

Ambiguity Aversion: Implications for the Uncovered Interest Rate Parity Puzzle (Job Market Paper)*

COSMIN L. ILUT[†]

December 2008

Abstract

A positive domestic-foreign interest rate differential predicts that the domestic currency will appreciate in the future. So, on average, positive profits can be made by borrowing in low-interest-rate currencies and lending in high-interest-rate currencies (a strategy known as the ‘carry trade’). Standard theory implies that capital inflows into high-interest-rate currencies should be so large that the positive profits from the carry trade are wiped out. The absence of inflows of such a magnitude is one way to characterize the well-known uncovered interest rate parity (UIP) puzzle. A standard, though controversial, resolution of the puzzle is that what limits capital inflows when domestic interest rates are high is an objective increase in risk in the domestic currency. This explanation has been challenged on the grounds that it is difficult to empirically detect this risk. The alternative explanation I pursue is that agent’s 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 agents’ distorted beliefs are derived formally from the assumption that they are ambiguity averse. In my model, agents do not know key parameters of the stochastic process driving the variables they forecast. In the presence of parameter uncertainty, ambiguity-averse agents compute forecasts using values of the parameters that (i) produce bad outcomes for the agent and (ii) are not implausible in a likelihood sense. The equilibrium of my model resembles the data in that a positive domestic interest rate differential predicts that the domestic currency will appreciate in the future. The reason capital inflows into high-interest-rate currencies are limited in the model is that agents tend to overstate the probability of a future depreciation. I show that my result cannot be duplicated in a simple model with risk aversion. In addition to providing a resolution to the UIP puzzle, the model predicts, consistent with the data, negative skewness and excess kurtosis for carry trade payoffs and positive average payoffs even for hedged positions. *JEL Classification*: D8, E4, F3, G1.

*I am grateful to the members of my dissertation committee Lawrence Christiano (chair), Martin Eichenbaum, Giorgio Primiceri and Sergio Rebelo for their continuous support and advice. I would also like to thank Peter Benczur, Eddie Dekel, Lars Hansen, Nicolas Lehmann-Ziebarth, Ricardo Massolo, Jonathan Parker, Tom Sargent, Tomasz Strzalecki and seminar participants at the Chicago Fed and Northwestern University for helpful discussions and comments. Financial support from the Center for International Economics and Development is greatly appreciated.

[†]Department of Economics, Northwestern University. Contact: cosminilut2009@u.northwestern.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 1. This implication is strongly counterfactual. In practice, UIP regressions (Fama (1984), Hansen and Hodrick (1980)) produce coefficient estimates well below 1 and sometimes even negative.¹ This anomaly is taken very seriously because the UIP equation is a property of most open economy models. The failure, referred to as the UIP puzzle or the forward premium puzzle², 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. The rational expectations assumption implies that agents are endowed with perfect knowledge about the true data generating process (DGP). The time-varying risk-premia interpretation views the positive profits from the carry trade strategy as a compensation for risk. This approach to the UIP puzzle has been criticized in two ways: survey evidence has been used to cast doubt on the rational expectations assumption³ 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 the complete knowledge of the true DGP and that they confront this uncertainty with ambiguity aversion.⁶ I model ambiguity aversion along the lines of the maxmin expected utility (or multiple priors) preferences as in Gilboa and Schmeidler (1989).

The model has several types of agents. The decision problem of a subset of the agents (I call them

¹Among recent studies see Burnside et al. (2008), Chinn and Frankel (2002), Chinn and Meredith (2005), Flood and Rose (1996), Gourinchas and Tornell (2004), Mark and Wu (1998), Sarno (2005) and Verdelhan (2006).

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

³For example, Froot and Frankel (1989), Chinn and Frankel (2002) and Bacchetta et al. (2008) decompose predictable excess returns into their currency risk premium and expectational error components. They find that almost all of the returns can be attributed to the latter.

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

⁵Eichenbaum and Evans (1995), Froot and Thaler (1990), Gourinchas and Tornell (2004) and Lyons (2001) argue that models where agents are slow to respond to news may explain the UIP puzzle. Bacchetta and van Wincoop (2008) offer a formalization of such a mechanism based on rational inattention. For details on the theoretical literature on distorted expectations and UIP see Appendix A.

⁶Lars Hansen and Thomas Sargent have introduced ambiguity aversion in the form of robust control in the macro-finance literature. For a thorough treatment of robust control see Hansen and Sargent (2008). For applications of robust control to exchange rate puzzles see Benigno (2007), Li and Tornell (2007) and Tornell (2003).

‘agents’) is modeled explicitly, and the behavior of the others (‘liquidity traders’) is taken as given. The supply of domestic and foreign bonds is fixed in domestic and foreign currency units, respectively. The liquidity traders adjust their demand for bonds to satisfy the market clearing condition. Agents are all identical and live for two periods.⁷ The representative agent begins the first period with no endowment. She buys and sells bonds in different currencies in order to maximize a negative exponential utility function of second period wealth. The problem of the agent is complicated by the fact that she is uncertain about a subset of the parameters in her environment. 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. As a result, the agent must solve a signal extraction problem when she wants to adjust her forecasts in response to a disturbance.

I follow and extend the setup in Epstein and Schneider (2007, 2008) by assuming that the agent does not know the variances of the innovations in the temporary and persistent components and she allows for the possibility that those variances change over time. As a result, the decision problem of the agent requires taking a stand on the parameter values of the model, as well as choosing a quantity of bonds to buy and sell. 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.

The main result of this paper is that ambiguity aversion has the potential to resolve the UIP puzzle. For the benchmark calibration, a higher domestic interest rate differential predicts a future appreciation of the domestic currency. Numerical simulations show that in large samples the UIP regression coefficient is negative and statistically significant. The key intuition for this can be explained by analyzing an impulse response experiment. In this experiment the economy starts from the steady state in which the interest rate differential equals to zero. At some date t there is an observed increase in the domestic interest rate and there are no other shocks following period t . Suppose that this increase is the result of a probabilistic combination of a persistent and temporary shocks whose magnitudes satisfy the signal to noise ratio from the true DGP. First, consider the case of rational expectations. The agent is then endowed with the perfect knowledge of this signal to noise ratio and correctly estimates the hidden persistent state and forecasts the future exchange rate. The rational expectations solutions would imply, as in the Dornbusch (1976) model of overshooting, a demand of domestic bonds that is large enough to create an immediate appreciation at date t and then a path of depreciation from period $t + 1$ onwards.

In contrast to the rational expectations assumption, in my model the agent is not certain about the probability that the increase in period t is the result of a temporary or a persistent shock. In equilibrium

⁷The agents in my model resemble those in Bacchetta and van Wincoop (2008), except that there they investigate rational inattention and I assume ambiguity aversion.

the agent borrows in the foreign interest rate and lends in the higher domestic interest rate. Next period the agent repays her debt by converting the investment proceeds back into foreign currency. From the investor's perspective, the worse-case scenarios are those in which the domestic currency depreciates. The agent uses the correct equilibrium relations and realizes that a large domestic depreciation occurs next period if the true shock is a temporary one. This happens because if the true shock is indeed temporary, next period the agent will demand less of the domestic bond. Hence, the concern about a future depreciation translates into the agent being worried that the true shock is temporary. The ambiguity averse agent places then more probability, compared to the true DGP, on the observed increase in the differential to be caused by the temporary shock. Under this belief, the agent at time t will demand less of domestic bonds than under the rational expectations rule.

Because the agent underestimates the hidden state by placing too much probability on the temporary shock compared to the true DGP, in period $t + 1$ she will perceive on average positive innovations about the hidden persistent state. Under her subjective beliefs these innovations are unexpected good news that at time $t + 1$ increase the estimate of the hidden state compared to the rational expectations case. This updating effect creates the possibility that in period $t + 1$ the agent finds it optimal to invest even more in the domestic bond because the higher estimate raises the present value of future payoffs of investing in the domestic bond. The increased demand will drive up the value of the domestic currency contributing to an appreciation between period t and $t + 1$. This type of impulse response corresponds to the empirically documented delayed overshooting puzzle⁸ in which following a positive shock to the domestic interest rate the domestic currency experiences a gradual appreciation for several periods instead of a path of depreciation as the UIP under rational expectations implies.⁹

More generally, when the agent is considering investing in the higher domestic interest rate she is concerned about a domestic currency depreciation and consequently about the estimate of the hidden state being low. Under this worst-case scenario she will choose to believe that it is more likely that the observed increases in the domestic differential have been generated by temporary shocks (low precision of signals) and decreases by persistent shocks (high precision of signals). In this sense, she underreacts to good news and overreacts to bad news. The intuition for the impulse response carries over when simulating the model with temporary and persistent shocks drawn every period from the true DGP.

The above intuition also applies for the case when the foreign interest rate is higher than the domestic rate. The agent then invests in the foreign bond and is concerned about a foreign currency depreciation. In that case, good news is an increase in the foreign rate. The same mechanism delivers the possibility of a gradual appreciation of the foreign currency subsequent to a positive shock to the foreign rate of interest. This highlights that the worst-case scenario is a function of the investment position.

⁸See Eichenbaum and Evans (1995), Faust and Rogers (2003), Grilli and Roubini (1996) and Scholl and Uhlig (2006).

⁹This intuition is related to Gourinchas and Tornell (2004) who show that if, for some unspecified reason, the agent systematically underreacts to signals about the time-varying hidden-state of the interest rate differential this can explain the UIP and delayed overshooting puzzle. The main difference from their paper is that I formally investigate the optimality of such distorted beliefs.

The explanation for the UIP puzzle proposed in this paper relies on placing some 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.¹⁰ 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.¹¹

The model is calibrated to data for eight developed countries which suggests a high degree of persistence of the hidden state and a relatively 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. 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. Studying other calibrations, I find that the UIP regression coefficient becomes positive, even though smaller than 1, if the true DGP is characterized by a significantly less persistent hidden state or much larger temporary shocks than the benchmark specification.

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 to 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. The underprediction is related to the dynamic interaction of the distorted beliefs and the true DGP that produces in equilibrium a more frequent occurrence of exchange rate realizations that benefit ex-post the carry trade strategy. This is in contrast with models that rely on peso events in which the agent overpredicts a large bad state that does not happen in a small sample. The peso problem interpretation views the positive ex-post payoffs as a manifestation of the lack of occurrence of this rare event.¹²

¹⁰The intuition still holds if the time-varying unobserved state is not the persistent component but the autocorrelation coefficient of the observed interest rate differential. In that case, the agent is concerned that this coefficient is low and interprets signals that imply a low estimate as reflecting a shock to the time-varying parameter and signals that imply a high estimate as temporary shocks to the differential. See Appendix F for details on the time-varying parameter case.

¹¹This is in contrast to unstructured uncertainty, for which, as shown for example in Strzalecki (2007), Barillas et al. (2008), the multiplier preferences used in Hansen and Sargent (2008) are equivalent to a higher risk aversion expected utility. The unstructured uncertainty places no restriction on the nature or location of possible misspecifications, except that the “distance” from the reference model is bounded by some cost function (in their model the relative entropy). A similar argument regarding unstructured uncertainty and robust filtering is made in Li and Tornell (2007). There is a literature on optimal monetary policy under structured uncertainty including, for example, Brock et al. (2007), Cogley and Sargent (2005), Giannoni (2007), Levin and Williams (2003) and Woodford (2006).

¹²See Farhi and Gabaix (2008) for an example of a rare-disaster explanation of the UIP puzzle. Their model is a theory of time-varying risk premia generated by rare events which can occur in small samples. For models addressing the UIP failure that rely on the typical interpretation of the peso event, i.e. one which does not happen in the sample, see Engel and Hamilton (1989), Evans and Lewis (1995) and Lewis (1989). A more detailed review is presented in Appendix A.

Second, in the model hedged positions can deliver positive mean payoffs. This is an important result because a recent empirical literature finds that even when the carry trade's downside risk is eliminated by using options, the hedged strategy generates positive payoffs.¹³ The difficulty in generating this result is related to the intuition that buying insurance against the downside risk produces on average negative payoffs that decrease the payoff of the hedged strategy. My model also implies this type of loss because of the overprediction of bad events. However, in my model this negative payoff does not completely offset the positive payoff of the unhedged carry trade. The reason is related to the more frequent occurrence of good states for the carry trade strategy under the objective probability distribution than under the equilibrium distorted beliefs. This is contrast to models in which peso events are associated with large losses from the carry trade strategy that do not occur in the sample but otherwise, for the non-peso events, the subjective and the probability distributions coincide. In those models, buying insurance eliminates the gains from the unhedged strategy.¹⁴The theory presented in this paper is also consistent with recent empirical findings about the conditional time-variation of risk-neutral moments for currency trading.¹⁵

Third, 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.¹⁶ In my model, the gradual appreciation arises from the slow incorporation of good news about the high interest rate currency. However, when this interest rate decreases compared to the market's expectation, it produces a relatively sudden depreciation because agents respond quickly to this type of news. The excess kurtosis is a manifestation of the 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 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 to the ambiguity averse version studied in Section 4. Section 5 describes the model implications for exchange rate determination and discusses alternative specifications. Section 6 concludes. In the Appendix I review some of the relevant theoretical literature and provide details on some of the model's equations.

¹³See Burnside et al. (2008) and Jurek (2008).

¹⁴Burnside et al. (2008) point out that the rare event could be a large loss from the strategy or a very high marginal utility when an otherwise small loss happens. In the first case we should expect to see zero payoffs for the hedged carry trade. This implication is rejected by the empirical evidence on the profitability of the hedged carry trade. They then conclude that it must be the extremely high marginal utility that characterizes such a negative event.

¹⁵See Jurek (2008) for an empirical analysis of risk-neutral moments.

¹⁶Traders describe this situation for the high interest rate currency as "exchange rates go up by the stairs and down by the elevator" (see Brunnermeier et al. (2008)).

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.

There are overlapping generations (OLG) of investors who each live 2 periods, derive utility from end-of-life wealth and are born with zero endowment. There is one good for which purchasing power parity 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 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 driving exogenous force is the process for the interest rate differential $r_t = i_t - i_t^*$. The true DGP can be described using a state-space model:

$$\begin{aligned} r_t &= H'x_t + \sigma_V v_t \\ x_t &= Fx_{t-1} + \sigma_U u_t \end{aligned} \tag{1}$$

The shocks u_t and v_t are white noises. Thus, at time t the observable differential is the sum of a hidden unobservable persistent (x_t) and a temporary component (v_t). The agent entertains 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 a CARA utility over end-of-life wealth, W_{t+1} , with a rate of absolute risk-aversion of γ .

$$V = \max_{b_t} \min_{\tilde{P} \in \Lambda} E_t^{\tilde{P}} [-\exp(-\gamma W_{t+1}) | I_t] \tag{2}$$

where I_t is the information available at time t and b_t is the amount of foreign bonds invested. 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 subjective probability distribution available to the agent. They decide 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 \tilde{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 $t + 1$. However, at the same time the agent has to repay the interest rate bearing borrowed amount of $b_t \exp(i_t)$ expressed in USD. 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 is then a function of the amount of bonds invested and the excess return:

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

To close the model I specify a Foreign bond market clearing condition similar to Bacchetta and van Wincoop (2008). There is a fixed supply B of Foreign bonds in the Foreign currency. In steady state the investor holds no assets since she has 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 1. I also normalize the steady state log exchange rate to $s^{SS} = 0$. The market clearing condition for Foreign bonds is then:

$$b_t = Be^{s_t} - B \tag{3}$$

where B is the steady state amount of Foreign bonds. Following Bacchetta and van Wincoop (2008) I also set $B = 0.5$, corresponding to a two-country setup with half of the assets supplied domestically and the other half by the rest of the world. By log-linearizing the RHS of (3) around steady state I get the market clearing condition¹⁷:

$$b_t = .5s_t \tag{4}$$

2.2 Model uncertainty

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 (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 worse-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.

