	0.183	-0.191	1.121
	(0.002)	(0.5)	(0.0014)	(0.001)	(0.05)	(0.1720	(0.434)
Portugal†	-0.002	0.45	0.0042	0.032	0.131	-0.071	2.385
	(0.003)	(0.25)	(0.0021)	(0.002)	(0.065)	(0.379)	(0.972)
Spain†	0.002	0.75	0.0032	0.032	0.102	-0.723	2.081
	(0.003)	(0.52)	(0.0024)	(0.002)	(0.076)	(0.350	(1.429)
Sweden	0.000	0.36	0.0059	0.030	0.199	-0.78	3.229
	(0.002)	(0.69)	(0.0015)	(0.002)	(0.058)	(0.35)	(1.476)
Switzerland	0.007	-1.40	0.0009	0.035	0.025	-0.23	0.668
	(0.003)	(0.68)	(0.0020)	(0.002)	(0.056)	(0.204)	(0.447)
UK	-0.002	-1.67	0.0057	0.030	0.191	-0.028	2.602
	(0.002)	(0.85)	(0.0015)	(0.002)	(0.050)	(0.374)	(0.99)
Average	0.001	-0.57	0.0041	0.031	0.133	-0.261	1.254

Notes: The first 2 columns report estimates of the regression:$S_{t+1}/S_t - 1 = \alpha + \beta(F_t/S_t - 1) + \varepsilon_{t+1}$. Both F_t and S_t are USD/FCU. Heteroskedasticity-robust standard errors are in parentheses. Carry trade payoffs are measured in USD, per dollar bet. The sample of monthly data is M1:1976 to M7:2008, except for countries (†) for which data ends in M12:1998 and Japan for which data begins on M7:1978.

Table 12: Standard and modified carry trade payoffs

	Standard carry trade payoffs			Modified carry trade payoffs		
Country	Mean	Standard Deviation	Sharpe Ratio	Mean	Standard Deviation	Sharpe Ratio
Austria†	0.0022	0.034	0.064	0.0043	0.033	0.129
	(0.0022)	(0.002)	(0.067)	(0.0026)	(0.002)	(0.077)
Belgium†	0.0069	0.033	0.208	0.0082	0.033	0.246
	(0.0020)	(0.002)	(0.062)	(0.0024)	(0.002)	(0.069)
Canada	0.0019	0.016	0.115	0.0033	0.015	0.215
	(0.0009)	(0.001)	(0.055)	(0.0009)	(0.001)	(0.061)
Denmark	0.0083	0.030	0.276	0.0088	0.030	0.289
	(0.0017)	(0.001)	(0.059)	(0.0020)	(0.002)	(0.066)
France†	0.0053	0.032	0.168	0.0071	0.032	0.219
	(0.0020)	(0.002)	(0.062)	(0.0022)	(0.002)	(0.069)
Germany†	0.0012	0.034	0.035	0.0030	0.034	0.089
	(0.0022)	(0.002)	(0.065)	(0.0027)	(0.002)	(0.078)
Ireland†	0.0053	0.032	0.165	0.0058	0.031	0.185
	(0.0023)	(0.002)	(0.072)	(0.0025)	(0.002)	(0.080)
Italy†	0.0027	0.030	0.091	0.0040	0.030	0.131
	(0.0021)	(0.002)	(0.07)	(0.0022)	(0.002)	(0.077)
Japan‡	0.0028	0.035	0.08	0.0017	0.035	0.048
	(0.0020)	(0.002)	(0.059)	(0.0024)	(0.003)	(0.070)
Netherlands†	0.0035	0.034	0.103	0.0051	0.033	0.154
	(0.0023)	(0.002)	(0.068)	(0.00260	(0.002)	(0.078)
Norway	0.0052	0.029	0.183	0.0067	0.029	0.230
	(0.0014)	(0.001)	(0.05)	(0.0018)	(0.002)	(0.063)
Portugal†	0.0042	0.032	0.131	0.0048	0.031	0.155
	(0.0021)	(0.002)	(0.065)	(0.0022)	(0.002)	(0.069)
Spain†	0.0032	0.032	0.102	0.0046	0.033	0.139
	(0.0024)	(0.002)	(0.076)	(0.0024)	(0.002)	(0.074)
Sweden	0.0059	0.030	0.199	0.0064	0.029	0.217
	(0.0015)	(0.002)	(0.058)	(0.00150	(0.002)	(0.055)
Switzerland	0.0009	0.035	0.025	0.0030	0.035	0.086
	(0.0020)	(0.002)	(0.056)	(0.0024)	(0.002)	(0.069)
UK	0.0057	0.030	0.191	0.0062	0.031	0.200
	(0.0015)	(0.002)	(0.050)	(0.0021)	(0.002)	(0.068)
Average	0.0041	0.031	0.133	0.0052	0.031	0.171

Notes: Similar notes as in Table 11 apply for the data. The standard carry trade payoffs are the same as in Table 11. The modified carry trade payoffs are obtained based on the strategy defined in (5.8), where μ=0.