¹⁷Bacchetta and van Wincoop (2008) analyze an alternative model with constant relative risk aversion in which agents are born with an endowment of one good and decide what fraction of it to invest in the foreign bond. The same equilibrium conditions are obtained as in this model except that those conditions are expressed in deviations from steady state.

Abusing notation let \tilde{P} also denote the arg min for the problem in (2). Thus, for future reference \tilde{P} is the optimal subjective probability distribution over the exogenous process for the interest rate differential.

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), 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 (1).¹⁸

Thus the agent uses the following state-space representation:

$$\begin{aligned} r_t &= H'x_t + \sigma_{V,t}v_t \\ x_t &= Fx_{t-1} + \sigma_Uu_t \end{aligned} \tag{5}$$

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

The information set is $I_t = \{r_{t-s}, s = 0, \dots, t\}$. 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) 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 the optimal belief \tilde{P} the agent will imagine what could be the worst-case realizations for $\sigma_{V,s}$ for the data that she observes.

This minimization then becomes selecting a sequence of

$$\sigma_V^t = \{\sigma_{V,s}, s \leq t : \sigma_{V,s} \in \Upsilon\} \tag{6}$$

in the product space $\Upsilon^t : \Upsilon \times \Upsilon \dots \Upsilon$. As in Epstein and Schneider (2007), the agent interprets this sequence as a “theory” of how the data was generated.

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 is the distorted model from the true DGP, I include the value σ_V in the set Υ . This does not necessarily imply that this is a priori known. If the agent uses maximum likelihood for a constant volatility model, her point estimate would be asymptotically σ_V . 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 worse-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.

The type of structured uncertainty I consider implies that the minimization in (2) is reduced to selecting a distorted sequence of the form (6). The optimization in (2) then becomes:

$$V = \max_{b_t} \min_{\sigma_V^t(r^t) \in \sigma_V^t} E_t^{\tilde{P}}[-\exp(-\gamma W_{t+1})|I_t] \tag{7}$$

¹⁸In Section 5.1 and Appendix F I discuss alternative specifications, including time-varying parameters.

where \tilde{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 .

2.3 Statistical constraint on possible distortions

An important question that arises in this setup is how easy is it 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.¹⁹ The ambiguity aversion models also constrain the minimization by imposing some cost function on this distance.²⁰ 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 ($L^{Dist}(r^t)$). The metric is the probability of model detection error which measures in this case how often $L^{DGP}(r^t)$ is smaller than $L^{Dist}(r^t)$.²¹ Hence, this shows how likely it is that the distorted sequence, treated as deterministic, produces a higher likelihood than the constant volatility model based on σ_V .

Given the set Υ and the desired level of error detection probability, it 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. Treating Υ and the level of error detection probability as parameters it amounts to solving for the closest integer n . For example if $n = 2$, as in the main parameterization, it means that the agent is in fact choosing only two dates where to be concerned that the realizations of $\sigma_{V,t}$ are different than σ_V .

This approach amounts to setting an average statistical performance of the distorted model. At each time t , $L^{Dist}(r^t)$ can be larger or smaller than $L^{DGP}(r^t)$, but on average it is higher than the latter with the selected fixed detection error probability (for example in the main parameterization, this is set to 0.17). An alternative, employed in Epstein and Schneider (2007) would be to fix a significance level for the likelihood ratio test which holds every period so that $L^{Dist}(r^t)$ is lower than $L^{DGP}(r^t)$ every period by some fixed amount, and allow the number of dates n to vary by period. Similar intuition and results are obtained.²² I choose to work with the first alternative for computational reasons and also to capture the idea that the distorted model is not always performing worse. Sometimes the distorted model looks

¹⁹See Anderson et al. (2003) and Hansen and Sargent (2008).

²⁰See Klibanoff et al. (2005) and Maccheroni et al. (2006) among others.

²¹This comparison is close to the detection error probability suggested in Hansen and Sargent (2008). The difference here is that I only consider the error probability when the reference model is the true DGP.

²²For the same set Υ as in the benchmark case, agents constrained by a significance level of 0.05 or 0.1 will be able to distort the variance only for a small number of times, i.e. n usually belongs to $\{1, 2, 3\}$.

even more plausible statistically than the reference model. 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.²³

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 the choose function of t and $n : {}^tC_n = t!/(n!(t - n)!)$. When the agent considers distorting a date she will choose low precision of the signal if that date’s innovation is good news for her investment and high precision if it is bad news. However, even this number becomes increasingly large as t increases. When the model is solved numerically, as described in Section 4.2, I will make the further assumption that the agent considers only distortions to the dates $t, \dots, t - m + 1$. That reduces the number of possible sequences to mC_n . In Section 5.1 I discuss the extent to which this affects the results.

It is important to emphasize that I impose the restriction on the distorted sequence to be different only for few dates from the reference model purely for reasons related to statistical plausibility. The same intuition applies if the agent is not constraint by this consideration. In that case, as in Epstein and Schneider (2008) the agent would interpret all past innovations that are good news as low precision signals and bad news as high precision signals. Given the set Υ that I consider in the benchmark parameterization such a sequence of signals would look very unlikely compared with the reference model. I then allow the agent to restrict attention only to a number of dates so that the two competing sequences have similar likelihoods.

2.4 Discussion of the setup

In this section I discuss some of the modeling choices. First, I assume a preference over wealth and take exogenous the interest rate differential. I want to capture the agents’ uncertainty about the evolution of this process which endogenously determines the exchange rate. In order to have uncertainty over the interest rate differential, I cannot simply use a model in which the differential is an endogenous variable. For example, in a classic CAPM model the inverse of the interest rate is the expected stochastic discount factor driven by the exogenous process for consumption. It is thus an endogenous variable that reflects the evolution of this expectation. In that case, it is hard to argue that the agents are not sure how this process occurs since it is reflecting their endogenous choices.

²³For 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).

To move away from this modeling implication, I could use a monetary model in which the government sets the interest rate, as in a Taylor rule.²⁴ In that model, one could imagine scenarios in which agents are uncertain about the way that the interest rate is set.²⁵ For example, the agents would need to distinguish between persistent shifts in the inflation target from transitory disturbances to the policy rule. Erceg and Levin (2003) study such a setup and use rational expectations by endowing the agents with the true DGP. Naturally, if agents act under the true distribution there is no room for systematically distorted beliefs to explain the UIP puzzle. In an ambiguity aversion model, agents are not endowed with this knowledge and are concerned about time-variation in the relative size of the persistent and transitory shock. The optimal gain is time-varying to reflect such concerns.²⁶

The type of uncertainty and the constraint on the set of possible distorted models is relatively new in the literature. I extend the model in Epstein and Schneider (2007, 2008) by considering a setup with ambiguous signals about time-varying hidden states. The reference model is a state-space constant volatility while the distorted one is a stochastic volatility representation of the data. The latter is different from a typical stochastic volatility in which the probabilities of drawing the realizations for the time-varying standard deviations are known and Bayesian inference occurs. Here the agent is not willing to place these probabilities, but rather, as the max min principle dictates²⁷, will choose these probabilities to be either 1 or 0.²⁸

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 my setup, the optimal choice will likely be different due to two reasons: the distance constraint will typically eliminate such a possibility and even if not, it is still the case that sequences with time variation might induce a lower utility for the agent. The fact that the likelihood comparison will be strong evidence against such a model is related to the idea that distortions in variances are easier to detect in the data. Within the model's setup the distorted sequence is identical to the reference one, except for a few dates in the observable data. Such a sequence would then be harder to ignore based on statistical significance.

Whether I assume uncertainty about the realizations for the variances of the observable shock or the trend shock is intuitively innocuous. The driving force is the agent's evaluation of the expected

²⁴See Engel et al. (2007) for a discussion and summary of evidence that news about economic fundamentals tend to be incorporated into the exchange rate market as predicted by an evolution of the interest rate determined through a Taylor rule.

²⁵Such a possibility is raised by the debate over the stability of the Taylor rule and the literature on US time-varying monetary shocks versus rules. See for example Clarida et al. (2000) and Sims and Zha (2006) for a discussion.

²⁶For more examples of models with the private-sector using econometric approaches to learn (such as discounted least-squares) about the interest rate rule see the references in Evans and Honkapohja (2001).

²⁷The "maxmin" preference, as for example in Gilboa and Schmeidler (1989), corresponds to an infinite level of uncertainty aversion as the agent chooses the worst-case scenario from a set of distributions. For smoothed ambiguity aversion models see Klibanoff et al. (2005), in which the agent does not choose the minimum of the set but rather weighs more the worse distributions.

²⁸A 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.

utility will be the expected return and much less the variance of the return. This is definitely the case with a risk neutral agent, but even in this setup with risk aversion, expected returns drive most of the portfolio decision. Expected returns are affected by the estimate for the hidden state which in turn depends on the time-varying signal to noise ratios. This means that it is not the specific place in which I assume uncertainty, the observation or state equation, that is important but the relative strength in the information contained in them. This avoids a problem that Li and Tornell (2007) have where they assume uncertainty only about the observation shock. There they need the assumption because the distortion only affects the variance of the returns. Uncertainty about the state equation will then manifest in choosing a higher variance for the persistent shock. This in turn generates a higher Kalman gain than the reference model which will imply overreaction to any type of news. In that case the model’s implications move even further away from rational expectations in explaining the puzzles.

As Hansen and Sargent (2007b) argue, when the agent only cares about the present or future value of the hidden state, a more relevant situation is that of no commitment to previous distortions. My model also investigates such a case. However, different from my setup they consider unstructured (global) uncertainty. This type of uncertainty places no restriction on the nature or location of possible misspecification. The distorted model can be any probability distribution, as long as the “distance” from the reference model is bounded by some cost function (in their model the relative entropy). The typical approach in macroeconomics and finance has been to include this entropy directly into the utility function as a cost function whose relative importance is controlled by a Lagrange multiplier on the relative entropy constraint. Importantly, as discussed for example in Strzalecki (2007), Barillas et al. (2008), these multiplier preferences are observationally equivalent to a higher risk aversion expected utility. In Appendix C I present some details for this equivalence in my model. As I show in Section 3, in my setup higher risk aversion combined with rational expectations does not provide an explanation for the puzzles. I then conclude that this type of uncertainty is not suited in this model for addressing the empirical findings.

2.5 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 , $\{r_s\}_{s=0,\dots,t}$. 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 (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 $\{r^t\}$, $t = 0, 1, \dots, \infty$ 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 (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 $s(r^t)$, $\sigma_V^*(r^t)$ and $f(r^{t+1})$, the demand for bonds $b(r^t)$ is the optimal solution for the max problem in (7).

2. *Optimality:* given $s(r^t)$, $b(r^t)$ and $f(r^{t+1})$, the distorted sequence $\sigma_V^*(r^t)$ is the optimal solution for the min problem in (7).

3. *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 (4).

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

Notice that the consistency of beliefs imposes that the agent uses the correct equilibrium relation between the exchange rate and the exogenous sequence of interest rate differentials in forming her subjective probability distribution. At time t the unknown realization is r_{t+1} whose variation affects s_{t+1} by the equilibrium relation. The rational expectations assumption imposes one model for the distribution of r_{t+1} . The uncertainty averse agent surrounds this reference distribution by a set of possible distributions which are indexed by the sequences σ_V^t defined in (6). Each sequence σ_V^t implies a subjective probability distribution over the future realizations of s_{t+1} . The sequence $\sigma_V^*(r^t)$ and demand $b(r^t)$ are a Nash equilibrium in the zero-sum game between the minimizing and maximizing agent.

3 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. $P = \tilde{P}$. Thus the agent fully trusts her model which also turns out to be the DGP. For ease of notation, I denote by $E_t(X) \equiv E_t^P(X)$, where P is the true probability distribution. The DGP is given by the constant volatility state space described in (1). The optimization problem is

$$V = \max_{b_t} E_t[-\exp(-\gamma W_{t+1})|I_t]$$

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. Appendix B shows that the FOC is

$$b_t = \frac{E_t(q_{t+1})}{\gamma \text{Var}_t(q_{t+1})} \quad (8)$$

The market clearing condition states that $b_t = .5s_t$. Combining the demand and the supply equation I get the equilibrium condition for the exchange rate:

$$s_t = \frac{E_t(s_{t+1} - r_t)}{1 + .5\gamma Var_t(s_{t+1})} \quad (9)$$

I call (9) the UIP condition in the rational expectations version of the model. If $\gamma = 0$ it implies the usual risk-neutral version $s_t = E_t(s_{t+1} - r_t)$. With $\gamma > 0$ it takes into account a risk premium, given the utility function, coming from the variance of the excess return.

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. Intuitively, solving for the exchange rate means iterating forward on the UIP equation and forming expectations about future differentials. Given the Gaussian and linear setup the optimal filter for the state-space in (1) is the usual Kalman Filter.

Let $\hat{x}_{m,n} = E(x_m|I_n)$ and $\Sigma_{m,n} = E[(x_m - E(x_m|I_n))(x_m - E(x_m|I_n))']$ denote the estimate and the MSE of the hidden state for time m given information at time n . As shown in Hamilton (1992) the estimates are updated according to the following recursion:

$$\hat{x}_{t,t} = F\hat{x}_{t-1,t-1} + K_t(y_t - H'F\hat{x}_{t-1,t-1}) \quad (10)$$

$$\Sigma_{t,t} = (I - K_t H')(F\Sigma_{t-1,t-1}F' + \sigma_U\sigma_U') \quad (11)$$

$$K_t = (F\Sigma_{t-1,t-1}F' + \sigma_U\sigma_U')H[H'(F\Sigma_{t-1,t-1}F' + \sigma_U\sigma_U')H + \sigma_V^2]^{-1} \quad (12)$$

where K_t is the Kalman gain.

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

$$s_t = \Gamma\hat{x}_{t,t} + \delta r_t \quad (13)$$

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, as detailed in Appendix B and denoting the time-invariant conditional variance $Var(s_{t+1}|I_t)$ by σ^2 the solution is

$$\delta = -\frac{1}{1 + .5\gamma\sigma^2} \quad (14)$$

$$\Gamma = -\frac{1}{1 + .5\gamma\sigma^2}H'F[(1 + .5\gamma\sigma^2)I - F]^{-1} \quad (15)$$

$$\sigma^2 = (\Gamma K + \delta)(\Gamma K + \delta)'Var(r_{t+1}|I_t) \quad (16)$$

with $Var(r_{t+1}|I_t) = H'F\Sigma F'H + H'\sigma_U\sigma_U'H + \sigma_V^2$.

To gain intuition, suppose that the state evolution is an AR(1), i.e. $F = \rho$. Then, denoting by $c = (1 + .5\gamma\sigma^2)$, the coefficients become $\delta = -\frac{1}{c}$ and $\Gamma = -\frac{\rho}{c(c-\rho)}$. This highlights 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 Γ will be typically a large negative

number. It shows that s_t reacts strongly to the estimate of the hidden state $\hat{x}_{t,t}$, 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

$$s_{t+1} - s_t = \Gamma(\rho - 1)\hat{x}_{t,t} + \Gamma K(r_{t+1} - \rho\hat{x}_{t,t}) + \delta(\rho\hat{x}_{t,t} + \rho\xi_t + \sigma_U u_{t+1} + \sigma_V v_{t+1} - r_t) \quad (17)$$

where $\xi_t = x_t - \hat{x}_{t,t}$ with $\xi_t \sim N(0, \Sigma)$ and independent of time t information.

Then taking expectations of (17) I get

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

where I make use of the rational expectations assumption that $E_t(\varepsilon_{t+1}|I_t) = 0$. Since $cov(\hat{x}_{t,t}, r_t) = Kvar(r_t)$, the UIP coefficient is

$$\hat{\beta} = K \frac{\rho(1-c)}{c(c-\rho)} + \frac{1}{c}$$

Since $c > 1$, to get a lower bound on $\hat{\beta}$ I set $K = 1$ so that $\hat{\beta}^L = \frac{1-\rho}{c-\rho} < 1$. The reason for $\hat{\beta}^L < 1$ is the existence of a rational expectations risk premium in this model. For the risk neutral case, $\gamma = 0$, $c = 1$ and $\hat{\beta}^L = 1$.

To investigate the magnitude of $\hat{\beta}^L$ under risk aversion, I report below some simple calculations based on the data. The data is explained in a later section. I estimate an AR(1) process for the interest rate differential. Table 1 reports the estimated standard deviation of the shock to the forward discount, the AR(1) coefficient and the empirical standard deviation of the exchange rate. I compute the implied standard deviation for the rational expectations model as above by varying the risk aversion parameter γ .

Using the data for $std_t(r_{t+1})$ and ρ , I substitute them in (14),(15),(16) where K is also set to 1. The conclusion that emerges is that with a low level of risk aversion the reaction of the exchange rate to the interest rate (equal in this case to $\Gamma + \delta$) is large and can generate significant variability in the exchange rate. 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, “beating a random walk” in forecasting is too strong a criterion for accepting an exchange rate model. They conclude that typically these models should have low forecasting power for the interest rate differential in predicting the exchange rate change. With a low risk aversion, the model implied $\hat{\beta}^L$ is smaller than 1, but very close to it. As γ is increased, $\hat{\beta}^L$ decreases, but this happens at the cost of reducing the impact of the interest rate on the exchange rate. I view this as a negative implication because it delivers a lower volatility of the exchange rate and a higher explanatory power for the interest rate differential. Even when $\gamma = 500$, the model implied $\hat{\beta}^L$ equals around 0.5. Note also that $\hat{\beta}^L$ cannot be negative and in order to bring it down to 0 an extremely large level of risk aversion is required. In fact this coefficient would then be 0 because the exchange rate would be basically constant.

This highlights why a model of unstructured uncertainty, like the multiplier preferences discussed in Section 2.4 and analyzed in Appendix C, does not fare well in this setup. That type of model is equivalent to a rational expectations framework but with higher risk aversion. This is not a solution to the forward premium puzzles because it fails to generate a negative UIP regression coefficient. Driving the coefficient to 0 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. Without appealing to noise as driving the exchange rate, that implication is certainly counter-productive.

4 The Distorted Expectations Model Solution

The main equations involved in solving the distorted expectations model are the optimization problem (7), the subjective state space representation (5) and the market clearing condition (4). As in the rational expectations case I substitute out $\log W_{t+1}$ by its first order approximation $b_t q_{t+1}$ and obtain the FOC for the maximization problem as:

$$E_t^{\tilde{P}} [q_{t+1} \exp(-\gamma b_t q_{t+1})] = 0 \quad (18)$$

The FOC (18) can be rewritten as

$$s_t = E_t^{\tilde{P}} \left[s_{t+1} \frac{W_{t+1}^{-\gamma}}{E_t^{\tilde{P}} W_{t+1}^{-\gamma}} \right] - r_t \quad (19)$$

$$\mu_{t+1} = \exp(-\gamma b(r^t) q_{t+1}) \quad (20)$$

where μ_{t+1} is the marginal utility for the end-of-life wealth $W(r^t, s_{t+1}) = b(r^t)[s_{t+1} - s(r^t) - r_t]$.

Equation (19) is also useful for thinking about the risk neutral measure versus the objective measure. It is worth emphasizing that the former differs from the latter due to two factors: risk premia and uncertainty premia. The relevant expectation in (19) can be rewritten as $E_t^{\tilde{P}} \left[\frac{d\tilde{P}_N}{d\tilde{P}} s_{t+1} \right]$ where \tilde{P}_N is the risk neutral measure and $\frac{d\tilde{P}_N}{d\tilde{P}}$ is the corresponding Radom-Nikodym derivative. The uncertainty premia is summarized by the difference between the distorted and the reference model, $E_t^{\tilde{P}_N} [s_{t+1}] = E_t^P \left[\frac{d\tilde{P}_N}{d\tilde{P}} \frac{d\tilde{P}}{dP} s_{t+1} \right]$. As I argued before, the risk premia corrections, i.e. $\frac{d\tilde{P}_N}{d\tilde{P}}$, do not account in this model for the empirical puzzles. The key mechanism is going through $\frac{d\tilde{P}}{dP}$, with \tilde{P} being distorted from the objective measure P .

4.1 The optimal distorted expectations

In presenting the solution I use the constraints on the sequence $\sigma_V(r^t)$ described in Section 2.3, which derive 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

being different than σ_V . For computational reasons, in Section 2.3 I also introduce a truncation on the possible distorted sequences that effectively means that the agent is concerned that only n out of the last m observations were generated by time varying volatilities.

Note that for a given deterministic sequence $\sigma_V(r^t) = \{\sigma_{V,s}, s = 0, \dots, t\}$ selected in (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 (10), (11). The difference with the constant volatility case is that the Kalman gain now incorporates the time-varying volatilities $\sigma_{V,t}^2$:

$$K_t = (F\Sigma_{t-1,t-1}F' + \sigma_U\sigma_U')H[H'(F\Sigma_{t-1,t-1}F' + \sigma_U\sigma_U')H + \sigma_{V,t}^2]^{-1}$$

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 that 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 to the period of the optimally chosen object of choice, i.e. in the sequence $\sigma_V(r^t)$ in the definition of the equilibrium.

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 to be 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 and filtering problem I denote by:

$I_j^t = \{r_s, H, F, \sigma_U, \sigma_{V,(t),s}, s = 0, \dots, j\}$ the information set that the filtering problem has at time j by treating as known the sequence $\sigma_{V,(t),s}, s = 0, \dots, j$. This sequence is optimally selected at date t . Thus,

$$\begin{aligned}\hat{x}_{i,j}^t &= E(x_i | I_j^t) \\ \Sigma_{i,j}^t &= E[(x_i^t - \hat{x}_{i,j}^t)(x_i^t - \hat{x}_{i,j}^t)' | I_j^t] \\ K_i^t &= (F\Sigma_{i-1,i-1}^t F' + \sigma_U\sigma_U')H[H'(F\Sigma_{i-1,i-1}^t F' + \sigma_U\sigma_U')H + \sigma_{V,(t),i}^2]^{-1}\end{aligned}\quad (21)$$

Thus $\hat{x}_{i,j}^t$ is the estimate of the hidden state for time i based on the sample $0, \dots, j$ by treating as known the sequence $\sigma_{V,(t),s}, s = 0, \dots, j$. This sequence is chosen at date t . K_i^t is the Kalman gain to be applied at time i by using the known realization of $\sigma_{V,(t),i}^2$. This value is an element of the sequence $\sigma_{V,(t),s}, s = 0, \dots, j$.

In order to solve the problem in (7) and (18) I endow the agent with a guess about the relationship between the future exchange rate and the estimates for exogenous process. Similar to the RE case, I will restrict attention to linear function. Let this guess be

$$s_{t+1} = \Gamma \hat{x}_{t+1,t+1}^{t+1} + \delta r_{t+1}\quad (22)$$

Note the fact that $\hat{x}_{t+1,t+1}^{t+1}$ is the estimate obtained by the date $t + 1$ agent which uses the time $t + 1$ distorted sequence $\sigma_V(r^{t+1})$.

For the minimization in (7) the agent needs to understand how the expected utility under the distorted model depends on $\sigma_V(r^t)$. For every possible $\sigma_V(r^t)$ the agent computes the implied estimates and relevant state variables at time $t + 1$. In comparing these cases, the agent is using the guess in (22).

To understand the solution we could imagine an approximation of the per period felicity function in (7) to $U = -\exp(-\gamma E_t^{\tilde{P}} W_{t+1} + \gamma^2 \text{Var}_t^{\tilde{P}} W_{t+1})$. As done in the solution to the portfolio choice problem, up to a second order, $W_{t+1} = b_t q_{t+1}$.

Thus, $E_t^{\tilde{P}} W_{t+1} = b_t (E_t^{\tilde{P}}(s_{t+1}) - s_t - r_t)$ and $\text{Var}_t^{\tilde{P}} W_{t+1} = b_t^2 \text{Var}_t^{\tilde{P}} s_{t+1}$. Given the guess in (22) and the Kalman filtering formulas

$$s_{t+1} = \Gamma(I - K_{t+1}^{t+1} H') F \hat{x}_{t,t}^{t+1} + (\Gamma K_{t+1}^{t+1} + \delta) r_{t+1} \quad (23)$$

The sequence $\sigma_V(r^t)$ does not affect the estimates $\hat{x}_{t,t}^{t+1}, K_{t+1}^{t+1}$ since these are formed based on the optimal choice of the time $t + 1$ agent.

4.1.1 Effect of higher observational noise on expected excess returns

Using (23), the exchange rate for the next period s_{t+1} is monotonic in the realization of the interest rate differential r_{t+1} . More specifically, in equilibrium it is a decreasing function of r_{t+1} . The direction of the monotonicity is controlled by $(\Gamma K_{t+1}^{t+1} + \delta)$ where Γ, δ are to be determined in equilibrium. As expected, the same intuition about these parameters 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. Thus, in equilibrium $(\Gamma K_{t+1}^{t+1} + \delta)$ is a negative number. By this result, $E_t^{\tilde{P}}(s_{t+1})$ is also decreasing in $E_t^{\tilde{P}}(r_{t+1})$.

The expected interest rate differential is given by the hidden state estimate $E_t^{\tilde{P}}(r_{t+1}) = H' F \hat{x}_{t,t}^t$. This is clearly increasing in the innovation $r_t - H' F \hat{x}_{t-1,t-1}^t$. In turn, the estimate $\hat{x}_{t,t}^t$ is updated by incorporating this innovation using the gain K_t^t . The latter is decreasing in the variance of the observation shock $\sigma_{V,(t),t}^2$. Intuitively, more noise in the measurement equation implies less information for updating the hidden state estimate. Combining these two monotonicity results, the estimate $\hat{x}_{t,t}^t$ is increasing in the gain if the innovation is positive. On the other hand, if $(r_t - H' F \hat{x}_{t-1,t-1}^t) < 0$ then $\hat{x}_{t,t}^t$ is decreased by having a larger gain K_t^t .

By construction, expected excess return $E_t^{\tilde{P}} W_{t+1}$ is monotonic in $E_t^{\tilde{P}}(s_{t+1})$. 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 $E_t^{\tilde{P}}(s_{t+1})$ will hurt her.

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

Proof. By combining the sign of the partial derivatives involved in $\frac{\partial E_t^{\tilde{P}} W_{t+1}}{\partial \sigma_{V,(t),t}^2}$. For details, see Appendix D.

The impact of $\sigma_{V,(t),t}^2$ on utility through the effect on the expected excess returns is given by the following intuitive mechanism. Suppose in equilibrium the agent invests in domestic bonds. She is then worried about a depreciation of the domestic currency, which in equilibrium happens if the estimated hidden state of the differential $\hat{x}_{t,t}^t$ is lower, but still positive. The variance $\sigma_{V,(t),t}^2$ affects the gain K_t^t . To get a lower estimate $\hat{x}_{t,t}^t$ the variance $\sigma_{V,(t),t}^2$ is increased if the innovation $(r_t - F\hat{x}_{t-1,t-1}^t)$ is positive and is decreased if the innovation is negative.

4.1.2 Effect of higher observational noise on expected variance of excess returns

The variance of excess returns is given by $Var_t^{\tilde{P}} W_{t+1} = b_t^2 Var_t^{\tilde{P}} s_{t+1}$. In turn, using the conjecture (23) and taking as given $\hat{x}_{t,t}^{t+1}, K_{t+1}^{t+1}$

$$Var_t^{\tilde{P}} s_{t+1} = (\Gamma K_{t+1}^{t+1} + \delta)(\Gamma K_{t+1}^{t+1} + \delta)' (Var_t^{\tilde{P}} r_{t+1}) \quad (24)$$

By the filtering formulas

$$Var_t^{\tilde{P}} r_{t+1} = H' F \Sigma_{t,t}^t F' H + H' \sigma_U \sigma_U' H + E_t^{\tilde{P}} (\sigma_{V,t+1}^2) \quad (25)$$

In Appendix D.1 I discuss the object $E_t^{\tilde{P}} \sigma_{V,t+1}^2$ and conclude that it is equal to $(\sigma_V^H)^2 \frac{n}{t} + (\sigma_V)^2 (1 - \frac{n}{t})$. When t is large then this expectation becomes σ_V^2 .

Proposition 2 *The expected variance of excess return, $Var_t^{\tilde{P}} W_{t+1}$ is increasing in $\sigma_{V,(t),t}^2$.*

Proof. The variance $Var_t^{\tilde{P}} W_{t+1}$ is increasing in the conditional variance of the differential r_{t+1} . By (25) the latter is increasing in $\sigma_{V,(t),t}^2$ through the effect on $P_{t,t}^t$. For details see Appendix D.

Intuitively, more noise in the observation 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 $Var_t^{\tilde{P}} r_{t+1}$ because $\frac{\partial P_{t,t}^t}{\partial \sigma_{V,(t),t}^2} > 0$.

The overall effect of $\sigma_{V,(t),t}^2$ on the utility V_t is then coming through two channels. One is the positive relationship between $\sigma_{V,(t),t}^2$ and the variance of the returns as in (24). As shown above, $\sigma_{V,(t),t}^2$ also influences V_t through the expected returns. The total partial derivative is then

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

The sign of this derivative is :

$$sign\left(\frac{\partial V_t}{\partial \sigma_{V,(t),t}^2}\right) = sign(b_t) sign(r_t - F\hat{x}_{t-1,t-1}^t) - sign\left(\frac{\partial Var_t^{\tilde{P}} r_{t+1}}{\partial \sigma_{V,(t),t}^2}\right)$$

From (25) the $sign(\frac{\partial Var_t^P r_{t+1}}{\partial \sigma_{V,(t),t}^2})$ is positive. Thus, if the sign of $[b_t(r_t - F\hat{x}_{t-1,t-1}^t)]$ is negative the two effects align because higher variance $\sigma_{V,(t),t}^2$ will imply lower expected excess returns. However, if the sign is positive the two directions are competing. To analyze this situation I show in Appendix C that in this setup the probability that the effect through the expected returns to dominate the one through the variance is almost equal to 1. I conclude that in this model the effect of $\sigma_V(r^t)$ on utility goes through its effect on expected returns.

The position b_t dictates in what direction is the agent pessimistic. If, for example, $b_t > 0$ she invests in foreign currency and is worried about a domestic appreciation. This will happen in equilibrium if the foreign rate is higher than the domestic one. Then the agent fears that the estimate of the hidden state of the domestic interest rate differential is higher, although still negative. To make this estimate higher, the optimal sequence $\sigma_V^*(r^t)$ involves choosing high values for $\sigma_{V,(t),s}^2$ when the innovations $r_{t-s} - F\hat{x}_{t-s-1,t-s-1}^t$ are negative and low values when these innovations are positive. The decision rule for choosing a distorted $\sigma_{V,(t),s}$ is:

$$\begin{aligned}\sigma_{V,(t),s} &= \sigma_V^H \text{ if } b_t(r_s - F\hat{x}_{s-1,s-1}^t) < 0 \\ \sigma_{V,(t),s} &= \sigma_V^L \text{ if } b_t(r_s - F\hat{x}_{s-1,s-1}^t) > 0\end{aligned}\tag{26}$$

In Section 2.3 I introduced the restriction that the agent only considers n dates out of the last m to be different from the reference model. Out of these possible sequences, the intuition given by (26) shows on what type of sequences the agent restricts attention because they affect negatively her utility. The number of such possible sequences is the choose function of m and n : ${}^m C_n = m!/(n!(m-n)!)$. Out of these relevant sequences the agent chooses the one that minimizes the estimate of the hidden state. If for example $m = n$, as in the benchmark parameterization, ${}^n C_n = 1$ and the decision rule in (26) implies the existence of only one such sequence. In the definition of the equilibrium I denoted by $\sigma_V^*(r^t)$ the solution to this minimization problem. This decision rule is taking b_t as given. The solution for this is similar to the RE case, except that the subjective probability distribution for the future excess return is the one implied by the optimal choice for $\sigma_V^*(r^t)$ found above.

4.2 Numerical solution procedure

The driving equilibrium relation is the no-arbitrage condition given by the FOC with respect to b_t in (18). To solve that problem the agent needs to form forecasts about the next period exchange rate. The conjecture in (22) is that $s_{t+1} = \Gamma\hat{x}_{t+1,t+1}^{t+1} + \delta r_{t+1}$ so that forecasts of $\hat{x}_{t+1,t+1}^{t+1}$ and r_{t+1} are needed. At time t the agent knows the equilibrium updating rule for time $t+1$:

$$\hat{x}_{t+1,t+1}^{t+1} = F\hat{x}_{t,t}^{t+1} + K_{t+1}^{t+1}(r_{t+1} - H'F\hat{x}_{t,t}^{t+1})$$

She will correctly understand the decision rule for $\sigma_V(r^{t+1})$ given by (26). For every possible realization of r_{t+1} she will have to solve the agent's time $t+1$ problem, who will in that case face the sample (r^t, r_{t+1}) .

For the resulting $\hat{x}_{t+1,t+1}^{t+1}$ there will be a s_{t+1} based on the conjecture (22). This distribution of s_{t+1} will imply the distribution for $q_{t+1} = s_{t+1} - s_t - r_t$ in (18). However, because of the asymmetric responses to innovations, normality is lost and I did not find any closed-form solution to deal with it. Hence, to recover the distribution for s_{t+1} I perform the procedure numerically.

I restrict attention to linear time-invariant conjectures as in (22). Suppose first the parameters Γ and δ are known. The solution to the distorted expectations equilibrium can be summarized by the following steps.

1. Make a guess about the sign of b_t to use in (26).
2. Use (26) 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 $\hat{x}_{t,t}^t$ and $\Sigma_{t,t}^t$.
3. Draw realizations for r_{t+1} from $N(H'F\hat{x}_{t,t}^t, Var_t^{\tilde{P}}r_{t+1})$, where $Var_t^{\tilde{P}}r_{t+1}$ is defined in (25). Form the sample $r^{t+1} = (r^t, r_{t+1})$. For each realization perform Steps 1 and 2 above to obtain the sequence $\sigma_V^*(r^{t+1})$.
4. For each realization in step 3 use $\sigma_V^*(r^{t+1})$ to compute $\hat{x}_{t+1,t+1}^{t+1}$ and use the conjecture in (22) to generate a realized s_{t+1} .
5. The distribution of s_{t+1} in step 5 defines the subjective probability distribution for the agent at time t . Use the FOC (18) to solve for s_t^* .
6. If $sign(s_t^*) = sign(b_t)$ the solution is $\sigma_V^*(r^t)$ and s_t^* . If not, switch the sign in step 1.
7. If there is no convergence on the sign of s_t^* and b_t , the solution is assumed to be $b_t^* = s_t^* = 0$.

The last point deserves some explanation. It is related to the Nash equilibrium solution between the minimizing and maximizing player in (7). Consider the solution for the sequence $\sigma_V^*(r^t)$ that minimizes the utility. This solution needs to take as given an investment position. When the guess is that the agent would like to invest in the seemingly higher rate currency she will choose a sequence $\sigma_V^*(r^t)$ to decrease the estimate of the differential. If based on this worse-case scenario the solution to the portfolio choice is to invest in the other currency it brings about a difficult situation for the agent. Initially she considers investing in a currency but once she takes into account that the model might be misspecified and uses the distorted model under this investment strategy her resulting estimates make her want to invest in the other currency. If, when switching direction, the same problem happens it means that there is no Nash equilibrium in pure strategies. I do not consider mixed strategies and instead I impose the solution in these cases to be $b_t = s_t = 0$. These are situations in which the hidden state estimate of the differential is very close to 0 under the true DGP and the uncertain averse agent is not willing to take any side in the strategy. For this b_t the agent does not invest in any bond so any solution to $\sigma_V^*(r^t)$ is a best response to b_t . Note that the particular assumed solution for $\sigma_V^*(r^t)$ does not have any impact on the results since in the calculations of excess returns only periods when b_t is different from zero are considered. For future reference I call this situation the “inaction” effect.

A concern is how to recover the parameters Γ and δ in the guess (22). The consistency of beliefs in the definition of equilibrium (see Section 2.5) requires that the guess about the law of motion be the correct relation on average. Because the distortions expectations model is solved numerically, this consistency

will require some approximations. I first start by using the values for Γ and δ for the RE case. Quite surprisingly, I find that the solution s_t^* from the numerical procedure and its implied value $\Gamma\hat{x}_{t,t}^t + \delta r_t$ by (22) are close. On average the difference is 0 in long samples. Nevertheless, one could still expect that $[s_t^* - (\Gamma\hat{x}_{t,t}^t + \delta r_t)]b_t < 0$ because the agent takes into account the asymmetric response to news for the next period. For example, when $b_t < 0$, that should bring the expected exchange rate slightly closer to 0. Thus s_t^* should be slightly higher than $\Gamma\hat{x}_{t,t}^t + \delta r_t$. However, the “inaction” effect mitigates this response. When $b_t < 0$ the agent realizes that in equilibrium it is very unlikely that $s_{t+1} > 0$, i.e. that agents switch their position in the carry trade next period. High realizations for r_{t+1} that would result under the RE in switching positions are now most likely to be in the “inaction” region and be characterized by $s_{t+1} = 0$. A similar intuition applies for $b_t > 0$. I find that the combined effect of these two directions is that s_t^* is close to $\Gamma\hat{x}_{t,t}^t + \delta r_t$. Although the RE cases still works relatively well as an approximation, I look for Γ and δ that minimize the distance between the two objects even in subsamples. I find that a reasonable approximation is characterized by a small change compared to the RE case which makes the s_t respond slightly less to the estimate $\hat{x}_{t,t}^t$. It would correspond to having a slightly less persistent state.

The market clearing condition states that $b_t = .5s_t$. In the rational expectations case I approximated $\log W_{t+1}$ by $b_t q_{t+1}$. Then using the normality of q_{t+1} I obtained the mean-variance solution in (8). In the distorted expectations model the excess return is no longer exactly normally distributed. However, I find that numerically the mean-variance approximation to the bond demand is very accurate. For intuition, I present this case and by using $b_t = \frac{E_t^{\tilde{P}}(q_{t+1})}{\gamma Var_t^{\tilde{P}}(q_{t+1})}$, I get

$$s_t = \frac{E_t^{\tilde{P}}(s_{t+1} - r_t)}{1 + .5\gamma Var_t^{\tilde{P}}(s_{t+1})} \quad (27)$$

I call (27) the UIP condition in the distorted expectations version of the model.

4.3 Options and risk neutral skewness

In this section I introduce options and define the risk-neutral probability distribution. These elements are needed for contrasting the model’s implications against the data. The asymmetric response to news that underlies the optimality of the distorted model is generating in this model another interesting feature: the negative skewness of the carry trade returns. This characteristic has sometimes been called “crash risk”. Burnside et al. (2008), Brunnermeier et al. (2008) and Jurek (2008) find strong evidence of this. Brunnermeier et al. (2008) argues that the mean profitability might be a compensation for the negative skewness. They also note that the data suggests that the negative skewness is endogenous. It is positively predicted by a larger interest rate differential. They argue that the sudden unwinding of the carry trade positions causes the currency to crash. In their view, the cause is liquidity shortages. In my model, the endogenous unwinding is caused by asymmetric response to news.

For analyzing the model implied risk-neutral $Skew_{t,t+1}^{\tilde{P}}$ I use the numerical procedure for solving the distorted expectations model and construct the distribution of excess returns $q_{t+1} = s_{t+1} - s_t - r_t$ based on

her distorted model. This distribution is used by the agent is solving for the optimal investment decision and thus generate the equilibrium exchange rate. The risk neutral measure is implied by the FOC in (19): $E_t^{\tilde{P}^N} [q_{t+1}] = E_t^{\tilde{P}} \left[\frac{d\tilde{P}^N}{d\tilde{P}} q_{t+1} \right]$.

In the unhedged version of the carry trade presented above, the agent's end-of-period wealth is exposed to both upside and downside risk. Of particular concern is the possibility of significant losses produced by large depreciations of the investment currency, i.e. for which the interest rate is higher. To eliminate the downside risk, an agent can use an option that provides insurance against the left tail of the returns.

A call option on the FCU will give the agent the right but not the obligation to buy foreign currency with domestic currency (USD) at a prespecified strike price k_t dollars per unit of FCU. Let $C(k_t)$ denote the price, including the time t interest rate, of the call option with strike price k_t . The net payoff in USD from investing in this option is²⁹:

$$z_{t+1}^C(k_t) = \max(0, s_{t+1} - k_t) - C(k_t)$$

Similarly, a put option on the FCU gives the agent the right, but not the obligation, to sell foreign currency for USD at a prespecified strike price k_t dollars per unit of FCU. Let $P(k_t)$ denoting the price, including the time t interest rate, of the put option with strike price k_t . The net payoff in USD from investing in this option is:

$$z_{t+1}^P(k_t) = \max(0, k_t - s_{t+1}) - P(k_t)$$

Options are priced according to the no-arbitrage conditions:

$$C(k_t) = E_t^{\tilde{P}} \left[\max(0, s_{t+1} - k_t) \frac{\mu_{t+1}}{E_t^{\tilde{P}} \mu_{t+1}} \right] \quad (28)$$

$$P(k_t) = E_t^{\tilde{P}} \left[\max(0, k_t - s_{t+1}) \frac{\mu_{t+1}}{E_t^{\tilde{P}} \mu_{t+1}} \right] \quad (29)$$

where μ_{t+1} is defined in (20).³⁰

Options are used by the agent as protection against negative realizations. When she decides to take advantage of the higher domestic interest rate by investing in the domestic currency, she is worried about a pronounced depreciation of the domestic currency. In that case her downside risk is eliminated by buying a call option on the FCU which pays when s_{t+1} is higher than the strike price k_t . Similarly, when the agent invests in the foreign currency she buys a put option on the FCU that pays when the foreign currency diminishes in value, i.e. when s_{t+1} is lower than k_t .

The unhedged carry trade strategy means borrowing in the low interest rate currency and investing in the high interest rate currency. The agents in the model use this strategy in equilibrium. Their 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 $s_t < 0$. In simulations,

²⁹The payoff should be expressed in terms of levels of the exchange rate as $z_{t+1}^C(K_t) = \max(0, \exp(s_{t+1}) - \exp(k_t)) - C(K_t)$. Using the approximation $\exp(x) - 1 \approx x$, when x is small, this becomes $z_{t+1}^C(k_t) = \max(0, s_{t+1} - k_t) - C(k_t)$.

³⁰This implies the put-call parity $C(k_t) - P(k_t) = f_t - k_t$. By the covered interest parity $f_t = s_t + r_t$.

the coefficient of correlation between s_t and r_t is around -0.97 .³¹ The payoff on a dollar bet for the unhedged carry trade strategy is:

$$\begin{aligned} z_{t+1} &= r_t - (s_{t+1} - s_t) \text{ if } b_t < 0 \\ z_{t+1} &= (s_{t+1} - s_t) - r_t \text{ if } b_t > 0 \end{aligned} \quad (30)$$

Then I define the return to the hedged carry trade, z_{t+1}^H , as the sum of the payoff to the unhedged carry trade and from buying an option at strike price k_t corresponding to the strategy of eliminating the downside risk :

$$\begin{aligned} z_{t+1}^H(k_t) &= z_{t+1} + z_{t+1}^C(k_t) \text{ if } b_t < 0 \\ z_{t+1}^H(k_t) &= z_{t+1} + z_{t+1}^P(k_t) \text{ if } b_t > 0 \end{aligned} \quad (31)$$

where z_{t+1} is the payoff to the unhedged carry trade defined in (30)³².

4.4 Parameterization

In the benchmark case, the reference model is a state space representation with constant volatilities as in (1). The observable is the interest rate differential r_t and the state evolution is an AR(4).³³ Thus $H' = [1 \ 0 \ 0 \ 0]$ and F is a 4×4 vector with only 4 unknown coefficients to reflect the transformation from an AR(4) to a VAR (see Appendix E for details). I obtain the parameters F, σ_V, σ_U by estimating the model in (1) on interest rate data:

$$\begin{aligned} r_t &= H'x_t + \sigma_V v_t \\ x_t &= Fx_{t-1} + \sigma_U u_t \end{aligned} \quad (32)$$

The data set is obtained from Datastream and consists of daily observations for the mean of bid and ask interbank spot exchange rates, 1-month forward exchange rates, and 1-month interest rates. All exchange rates are quoted in FCUs per British pound. I convert daily data into nonoverlapping monthly observations. The data set covers the period January 1976 to December 2006 for spot and forward exchange rates and January 1981 to December 2006 for interest rates. The countries included in the data set are listed in Table 4. This data is a subset of the one used Burnside et al. (2008) (henceforth BEKR) where they also investigate bid-ask spreads.

Table 5 reports results for the Maximum Likelihood estimation of (32) on interest rate differential data. First, I find a very high degree of persistence in the state evolution. The table reports values for

³¹The 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 0. In these very few case I assume that the payoff to the carry trade is 0.

³²By the put-call parity the hedged strategy is equivalent to buying a put option when $b_t < 0$ and a call option when $b_t > 0$.

³³The AR(4) was selected based on the Akaike Information Criterion. Results are available on request from the author.

the long-run autocorrelation of the hidden state, denoted by $\sum \rho$, which is defined as the sum of the AR coefficients for the transition equation. Second, I find evidence for a positive σ_V with some heterogeneity in its statistical significance.³⁴ Interestingly, when using the data for the forward discount for the period 1976-2007 all country pairs are characterized by a larger σ_V than the one implied by the interest rate differential data set. These latter results seem to be driven both by the larger sample and the fact that the forward discount data suggests that the observation shocks are even more important. This might be an artifact of averaging over bid-ask quotes, which is done for both the interest rate and forward rate. A priori it is not clear which variable to use, since covered interest rate parity holds in the sample.³⁵

The benchmark parameterization is a calibrated version of (32) which averages across these different findings. In robustness checks I also use the other available data and find that the main conclusions hold. The results would be significantly weaker in the case in which the true σ_V would be several times larger than σ_U . I discuss this implication in Section 5.1.

Table 2 reports the benchmark parameterization. These values imply that the steady state Kalman weight on the innovation used to update the estimate of the current state is 0.86, 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.³⁶

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 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 ($L^{Dist}(r^t)$).

As discussed in Section 2.3 for computational reasons I need to make a restrictive assumption about the distorted sequence so that at time t the agent considers distortions only for dates $t - m + 1$ through t . This truncation is necessary to decrease the number of possible combinations for the alternative sequences that the agent can entertain. The benchmark parameterization is $m = n$. That reduces the number of possible sequences to $n!$. After taking into account the intuition for incorporating good news and bad news, as in (26), the agent obtains the optimal sequence. The reason that m is chosen to be a small number is the very intensive computational burden that occurs in steps 1-4 of the numerical procedure explained in Section 4.2 for obtaining the subjective distribution of the future values of the exchange

³⁴Performing a similar exercise, Gourinchas and Tornell (2004) find that there is no evidence of a positive σ_V in the interest rate differential data they analyze. There the sample is 1986-1996. I find that by including earlier data, which is characterized by more sudden movements, the evidence becomes stronger in favor of a positive σ_V .

³⁵See Burnside et al. (2008) for a thorough analysis of this data. Their data set is wider in coverage and they find that there is no evidence of CIP failure once accounting for bid-ask spreads.

³⁶Also note that when a distortion is occurring the Kalman gain does not instantly adjust to the steady gain but it evolves according to the deterministic chosen sequence of variances. With time-varying volatilities in the observation equation, such adjustments are nevertheless quick.

rate. There, a large number of draws are made out of the subjective distribution for future interest rate differential and for each of these the corresponding optimal distorted sequence needs to be solved for. If there is a large number of such sequences this computation becomes increasingly complex. In Section 5.1 I discuss the small extent to which this affects the results.

Table 3 reports some statistics for the likelihood comparison $L^{Dist}(r^t) - L^{DGP}(r^t)$ for various cases. Column (1) reports the mean and the standard deviation of this difference for the benchmark parameterization in which $m = n = 2$. It also shows that for about 17% of cases $L^{DGP}(r^t)$ is smaller than $L^{Dist}(r^t)$. This statistic is referred to as the probability of model detection error. Column (2) reports these values for the case in which $n = 2$ and $m = 10$. The results are very similar to the benchmark specification with the $L^{Dist}(r^t)$ being slightly higher on average in this case. Column (3) considers the situations in which there is no restriction on the number of periods for which the agent distorts the sequence so that any previous period characterized by good news is attached a low precision of signals and bad news a high precision of signals. 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 accross simulations is around 170 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.

The parameters δ, Γ, σ^2 that characterize the solution to the rational expectations model are given by equations (14), (15) and (16) respectively. As mentioned in Section 4.2 for the distorted expectations model I need to use a numerical approach to finding the equilibrium. I find that a good approximation is that the δ, Γ, σ^2 can be computed using the same equations but with a slightly less persistence in the state-equation. Thus, I compute these parameters for an \tilde{F} that differs from F above in its first component: $\tilde{F}(1, 1) = 1.538$. The intuition why the numerical procedure delivers such a result is that in the distorted expectations model the exchange rate in fact responds less to the same state variables as in the rational expectations case. This response is controlled by the persistence of the state equations. A larger maximal eigenvalue for F produces a larger eigenvalue for Γ .

For the absolute risk aversion parameter γ I use in the benchmark parameterization a value of 10. A higher value allows for risk aversion to play a slightly larger role, given the known extreme sensitivity of portfolio choice to expected excess returns in this setup. In fact, I find in robustness checks that as long as γ is in a standard range the main results are insensitive to its value.

5 Results

In this section I present the main implications that the distorted expectations model has for exchange rate puzzles. Figure 1 and 2 illustrate the exchange rate path and the estimate of the hidden state of the domestic interest rate differential under the distorted expectations model and the rational expectations (RE). 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 $0 < s_t < s_t^{RE}$ and $0 > s_t > s_t^{RE}$ where s_t^{RE} denotes the exchange rate under RE.

Figures 1 and 2 also show that the exchange rate under ambiguity is reacting 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), henceforth BNP). To investigate this asymmetric response further, Table 8 computes correlations between the exchange rate under ambiguity and under RE. The unconditional correlation in levels is very high while in first differences is lower. Columns (3) and (4) indicate 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. Similarly, for the estimate of hidden states, Table 9 shows that the correlation between the estimate in first difference under ambiguity and under RE is much weaker when the interest rate of the investment currency tends to increase compared to states in which this interest rate decreases.

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. 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 unhedged carry trade strategy; 3) predicts a positive excess return from the *hedged* carry trade strategy; 4) predicts a negatively skewed excess return for the unhedged carry trade strategy; 5) is associated with a more negative risk-neutral skewness of expected excess returns; 6) is followed by a gradual appreciation of the domestic currency.

5.1 The UIP puzzle and positive unhedged carry trade returns

In the distorted expectations model a risk adjusted version of UIP holds ex-ante and the expected excess returns under the risk neutral measure are zero (see (19)). The UIP regression is:

$$s_{t+1} - s_t = \beta r_t + \varepsilon_{t+1} \quad (33)$$

In the distorted model, with $\mu_{t+1} = \exp(-\gamma b_t q_{t+1})$ denoting the marginal utility and $\lambda_{t+1} = \frac{\mu_{t+1}}{E_t^{\tilde{P}} \mu_{t+1}}$, the expected exchange rate difference equals

$$E_t^{\tilde{P}}(s_{t+1}) - s_t = E_t^{\tilde{P}}(s_{t+1}) - E_t^{\tilde{P}}[s_{t+1}\lambda_{t+1}] + r_t$$

If the true DGP is the distorted model then $E_t^{\tilde{P}}(\varepsilon_{t+1}|I_t) = 0$ and

$$\hat{\beta} = 1 + \frac{cov(cov(s_{t+1}\lambda_{t+1}), r_t)}{var(r_t)}$$

The message of the time-varying risk premia story is that the term $\frac{cov(cov(s_{t+1}\lambda_{t+1}),r_t)}{var(r_t)}$ cannot deliver a significantly $\hat{\beta} < 1$. If the true DGP is the distorted model, the results established for the rational expectations apply here too except that the log excess return is not normally distributed. However, as I presented before, I find numerically that the mean variance result is a very good approximation for the optimal bond solution which delivers

$$E_t^{\tilde{P}}(s_{t+1}) - s_t = E_t^{\tilde{P}}(s_{t+1}) - \frac{E_t^{\tilde{P}}(s_{t+1} - r_t)}{1 + .5\gamma Var_t^{\tilde{P}}(s_{t+1})}$$

As explained at length in the RE case, the conditional variance $Var_t^{\tilde{P}}(s_{t+1})$ has a very small effect in obtaining a coefficient smaller than 1. In the distorted expectations model, the same result applies, with the observation that $Var_t^{\tilde{P}}(s_{t+1})$ is even slightly lower given the implied smaller response of the exchange rate to the fundamentals. For reasonable values of γ the lower bound on $\hat{\beta}$ would be also around 0.999. Hence in presenting the results I will make use of the fact that the $1 + .5\gamma Var_t^{\tilde{P}}(s_{t+1})$ is approximately equal to 1. The fact that in my setup the covariance term is almost zero is consistent with findings that study currency risk-premia, reviewed in Sections 1 and A. In a richer environment the ambiguity aversion theory presented in this paper can be incorporated into models that take more seriously the risk premia corrections. With the covariances becoming effectively zero the risk neutral UIP implies that $\hat{\beta} = 1$ since

$$E_t^{\tilde{P}}(s_{t+1}) - s_t = r_t$$

If the DGP is different from the distribution implied \tilde{P} the coefficient $\hat{\beta}$ can be significantly lower than 1. 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 $\hat{\beta}$.

Table 6 presents the estimated $\hat{\beta}$ for the model implied UIP regression in (33) in 1000 repeated samples of $T = 300$ and 500 repeated samples of $T = 3000$. Column (1) is the benchmark specification and shows that, for small T , the model is generating a negative average $\hat{\beta}$, even if not significant statistically. As the sample size is increased and the standard errors reduce, the average and median estimate become significant. This highlights that the results of the model are not limited to small sample 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 3 and 4 are histograms of the UIP coefficients and their t-statistics significance across many repeated samples. Figure 3 (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. As Figure 4 (top panel) shows these are usually not significant statistically. 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. Also these coefficients are significant in large samples

Column (2) of Table 6 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, 0.95 and 0.03 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 $\hat{\beta}$: the underreaction makes the coefficient become negative and the overreaction effect pushes it to be larger than 1. The combined effect depends on how far apart are the distorted gains 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 0 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 Table 6.B., 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 6 shows that when there is significantly less persistence in the state evolution the model cannot account for the UIP puzzle. In Table 6.C the long run autocorrelation of the 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.

This shows 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 6 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 6 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 larger negative 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 3 reports this case generates sequences that become increasingly less likely as the sample size grows.

In the benchmark parameterization I also make an assumption about the number m described in Section 2.3 and 4.4. This number effectively constrains the agent to only consider n dates out of the last m to be different from the reference model. I investigate this assumption by solving for the optimal distorted sequence when $m = 10$. As the original motivation for having a lower m suggests, I am not

able to simulate the future distribution of exchange rates for this high m due to computational intensity. I find that when $m = 10$ the optimal dates chosen by the agent are very close to time t . The model can be simulated to investigate the UIP coefficient by using the equilibrium relation for the exchange rate in (22). I find, as reported in Column (5) of Table 6 that for $m = 10$ the results are slightly weaker. As Column (2) of Table 3 shows, the error detection probability is also relatively larger suggesting that setting $m = 2$ imposes additional statistical penalty on the distorted sequence. Thus, compared to $m = 10$, in the benchmark case of $m = 2$ the model implies that the agent is using a distorted sequence that is generating slightly stronger results for the UIP puzzle while being slightly easier to distinguish statistically from the reference sequence. I find this result intuitive and conclude that the benchmark specification provides results that are robust to the assumption of $m = 2$.

The payoff on a dollar bet for the unhedged carry trade strategy is defined in (30). The model implied non-annualized monthly carry trade returns are described in Table 7.A. They are characterized by a positive mean, negative skewness and excess kurtosis. The excess returns 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 7.A. is 0.0017. For the data analyzed in Table 4 the average mean payoff for the carry trade strategy is 0.0033. This does not take into account transactions costs. BEKR analyze a more extensive 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 also shows that the average standard deviation of the model implied unhedged carry trade payoffs is around 0.011. For the data analyzed in Table 4 the average standard deviation is 0.03. BEKR report average values ranging from 0.028 to 0.031.

Thus, compared to empirical evidence on the unhedged carry trade payoff, my model delivers mean returns 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 about a third of its empirical counterpart. Naturally, the model is characterized by higher Sharpe ratios than in the data due to the considerably lower standard deviation. The average Sharpe ratio for the model is 0.16 while empirically BEKR find an average value around 0.1.

Figure 5 plots the histogram of the realized unhedged carry trade returns obtained for $N = 1000$ samples of size $T = 300$. The mean of these returns is positive and compared to a normal distribution with the same mean and variance the returns are negatively skewed and have significant excess kurtosis.

5.2 Negative skewness and excess kurtosis of unhedged carry trade returns

The results reported in Table 7.A. indicate that the returns to the carry trade are on average negatively skewed. The degree of skewness is slightly larger than the one found in the data for the carry trade returns analyzed in Table 4. The model implies a negative skewness of -0.42 while the average for the countries in Table 4 is -0.24. BEKR report an average for the individual country pairs of around -0.26.

To investigate the properties of the realized skewness of excess returns I construct the following tests. The first, more cross-sectional in nature, is similar to that of BNP. It involves checking whether periods (countries in BNP) characterized by a higher domestic currency also experience a negative skewness in the excess returns. To that end, I simulate the model for $T = 300$ and for each t collect r_t and the realized $ex_{t+1} = r_t - (s_{t+1} - s_t)$. I sort the excess returns ex_{t+1} according to the sign of r_t . Denote by ex_{t+1}^+ the returns when $r_t > 0$ and by ex_{t+1}^- when $r_t < 0$. Consistent with the predictability of excess returns 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 (-0.4) and that of ex_{t+1}^- is positive (0.4).

A test for a time varying dimension is to simulate the model and at each date t to have $N = 20000$ draws from the DGP process for the date $t + 1$ realizations of r_{t+1} . This is similar to the numerical procedure used to solve the distorted model but here the draws are from the DGP and not from the subjective distribution. 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 returns $ex_{t+1} = r_t - (s_{t+1} - s_t)$ 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. C reports the results of the regression

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

The model generates a negative significant $\hat{\beta}_2$ 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, this implies 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 smaller reaction to positive shocks the excess returns are negatively skewed. Thus “crash risk ” in this model is endogenous and it happens when negative shocks hit an otherwise positive estimate of the hidden state.

In this model agents at time t expect negatively skewed returns at $t + 1$. They understand that the pricing at $t + 1$ will involve asymmetric response to news which generates negative skewness. Given that in this model expected excess returns are driving most of the portfolio decision skewness per se does not seem to affect this decision. It does matter indirectly in forming expectations about the excess returns.

Excess returns under the distorted model are always expected to be negatively skewed and under the DGP they will in fact be. Their skewness under the DGP depends on the size of the interest rate differential. Consider the following scenario. At time t there is a high positive domestic interest rate differential. That means that the expected exchange rate tomorrow will also have a higher absolute level, although lower than the time t one by UIP. If the innovations $r_{t+1} - H'F\hat{x}_{t,t}^{t+1}$ in the Kalman Filter at time $t + 1$ are negative, they are incorporated with a high gain while if they are positive with a low gain. As explained above because $H'F\hat{x}_{t,t}^{t+1}$ is a downward biased estimate of the DGP $E_t r_{t+1}$ these innovations will be positive on average. However, when negative they will strongly affect the estimate and the exchange rate.

To study the properties of the agents’ expectations, Jurek (2008) computes the risk-neutral implied

skewness from options data. In my setup that corresponds to the object $Skew_{t,t+1}^{\tilde{P}}$, defined in Section 4.3. It is worth noting again that the essential difference between the risk-neutral and the realized probabilities is not the risk correction from marginal utilities, but the distorted probabilities. This is at the heart of the mechanism of the paper.

Table 7.C also reports results for the regression

$$Skew_{t,t+1}^{\tilde{P}} = \beta_3 r_t + \xi_{3,t+1}$$

The model implied $\hat{\beta}_3$ is also negative and very significant. Thus there is a negative correlation between the high interest rate differential and the skewness of the excess returns under the distorted model. Again, this happens because agents take into account the asymmetric response to news for the next period and thus expect a skewed distribution. This negative relation is also present when the test is cross-sectional, as with the realized skewness. Jurek (2008) also confirms this cross-sectional implication, i.e. periods of higher domestic rate are also characterized on average by a negative risk-neutral skewness. Interestingly, comparing the point estimates of $\hat{\beta}_2$ and $\hat{\beta}_3$ a higher r_t has a larger conditional impact on the predicted realized skewness than on the risk-neutral skewness.

5.3 Positive hedged carry trade returns and options

The hedged carry trade is characterized by no downside risk. To measure the hedged carry trade I use the definition in (31) and as in BEKR I use at-the-money options by setting $k_t = s_t$. The results are very similar with another interpretation of at-the-moneyness, namely $k_t = f_t$. Table 7. B. presents the results for the hedged carry trade. The average payoff is around 0.0012 which is smaller than the payoff for the unhedged carry trade. BEKR also find this relation in the data. The standard deviation of the hedged carry trade is lower and the resulting Sharpe ratio is slightly higher than for the unhedged carry trade.

For intuition, take the case of the foreign currency being at a forward premium, i.e. $r_t > 0$ and $s_t < 0$. Buying a call option on the FCU is relatively expensive because the probability of exercising the option is relatively large. By using the distorted model, the agent believes that there will be a more depreciated domestic currency than what actually happens at time $t + 1$. This intuitively means that buying a call option delivers losses ex-post. If one compares the average means of the payoffs, the average loss from buying the insurance options is about 30% of the unhedged payoff consistent with findings in Jurek (2008).³⁷ BEKR report values that imply that the average individual country pair loss resulting from eliminating the downside risk is about 20% of the unhedged payoff. Because these losses are not that large the hedged carry trade still has positive average returns. There are several reasons for this.

First, there is an interesting effect that mitigates the increase in price. At the time of pricing agents understand the equilibrium conditions and realize that next period the “inaction effect” can take place. This effect, defined in Section 4.2, means that if the estimate of the hidden state at time $t + 1$ is close

³⁷Jurek (2008) uses a more complicated version of hedging that also takes into account the delta exposure of the option.

enough to 0, agents will choose not to invest at all and $s_{t+1} = 0$. Since at time t the estimate of the hidden state is positive agents foresee that there are innovations at time $t + 1$ that are negative enough to translate into negative estimates at time $t + 1$. In the rational expectations world, these negative estimates would immediately imply a $s_{t+1} > 0$ and very large depreciation of the foreign currency. However, in the present model these situations are likely to generate the “inaction effect” because agents are cautious seeing this switch in the sign of the estimate. As such, these innovations would generate $s_{t+1} = 0$ instead of their positive rational expectations counterpart.

Second, for the cases in which the “inaction effect” is not active, the large depreciations that the agent worries about will mostly still occur in equilibrium. When negative news arrives, it will be strongly incorporated in the exchange rate both under the objective and subjective probability distribution. The key difference between the agent’s ex-ante probability distribution and the realized one is that the average exchange rate is slightly less depreciated in the latter. This lower average is caused by a slight shift of mass of exchange rate realizations that are around the mean of the subjective distribution to values that imply an appreciated currency. At time $t + 1$ the agent will receive on average positive innovations about the interest rate which cause the exchange rate realizations to be tilted toward appreciation. However, even under this objective distribution, when negative news arrive there is a strong reaction of the exchange rate toward depreciation. Thus a strong negative skewness still remains under the realized distribution. As noted in Section 5.2 a higher interest rate differential predicts a larger skewness for the realized distribution than for the risk-neutral one.

A way to think about the reason why the model generates positive average payoffs both for unhedged and hedged strategies is to note that the subjective probability distribution differs from the objective one by overpredicting bad events and underpredicting good events. Buying insurance against the downside risk produces on average losses because of the exaggeration of the probability of bad states under the subjective probability distribution for the exchange rate realizations. However, the hedged position delivers positive payoffs because of the more frequent occurrence of the good events under the objective probability distribution than under the equilibrium distorted beliefs.

The theory proposed here does not rely on a peso event explanation in the sense used in the literature.³⁸ Such an event is interpreted to be one that does not materialize in the sample but is characterized by some extreme negative effect on the agent’s utility. As BEKR point out that this extreme negative event could be a large loss from the strategy or a very high marginal utility when an otherwise small loss happens. To see the reasons why the peso problem can explain the positive payoffs to the unhedged carry trade notice that ex-ante the investor has to be compensated for this negative possibility. When this event does not happen in the sample then ex-post the payoffs look systematically positive. If the peso event is associated with a large loss this would necessarily mean that the hedged carry trade would have zero payoffs. In that case buying an option as insurance against downside risk should be producing systematic negative

³⁸This explanation has been put forward by recent papers such as Farhi and Gabaix (2008) for the UIP puzzle and Barro (2006) and Gabaix (2008) for the stock market. In their models there is a rare disaster that can happen with a small, usually time-varying probability. See also Gourio (2008) for a discussion of rare disasters models.

payoffs that outweigh exactly the gains from the unhedged carry trade. They then conclude that it must be the extremely high marginal utility that characterizes such a negative event. Clearly, in my model I cannot provide such a solution.

The peso explanation assumes that the only difference between the subjective and the objective distribution is the rare disaster that does not materialize in the sample under the latter distribution. In states of the world when there is no peso event the two distributions completely coincide. In my model this is not the case. First, there is no peso problem in this model in the sense that under the subjective distribution there is no event that entails some disaster for the agent compared to the support of the distribution. In fact, due to the “inaction effect” explained above, for some realizations that would imply significant domestic depreciation under the rational expectations model the exchange rate in the subjective distribution is less depreciated. Agents realize that enormous depreciations will not occur in equilibrium if they require agents switching positions from domestic to foreign bonds. Second, the mechanism presented here delivers a realized distribution for the exchange rates that can have a different shape from the subjective one. Different than the peso explanation, there is no reasons to expect that any property of these distributions is the same. Their mean differ thereby generating the ex-post profitability of the unhedged carry trade. They are both negatively skewed and have excess kurtosis. As explained above the model can deliver an explanation for the small losses of buying insurance against the downside risks and implicitly for the positive payoffs of the hedged strategy because of the properties of these distributions.

5.4 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) investigate the robustness of these identified VAR results. They conclude that the delayed overshooting result is sensitive to the recursive identification assumptions and that the peak of the exchange rate response is imprecisely estimated. Scholl and Uhlig (2006) uses sign restrictions and also finds evidence for delayed overshooting. For the country pairs he analyzes the estimated peak occurs within a year or two. Although they differ in their estimates of the length of

the delayed overshooting effect all 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 described in detail in Gourinchas and Tornell (2004). There they posit that if agents act on a certain type of distorted beliefs that can explain the conditional and unconditional UIP puzzles. The difference is that here I provide a model that explains the origin and optimality of such beliefs.

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, $r_{t-1} = 0$, $b_{t-1} = 0$, $E_{t-1}^{\tilde{P}} s_t = E_{t-1}^P s_t = 0$. At time t there is a positive shock to the interest rate differential. Given the setup of the model, this shock is a decrease in the foreign interest rate but the same conclusions hold if the foreign rate is treated as constant and the domestic rate would increase.

To allow for a clearer intuition I present the case in which the transition equation is an AR(1). The autocorrelation coefficient is equal to the largest eigenvalue of the matrix F in the benchmark parameterization (see Section 4.4). Thus, I use the DGP

$$\begin{aligned} r_t &= x_t + \sigma_V v_t \\ x_t &= \rho x_{t-1} + \sigma_U u_t \end{aligned} \tag{34}$$

where $\rho = 0.9718$. The values for σ_V and σ_U are the same as in Section 4.4. For these values, by applying the formulas for the rational expectations solution in (14) and (15) I find that $\Gamma^{RE} = -31.61$ and $\delta^{RE} = -0.9986$. As discussed in Section 4.2 and Section 4.4 for the distorted expectations solution the parameters Γ and δ are based on a slightly lower ρ . This is to reflect that the conjecture about the exchange rate evolution in the equilibrium with ambiguity aversion takes into account the asymmetric response to news. In the benchmark parameterization Γ and δ are found using (14) and (15) but with a $\tilde{\rho} = 0.9705$. Then $\Gamma = -30.28$ and $\delta = -0.9987$. Note that compared to rational expectations the agent uses the correct ρ in forming her estimates but that the exchange rate responds less to the same estimate.

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. Let $v_t = u_t = 1$. Then $r_t = \sigma_U + \sigma_V$ and $x_t = \sigma_U$. The next periods shocks are set equal to zero. The rational expectations solution is then: $s_t^{RE} = \Gamma^{RE} \hat{x}_{t,t}^{RE} + \delta^{RE} r_t$ and the distorted expectations solution is $s_t = \Gamma \hat{x}_{t,t}^t + \delta r_t$, where $\hat{x}_{t,t}^{RE}$ is the estimate of the hidden state x_t under rational expectations and $\hat{x}_{t,t}^t$ is the estimate under the ambiguity aversion model.

Figure 6 plots the dynamic response of the exchange rate to the observed increase in r_t . The rational expectations features the Dornbusch 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 that means 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. By underestimating the true hidden state she observes a higher than expected r_{t+1} . Her updated estimate at time $t + 1$ is:

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

In this setting $\widehat{x}_{t+1,t+1}^{t+1}$ is higher than $\widehat{x}_{t,t}^t$ and a further appreciation occurs. Because $r_t - \rho \widehat{x}_{t-1,t-1} > 0$ and using the notation I introduced in Section 4.1, $\sigma_{V,(t+1),t} = \sigma_{V,(t+1),t+1} = \sigma_V^H$. At time $t + 2$ if the agent can distort only the last 2 periods as in the benchmark specification then $\sigma_{V,(t+2),t} = \sigma_V$ and $\sigma_{V,(t+2),t+1} = \sigma_{V,(t+2),t+2} = \sigma_V^H$. This corresponds to setting $m = n$. If m is allowed to be significantly larger than n so that if the agent distorts 2 periods but chooses what exactly these are, then I find that $\sigma_{V,(t+2),t} = \sigma_{V,(t+2),t+2} = \sigma_V^H$ and $\sigma_{V,(t+2),t+1} = \sigma_V$. If there is no restriction on n , so that the agent can distort any period then $\sigma_{V,(t+2),t} = \sigma_{V,(t+2),t+1} = \sigma_{V,(t+2),t+2} = \sigma_V^H$. A similar argument applies for the future periods. Eventually the estimate of the hidden state converges to the rational expectations one.

The top plot in Figure 6 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 and $m = n$. There the peak occurs 2 periods later than in the rational expectations model. There is a gradual appreciation until period 3 followed by a depreciation after the estimates of the hidden state converge under the distorted and rational expectations. Note that for all cases in Figure 6 even after converging the exchange rate under ambiguity and rational expectations differ slightly because of the lower weight put on the estimate of the hidden state in the former. The second plot shows the evolution under the specification of $n = 2$ periods but $m = 10$. In that case the peak is 3 periods later. However these two specifications deliver very similar results. 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 17 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 on 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 generated likelihoods that become increasingly lower than the ones under the reference model.

The conclusion from Figure 6 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 because the agent is limited in distorting the time-varying precision of signals by statistical plausibility considerations. When these are absent, the model delivers significantly longer delayed overshooting.

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. In 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 in 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 1, 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 that imply 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 returns: 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 returns. The benchmark calibration implies positive but smaller and less variable excess returns 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 most often small excess returns. The model also implies equilibrium positive mean payoffs for the hedged positions due to the more frequent occurrence of good events for the investment strategy under the objective probability distribution than under the equilibrium distorted beliefs.

The theory proposed in this paper can be extended to other situations which involve signal extraction. For example, the model of rational inattention in Bacchetta and van Wincoop (2008) delivers underreaction to news in the foreign exchange market but symmetric responses to good and bad news. Investigating the link between information processing under ambiguity aversion and rational inattention seems a fruitful future direction of research. Another application could be analyzing the business cycle property that when a boom ends, the downturn is generally sharp and short but when growth resumes, the boom is more gradual.³⁹ The model proposed in this paper would produce an explanation for this observation intuitively similar to the delayed overshooting response: the boom is more gradual because agents tend to underreact to good news which implies that under their subjective probability distribution they continue to receive positive news for more periods than the rational expectations model predict. The downturn is abrupt because agents tend to overreact to bad news. A model in which an ambiguity averse agent has access to many signals about a common unobserved factor, as for example in the monetary policy model of Bernanke and Boivin (2003), would also have interesting implications. Compared to the rational expectations case, an ambiguity averse decision-maker that does not trust the precisions of these signals tends to place too much probability on signals that imply a lower expected utility.

³⁹This observation has been investigated in a rational expectations framework by Van Nieuwerburgh and Veldkamp (2006). They argue that more (less) production generates higher (lower) precision of information. Thus when the boom ends the precise estimates of the slowdown prompt a sudden recession. When growth resumes the low production yields noisy estimates of recovery which makes booms more gradual than downturns.

A Review of literature

In this appendix I review some of the literature on UIP, negative skewness of returns and robust filtering. In the literature on UIP, for my purposes, a fundamental assumption that distinguishes between different models is the assumption of rational expectations.

The theoretical literature on non-rational expectations and UIP is relatively scarce. Froot and Thaler (1990) and Lyons (2001) informally argue that models where agents are slow to respond to news may explain the UIP puzzle. The intuition is that if investors respond gradually to news about a higher domestic interest rate, there will be a continued reallocation of portfolios towards domestic bonds and an appreciation of the domestic currency following the positive shock. Bacchetta and van Wincoop (2008) offer a formalization of this intuition based on rational inattention. If information is costly to acquire and to process, some investors optimally choose to be inattentive and revise their portfolios infrequently.

An alternative way to obtain a gradual response to news is to consider the case in which the observable interest rate differential as a sum of hidden temporary and persistent components. If agents misperceive noise to signal ratio by attributing more probability to the temporary shocks than under the true data generating process this implies underreaction to news. Such a mechanism is discussed informally in Eichenbaum and Evans (1995) and in depth in Gourinchas and Tornell (2004). The latter posit this type of distortion and investigate some of its properties in exchange rate determination. Gourinchas and Tornell (2004) find that there are levels of misperception which are consistent with survey data that can explain the UIP puzzle. Their paper does not provide a model under which such a distortion would be optimal.

To my best knowledge, there are two papers that investigate such optimality. First, Tornell (2003) uses a mixed approach of optimal and robust control to show that unstructured uncertainty about the interest rate differential can deliver such a result. To explain the forward premium puzzle, one key assumption in his model is that the agents' degree of uncertainty aversion is decreasing through time. If this aversion is interpreted as a preference parameter as does the decision-theory literature it is hard to imagine why it would be decreasing. Even if it reflects the possibility that information on a market is becoming more precise as time passes it should imply that the puzzle should disappear in long samples. Second, Li and Tornell (2007) analyze a case in which uncertainty about the process is only allowed to affect the variance of the returns. In that situation, if and only if the agent is concerned about uncertainty in the observation equation of a state-space model, it is optimal for the agent to believe that the variance of the temporary shock is larger.

My model is related to this intuition as it analyzes ambiguity aversion for an agent that needs to estimate a hidden state. The agent is uncertain about the variances in the model, either in the observation or state equation. In fact, the model can be extended to uncertainty about the parameters. Independent

of the place where this uncertainty is assumed, the key general result is that the agent will choose to believe in a DGP that hurts her utility by making worse news for her investment strategy be more persistent and good news more temporary. She will overreact to bad innovations and underreact to good innovations. This still produces an explanation for the forward premium puzzle because the underreaction effect dominates for a case in which the true DGP features temporary shocks that are relatively small.

There have been several other approaches for explaining the UIP puzzle through expectational errors. For example, Frankel and Froot (1987) assume behavioral biases on the part of two types of investors, “chartists” and “fundamentalists” who have different horizons for holding assets. They find that this model can explain some of the myopic expectations apparent from survey data. Chakraborty and Evans (2008) study a model in which agents form expectations using econometric learning instead of being endowed with the objective probability distribution. They prove that discounted least-squares learning can produce an asymptotic UIP coefficient of 0. In simulations they find that for small samples the coefficient can become negative. Recently, Burnside et al. (2007) relax the assumption of frictionless markets and use a model with microstructure frictions in which there are informed and uninformed traders. They show that the adverse selection between market makers and traders, combined with some assumptions on the behavior of the uninformed agent, can explain the forward premium puzzle.

The larger part of the literature addressing the UIP puzzle has maintained the assumption of rational expectations. In such a model, there could be, broadly speaking, two explanations for the forward premium puzzle: time-varying risk premia and rational expectational errors. The time-varying risk premia literature is facing the challenge of producing extremely large and variable pricing kernels. As many have observed, a representative agent model with standard utility functions cannot generate large and variable risk premia (see Backus et al. (1993), Bekaert (1996), Canova and Marrinan (1993), Engel (1996), Lewis (1995)). Among the recent studies, Burnside et al. (2008) document that linear stochastic discount factors built from conventional measures of risk fail to explain the excess returns from currency speculation. They argue that this failure reflects the absence of a statistically significant correlation between the payoffs to the carry trade and traditional risk factors. Lustig and Verdelhan (2007) find, on the contrary, that real aggregate consumption growth risk can explain the excess returns on currency speculation. This disagreement prompted a debate between Burnside (2007) and Lustig and Verdelhan (2008).

Given the difficulties facing the time-varying risk premia approach, one strand of literature has investigated non-standard utility functions combined with assumptions about the time-variation of the data generating process. Time-variation in habit persistence is investigated in Bekaert (1996) and Verdelhan (2006). The former assumes conditionally heteroskedastic shocks while the latter pro-cyclical real interest rates. Bansal and Shaliastovich (2007) use Epstein-Zin preferences and rely on a perfect cross-country correlation among shocks to the long run components of consumption growth rates to reproduce the UIP puzzle. Using the rare disaster framework of Barro (2006), Farhi and Gabaix (2008) analyze whether time-variation in disaster risk can explain the UIP puzzle. They express their results in terms of a “resilience” variable which combines the effect of time-variation in the disaster probability, size of consumption disaster and size of individual country’s productivity in the disaster. When this variable is

mean-reverting they are able to explain the predictability of returns using a standard utility.

Another direction has been to study limited participation models which separate the consumption of the marginal investor from the aggregate consumption. Alvarez et al. (2008) examine an endogenously segmented asset market. In their model, higher money growth produces higher inflation which increases the agents participation in the asset market. This, in turn, decreases risk premia and can generate the forward premium anomaly.

In the rational expectations framework there could still be expectational errors coming from two sources: learning about a possible past shift in the distribution or forming expectations about a future shift. Lewis (1989) uses Bayesian learning about the probability of a past regime shift to provide an explanation for the forward premium puzzle. However, the model could not explain the persistence of prediction errors, since the magnitude of the error shrinks over time to zero. The second source of errors can be coming from expectations of future events that do not materialize in the sample. This is known as the peso problem. Kaminsky (1993) studies a model in which there are infrequent switches from contractionary to expansionary monetary policy. She provides evidence for the USD-GBP pair for 1976-1987 that investors' expectations are consistent with the model. Evans and Lewis (1995) show that a reasonably calibrated regime switching model similar to Engel and Hamilton (1989) can produce a negative coefficient in Fama regressions in small samples. These biases disappear, however, in large samples.

The asymmetric response implied by the optimal behavior is also providing an explanation for what has been called "currency crashes" (Brunnermeier et al. (2008)) or equivalently the negative skewness of currency excess returns. Brunnermeier et al. (2008) argue that realized skewness is related to the rapid unwinding of currency positions, precipitated by shocks to funding liquidity. Brunnermeier and Pedersen (2008) show theoretically that securities that speculators invest in have a positive average return, which is a premium for providing liquidity, and a negative skewness. The latter arises from an asymmetric response to fundamental shocks: shocks that lead to losses are amplified when speculators hit funding constraints and unwind their positions, further depressing prices and entering into a downward "liquidity spiral". Conversely, shocks that lead to speculator gains are not amplified. In a related paper, Plantin and Shin (2008) provide a game-theoretic motivation of how strategic complementarities can endogenously generate currency crashes. Their model is silent on the source of the UIP failure because in equilibrium there are no premia to motivate excess returns. Veronesi (1999) provides a model of rational expectations in which there is overreaction to bad news in good states and underreaction to good news in bad times. The key ingredient is the assumption of two hidden states so that the conditional variance of future realizations is affected by the current observation. For the typical Kalman filter used also in my model this conditional variance is deterministic and converges to a steady state.

Hansen and Sargent (2008, Ch.17) study a robust estimation problem with commitment to previous distortions in which the agent wants to minimize the estimation mean square error but is faced with uncertainty in both the observation and transition equation. They find that the robust filter flattens the decomposition of variances across frequencies by accepting higher variances at higher frequencies

in exchange for lower variances for lower frequencies. This implies that a preference for robustness increases the Kalman gain. Intuitively, the agent is more concerned about the model being misspecified at low-frequencies. Hansen and Sargent (2005) analyze more general settings of robust estimation with commitment.

Hansen and Sargent (2007b) study robust estimation problems without commitment. They use multiplier preferences to analyze unstructured uncertainty. Hansen and Sargent (2007a) investigate a setup where the hidden state comprises a discrete set of models and the agent slants pessimistically the posterior probabilities over these models. My model also studies a robust estimation problem without commitment but focuses on structured uncertainty. I analyze a type of uncertainty similar to Epstein and Schneider (2007, 2008) who investigate learning under ambiguity about a constant parameter. They show that if there is only ambiguous prior information about the parameter then with repeated signals about the parameter this uncertainty disappears in the long run. They point out that if there is ambiguity about the precision of the signal then the agent in the long-run is still uncertain about future realizations of the signals. In their model agents underreact to good news and overreact to bad news. I extend that setup to time-varying hidden states.

B Rational expectations model solution

In this section I provide some details for the rational expectations model solution. The optimization problem is

$$\max_{b_t} E_t[-\exp(-\gamma b_t q_{t+1})]$$

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

$$E_t[q_{t+1} \exp(-\gamma b_t q_{t+1})] = 0$$

Use the approximation of $q_{t+1} \simeq \exp(q_{t+1}) - 1$ and a second order approximation around $q_{t+1} = 0$ to the function $[\exp(q_{t+1}) - 1] \exp(-\gamma b_t q_{t+1})$ which delivers:

$$0 = q_{t+1} - \gamma b_t q_{t+1}^2 + o(3)$$

where $o(3)$ is the error term up to order 3. If, as in the RE setup, the variable q_{t+1} is log-normally distributed then the third order error term equals 0. Thus, I get that the FOC implies the mean-variance solution as in (8):

$$b_t = \frac{E_t(q_{t+1})}{\gamma \text{Var}_t(q_{t+1})}$$

Let the guess about the exchange rate be

$$s_t = \Gamma \hat{x}_{t,t} + \delta r_t$$

where $\hat{x}_{t,t} = E(x_t | I_t)$. The filtering notation for the RE case was introduced in Section 3. The corresponding Kalman filter recursion correspond to equations (10), (11) and (12). For simplicity I

assume convergence on the Kalman gain ($K_t = K$) and the estimate of the covariance matrix of the hidden state ($\Sigma_{t,t} = \Sigma$). Then:

$$\begin{aligned} E_t(s_{t+1}|I_t) &= (\Gamma + \delta H')F\hat{x}_{t,t} \\ Var_t(s_{t+1}|I_t) &= (\Gamma K + \delta)(\Gamma K + \delta)'Var_t(r_{t+1}|I_t) \end{aligned}$$

where the conditional variance of the observed interest rate differential is the time invariant

$$Var(r_{t+1}|I_t) = H'F\Sigma F'H + H'\sigma_U\sigma_U'H + \sigma_V^2 \quad (35)$$

Denoting $Var_t(s_{t+1}|I_t)$ by σ^2 the UIP condition (9) states

$$s_t = \frac{(\Gamma + \delta H')F\hat{x}_{t,t} - r_t}{1 + (\gamma\sigma^2/2)}$$

Now, I verify the conjecture in (9) and solve for the unknown coefficients Γ, δ :

$$\delta = -\frac{1}{1 + (\gamma\sigma^2/2)} \quad (36)$$

$$\Gamma = -\frac{1}{1 + (\gamma\sigma^2/2)}H'F[(1 + (\gamma\sigma^2/2))I - F]^{-1} \quad (37)$$

The variance of the exchange rate is

$$\sigma^2 = (\Gamma K + \delta)(\Gamma K + \delta)'Var(r_{t+1}|I_t)$$

with $Var(r_{t+1}|I_t)$ given by (35), so there are 3 equations in 3 unknowns δ, Γ, σ^2 .

Table 1. Rational expectations model

Risk aversion	δ	Γ	$std_t(s_{t+1})$	$\widehat{\beta}^L$
$\gamma = 2$	-0.9994	-38.14	0.0235	0.9784
$\gamma = 10$	-0.9976	-35.5	0.0219	0.9125
$\gamma = 50$	-0.992	-29.15	0.0181	0.7535
$\gamma = 500$	-0.971	-17.25	0.0109	0.455

These results refer to the discussion in Section 3 and are obtained for an AR(1) representation of the USD-GBP interest rate differential for the period 1976-2007 for which $std_t(r_{t+1}) = 0.0006$, $\rho = 0.97$. The empirical standard deviation for this sample of the exchange rate is $std_t(s_{t+1}) = 0.029$.

C Multiplier preferences and unstructured uncertainty

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

$$\max_b \min_{\tilde{P} \in \Phi} E^{\tilde{P}}[U(c(b; \varepsilon))] + \theta R(\tilde{P}/P) \quad (38)$$

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 $R(\tilde{P}/P)$ is the relative entropy of probability measure \tilde{P} with respect of measure P :

$$R(\tilde{P}/P) = \begin{cases} \int_{\Omega} \log\left(\frac{d\tilde{P}}{dP}\right) d\tilde{P} & \text{if } \tilde{P} \text{ is absolutely continuous w.r.t } P \\ \infty & \text{otherwise} \end{cases} \quad (39)$$

Hansen and Sargent (2008) 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) and Barillas et al. (2008) the problem in (38) is equivalent to⁴⁰:

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

Note that now the subjective probability distribution becomes the reference model P . Thus the problem is restated in terms of maximizing an expected utility, under this reference model, with a higher risk aversion. In the present case, where $U(c(b; \varepsilon)) = -\exp(-\gamma W_{t+1})$ using the multiplier preferences would transform the utility to an exponential of the original utility, with the maximization being performed under the reference model P .

$$\max_b E^P \left[-\exp\left(\frac{1}{\theta} (\exp(-\gamma W_{t+1}))\right) \right]$$

To illustrate further these implications consider the investment problem studied in this paper under risk neutrality and unstructured uncertainty so that $U(c(b; \varepsilon)) = b_t(s_{t+1} - s_t - r_t)$. For simplicity, assume also that the transition equation is an AR(1). As in Hansen and Sargent (2007b) consider the problem in which the agent takes the process in (1) as an approximated model and she surrounds it with a set of alternative models such as:

$$\begin{aligned} r_{t+1} &= x_{t+1} + \sigma_V v_{t+1} + \epsilon_{t+1}^V \\ x_t &= Fx_{t-1} + \sigma_U u_t + \epsilon_t^U \end{aligned}$$

The shocks ϵ_t^V and ϵ_t^U 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 $N(Fx_t + \epsilon_{t+1}^V + \epsilon_{t+1}^U, \sigma_U^2 + \sigma_V^2)$. Under RE, the hidden state x_t is distributed $N(\hat{x}_{t,t}, \Sigma_{t,t})$. In this setting, Hansen and Sargent (2007b) propose two robustness corrections: one that distorts r_{t+1} , conditional on x_t , through the mean of $(\epsilon_{t+1}^V + \epsilon_{t+1}^U)$ and another that distorts the distribution over the hidden state. Hansen and Sargent (2007b) 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 (1). Hence x_t is distributed as $N(\hat{x}_{t,t}^{RE} + \varepsilon_t, \Sigma_{t,t})$ where $\hat{x}_{t,t}^{RE}$ is the estimate under Kalman Filter for (1) and ε_t is the arbitrary unknown conditional mean distortion of the hidden state.

⁴⁰The equivalence is true in a Savage setting. This result is well known in the decision theory literature. 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.

These alternative models are constrained to be close to the approximating model by using the conditional relative entropy defined in (39) above. After taking into account these distortions the maximization occurs under the transformed conditional distribution

$$r_{t+1} \stackrel{\tilde{P}}{\sim} N(F\hat{x}_{t,t}^{RE} + F\varepsilon_t + \epsilon_{t+1}^V + \epsilon_{t+1}^U, F^2\Sigma_{t,t} + \sigma_U^2 + \sigma_V^2). \quad (40)$$

Note that under the reference model

$$r_{t+1} \stackrel{P}{\sim} N(F\hat{x}_{t,t}^{RE}, F^2\Sigma_{t,t} + \sigma_U^2 + \sigma_V^2) \quad (41)$$

The relative entropy defined in (39) for the implied distributions \tilde{P} in (40) and P in (41) is

$$R(\tilde{P}/P) = \frac{(F\varepsilon_t + \epsilon_{t+1}^V + \epsilon_{t+1}^U)^2}{2(F^2\Sigma_{t,t} + \sigma_U^2 + \sigma_V^2)}$$

Denote the overall distortion $F\varepsilon_t + \epsilon_{t+1}^V + \epsilon_{t+1}^U$ by ω_{t+1} and use that $Var_t^P(r_{t+1}) = F^2\Sigma_{t,t} + \sigma_U^2 + \sigma_V^2$. The multiplier preferences defined in (38) can then be applied here to maximize over b_t and minimize over ω_{t+1} :

$$\max_{b_t} \min_{\omega_{t+1}} E_t^{\tilde{P}}[1 + b_t(s_{t+1} - s_t - r_t)] + \theta \frac{\omega_{t+1}^2}{2Var_t^P(r_{t+1})} \quad (42)$$

To solve for an equilibrium in this setup use a guess and verify approach and conjecture that the solution for s_{t+1} is

$$s_t = \Gamma\hat{x}_{t,t}^{RE} + \delta r_t$$

with unknown coefficients Γ and δ . Then $E_t^{\tilde{P}}(s_{t+1}) = \Gamma E_t^{\tilde{P}}(\hat{x}_{t+1,t+1}^{RE}) + \delta E_t^{\tilde{P}}(r_{t+1})$. Using the fact that the estimate at time $t + 1$ is formed by the Kalman filter updating formulas, I get:

$$\begin{aligned} E_t^{\tilde{P}}(s_{t+1}) &= \Gamma F(1 - K)\hat{x}_{t,t}^{RE} + (\Gamma K + \delta)E_t^{\tilde{P}}r_{t+1} \\ &= (\Gamma + \delta)F\hat{x}_{t,t}^{RE} + (\Gamma K + \delta)\omega_{t+1} \end{aligned}$$

Replacing $E_t^{\tilde{P}}(s_{t+1})$ in (42) and taking the FOC with respect to ω_{t+1} I obtain

$$\omega_{t+1} = -\frac{Var_t^P(r_{t+1})}{\theta} b_t (\Gamma K + \delta)$$

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

$$s_t = [1 + (\Gamma K + \delta)^2 \frac{Var_t^P(r_{t+1})}{2\theta}]^{-1} [(\Gamma + \delta)F\hat{x}_{t,t}^{RE} - r_t]$$

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

$$\Gamma = [1 + \frac{Var_t^P(s_{t+1})}{2\theta}]^{-1} (\Gamma + \delta) F \quad (43)$$

$$\delta = -[1 + \frac{Var_t^P(s_{t+1})}{2\theta}]^{-1} \quad (44)$$

The solution to (43) and (44) is identical to the one in (37) and (36) when $\gamma = \theta^{-1}$. Hence distorting the conditional mean of the interest rate differential process by considering unstructured uncertainty in the multiplier preferences with the uncertainty aversion parameter θ produces the same solution as solving the model under rational expectations and having a risk averse agent with an absolute rate of risk aversion of θ^{-1} .

D Distorted expectations model equations

Proposition 1. Expected excess return, $E_t^{\tilde{P}} W_{t+1}$ is monotonic in $\sigma_{V,(t),t}^2$. The sign is given by the sign of $b_t(r_t - H'F\hat{x}_{t-1,t-1}^t)$.

Proof: The Perceived Law of Motion is given by (23)

$$s_{t+1} = \Gamma \hat{x}_{t+1,t+1}^{t+1} + \delta r_{t+1}$$

By the Kalman filter

$$\hat{x}_{t+1,t+1}^{t+1} = F \hat{x}_{t,t}^{t+1} + K_{t+1}^{t+1}(r_{t+1} - H'F \hat{x}_{t,t}^{t+1})$$

For more intuition and easier derivation consider the example if \hat{x}_t is a vector of (1×1)

Then, taking as given $\hat{x}_{t,t}^{t+1}$ and noting that K_{t+1}^{t+1} is positive, s_{t+1} is monotone in r_{t+1} . In equilibrium $\Gamma K_{t+1}^{t+1} + \delta < 0$, so s_{t+1} is a decreasing function of r_{t+1}

$$\frac{\partial E_t^{\tilde{P}}(s_{t+1})}{\partial E_t^{\tilde{P}} r_{t+1}} < 0$$

For the Kalman gain

$$\begin{aligned} K_t^t &= (F \Sigma_{t-1,t-1}^t F' + \sigma_U^2) [F \Sigma_{t-1,t-1}^t F' + \sigma_U^2 + \sigma_{V,(t),t}^2]^{-1} \\ \frac{\partial K_t^t}{\partial \sigma_{V,(t),t}^2} &= -(F \Sigma_{t-1,t-1}^t F' + \sigma_U^2) [F \Sigma_{t-1,t-1}^t F' + \sigma_U^2 + \sigma_{V,(t),t}^2]^{-2} < 0 \end{aligned} \quad (45)$$

From the Kalman Filter formulas $E_t^{\tilde{P}} r_{t+1} = H'F \hat{x}_{t,t}^t$. and $\hat{x}_{t,t}^t$ is a function of the sequence $\sigma_V(r^t)$

$$\hat{x}_{t,t}^t = F \hat{x}_{t-1,t-1}^t + K_t^t(r_t - H'F \hat{x}_{t-1,t-1}^t)$$

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

$$\begin{aligned} \frac{\partial E_t^{\tilde{P}} W_{t+1}}{\partial \sigma_{V,(t),t}^2} &= \frac{\partial E_t^{\tilde{P}} W_{t+1}}{\partial E_t^{\tilde{P}} s_{t+1}} \frac{\partial E_t^{\tilde{P}}(s_{t+1})}{\partial E_t^{\tilde{P}} r_{t+1}} \frac{\partial \hat{x}_{t,t}^t}{\partial \sigma_{V,(t),t}} \\ \text{sign}\left(\frac{\partial E_t^{\tilde{P}} W_{t+1}}{\partial \sigma_{V,(t),t}^2}\right) &= \text{sign}(b_t(r_t - F \hat{x}_{t-1,t-1}^t)) \end{aligned}$$

This establishes *Proposition 1*.

Proposition 2. The expected variance of excess return, $Var_t^{\tilde{P}} W_{t+1}$ is increasing in $\sigma_{V,(t),t}^2$.

Proof: The variance $Var_t^{\tilde{P}} W_{t+1} = b_t^2 Var_t^{\tilde{P}} s_{t+1}$. In turn, using the conjecture (23) and taking as given $\hat{x}_{t,t}^{t+1}, K_{t+1}^{t+1}$

$$Var_t^{\tilde{P}} s_{t+1} = (\Gamma K_{t+1}^{t+1} + \delta)(\Gamma K_{t+1}^{t+1} + \delta)' Var_t^{\tilde{P}} r_{t+1}$$

By the Kalman filtering formulas I get

$$Var_t^{\tilde{P}} r_{t+1} = H' F \Sigma_{t,t}^t F' H + H' \sigma_U \sigma_U' H + E_t^{\tilde{P}}(\sigma_{V,(t),t}^2)$$

Use the formula in (23) for the Kalman gain and the recursion:

$$\Sigma_{t,t}^t = \Sigma_{t,t-1}^t - K_t H' \Sigma_{t,t-1}$$

It is easy to see that if \hat{x}_t is a vector of (1×1) :

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

Thus,

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

This establishes *Proposition 2*.

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

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

The sign of this derivative is :

$$sign\left(\frac{\partial V_t}{\partial \sigma_{V,(t),t}^2}\right) = sign(b_t) sign(r_t - F \hat{x}_{t-1,t-1}^t) - sign\left(\frac{\partial Var_t^{\tilde{P}} r_{t+1}}{\partial \sigma_{V,(t),s}^2}\right)$$

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

$$\begin{aligned} \frac{\partial V_t}{\partial \sigma_{V,(t),t}^2} &= \frac{\partial V_t}{\partial E_t^{\tilde{P}} s_{t+1}} \frac{\partial E_t^{\tilde{P}} s_{t+1}}{\partial E_t^{\tilde{P}} r_{t+1}} \frac{\partial E_t^{\tilde{P}} r_{t+1}}{\partial \sigma_{V,(t),t}^2} + \frac{\partial V_t}{\partial Var_t^{\tilde{P}} s_{t+1}} \frac{\partial Var_t^{\tilde{P}} s_{t+1}}{\partial Var_t^{\tilde{P}} r_{t+1}} \frac{\partial Var_t^{\tilde{P}} r_{t+1}}{\partial \sigma_{V,(t),t}^2} \\ &= -(\Gamma K_{t+1}^{t+1} + \delta) F b_t (r_t - F \hat{x}_{t-1,t-1}^t) (F \Sigma_{t-1,t-1}^t F' + \sigma_U^2) [F \Sigma_{t-1,t-1}^t F' + \sigma_U^2 + \sigma_{V,(t),t}^2]^{-2} + \\ &\quad + (1 - \gamma) (\Gamma K_{t+1} + \delta)^2 b_t^2 F^2 [F^2 \Sigma_{t-1,t-1}^t + \sigma_U^2]^2 [F^2 \Sigma_{t-1,t-1}^t + \sigma_U^2 + \sigma_{V,(t),t}^2]^{-2} \end{aligned}$$

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

$$\Pr\left[\frac{\partial V_t}{\partial \sigma_{V,(t),t}^2} > 0 \mid (b_t \xi_t > 0)\right] = \Pr[b_t (r_t - F \hat{x}_{t-1,t-1}^t) > (1 - \gamma) (\Gamma K_{t+1} + \delta) F (F^2 \Sigma_{t-1,t-1}^t + \sigma_U^2) \mid (b_t \xi_t > 0)]$$

which equals

$$\Pr b_t \xi_t > (1 - \gamma)(\Gamma K_{t+1} + \delta) F b_t^2 (F^2 \Sigma_{t-1,t-1}^t + \sigma_U^2) (F^2 \Sigma_{t-1,t-1}^t + \sigma_U^2 + \sigma_{V,(t),t-1}^2)^{-0.5} | (b_t \xi_t > 0)]$$

To see more clearly the extremely low probability that the variance effect dominates imagine that there are no observation shock. This delivers the upper bound on the probability. Then $K_{t+1} = 1$, $\Sigma_{t-1,t-1}^t = 0$ and $\sigma_{V,(t),s}^2 = 0$. In this case I have:

$$\Pr(b_t \xi_t > (1 - \gamma)(\Gamma + \delta) F b_t^2 \sigma_U | (b_t \xi_t > 0))$$

If $b_t > 0$ it becomes

$$\Pr(\xi_t > (1 - \gamma)(\Gamma + \delta) F b_t \sigma_U | (\xi_t > 0))$$

In the benchmark AR(1) simulation $F = \rho = 0.97$, $\Gamma + \delta = \frac{\rho}{\rho-1} = -32$, $1 - \gamma = -9$, $\sigma_U = 0.0006$. If $b_t = 1$ the RHS would be a value around 0.167. The probability would then be 0.066. In simulations b_t has the same magnitude as $.5s_t$, whose standard deviation in the model is 0.03. Thus in the model, the above probability is basically 0. A similar calculation applies for $b_t < 0$.

D.1 Restrictions on distorted sequences

In Section 2.3 I introduce the constant n as the number of dates at which the agent is distorting a sequence of previous variances of the observation shock. This constant can be interpreted as the number of times the agent is willing to entertain that the variance of the observation shock differs from the reference model. It is thus independent of the sample size. However, it can also be thought of as the probability that for a given date this different value occurs. This probability interpretation has an important implication for the expected value of the standard deviation at date $t + 1$, $E_t^{\tilde{P}} \sigma_{V,t+1}$. Indeed,

$$E_t^{\tilde{P}} \sigma_{V,t+1} = \Pr(\sigma_{V,t+1} = \sigma_V^H) \sigma_V^H + \Pr(\sigma_{V,t+1} = \sigma_V^L) \sigma_V^L + \Pr(\sigma_{V,t+1} = \sigma_V) \sigma_V$$

I choose to consider the case in which the agent treats the expected value about the future in a way that is consistent with her view about the past. The agent believes that that $\sigma_{V,s} \neq \sigma_V$ only for n out of t times in the sample of observable data she has.⁴¹ Thus, in forming $E_t \sigma_{V,t+1}$, the agent uses $\Pr(\sigma_{V,t+1} = \sigma_V^H) + \Pr(\sigma_{V,t+1} = \sigma_V^L) = n/t$. There are many combinations of these probabilities that satisfy this criterion. If the object matters for the utility of the agent, the uncertainty aversion principle dictates that she would choose the value that makes her worse off. In this model, where agents dislike the variance of the excess returns, they will choose at time t based on the sample size of t to believe that next periods realizations for the variance $\sigma_{V,t+1}$ are equal to σ_V with probability $1 - n/t$ and equal to σ_V^H with probability n/t .

⁴¹The only important implication that this assumption has is on the pricing of options. Allowing for $E_t^{\tilde{P}} \sigma_{V,t+1}$ to equal σ_V^H would result in extremely expensive options which we do not see in the data.

For simplicity, I will not keep track specifically of this choice when referring to the minimization part in (7). When this object appears in the forecast I make use of the above argument and conclude that

$$E_t^{\tilde{P}} \sigma_{V,t+1} = \sigma_V^H \frac{n}{t} + \sigma_V \left(1 - \frac{n}{t}\right) \quad (46)$$

Because this object will have a very small influence on the results and I could directly consider the case of t large, it is inconsequential to use $E_t^{\tilde{P}} \sigma_{V,t+1} = \sigma_V$.

E Benchmark parameterization

The state space for the true DGP is defined in (32) as

$$\begin{aligned} r_t &= H'x_t + \sigma_V v_t \\ x_t &= Fx_{t-1} + \sigma_U u_t \end{aligned}$$

where r_t is the observable interest rate differential, x_t is a (4×1) vector, $H' = [1 \ 0 \ 0 \ 0]$, $\sigma_U = [\sigma_U^{DGP} \ 0 \ 0 \ 0]$ and F is a (4×4) matrix:

$$F = \begin{bmatrix} \rho_1 & \rho_2 & \rho_3 & \rho_4 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

The benchmark parameterization is:

Table 2. Benchmark specification

σ_V	σ_V^H	σ_V^L	σ_U^{DGP}	ρ_1	ρ_2	ρ_3	ρ_4	γ
0.00025	0.0037	0.000166	0.0005	1.54	-0.52	-0.34	0.3	10

The implied steady state Kalman gain for the true DGP, low precision and high precision signal is $K^{DGP} = [0.8584 \ 0.1735 \ -0.0177 \ -0.0411]'$, $K^H = [0.166 \ 0.14 \ 0.108 \ 0.0779]'$ and $K^L = [0.9989 \ 0.0017 \ 0 \ 0]'$.

As discussed in Section 2.3 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 and under the distorted optimal sequence. To compute the log-likelihood I use the Kalman Filter. For a given deterministic sequence $\sigma_V^t = \{\sigma_{V,s}, s \leq t\}$ the filter generates the objects

$$\begin{aligned} \Sigma_{s,s-1} &= F\Sigma_{s-1,s-1}F' + \sigma_U\sigma_U' \\ K_s &= \Sigma_{s,s-1}H(H'\Sigma_{s,s-1}H + \sigma_{V,s}^2)^{-1} \\ \hat{x}_{s,s} &= F\hat{x}_{s-1,s-1} + K_s(r_s - H'F\hat{x}_{s-1,s-1}) \\ \Sigma_{s,s} &= \Sigma_{s,s-1} - K_sH'\Sigma_{s,s-1} \end{aligned}$$

Let Ω_s denote the variance $H'\Sigma_{s,s-1}H + \sigma_{V,s}^2$. The likelihood for the observation r_s conditional on the history of observables r^{s-1} is:

$$f(r_s/r^{s-1}) = (2\pi)^{-.5} \det(\Omega_s)^{-.5} \exp(-.5((r_s - H'F\hat{x}_{s-1,s-1})'(\Omega_s)^{-1}(r_s - H'F\hat{x}_{s-1,s-1}))$$

The sample log likelihood is then the sum $\sum_{i=1}^t \log f(r_s/r^{s-1})$. Let the log-likelihood generated by the reference sequence $\sigma_V^t = \{\sigma_{V,s} = \sigma_V, s \leq t\}$ be denoted by $L^{DGP}(r^t)$ and the one generated by the equilibrium optimal sequence $\sigma_V^t = \sigma_V^*(r^t)$ using the parameterization in Table 2 be denoted by $L^{Dist}(r^t)$.

Table 3 Likelihood comparison: $L^{Dist}(r^t) - L^{DGP}(r^t)$

	(1)	(2)	(3)
<i>Mean</i>	-1.48 (0.06)	-1.39 (0.05)	-138 (3.2)
<i>St.dev.</i>	1.13 (0.03)	1.06 (0.03)	3.2 -
<i>%positive</i>	0.167 (0.021)	0.178 (0.027)	0 -

Table 3 reports the mean difference and the standard deviation of $L^{Dist}(r^t) - L^{DGP}(r^t)$ for a sample of $T = 300$. It also reports the percent of times in which $L^{Dist}(r^t) - L^{DGP}(r^t) > 0$. This is referred to in the model as the detection error probability. The standard errors for each statistics across $N = 1000$ simulations is reported in the parentheses. Column (1) is the benchmark specification in which $m = n = 2$ so that the agent distorts the reference sequence only for the last 2 periods. In Column (2) $n = 2$ but $m = 10$ so that the agent distorts 2 out of the last 10 periods. In Column (3) $n = t$ so the agent can distort any period in the sample. In this case the difference $L^{Dist}(r^t) - L^{DGP}(r^t)$ is increasing with the sample size so I report the mean difference for $t = T = 300$: $L^{Dist}(r^T) - L^{DGP}(r^T)$. The standard deviation is then computed across the $N = 1000$ simulations.

F Time-varying parameters

In this section I analyze a setup with time-varying parameters and point out that the same intuition applies. Consider the setup

$$r_t = \rho_t r_{t-1} + v_t$$

$$\rho_t = \rho_{t-1} + u_t$$

where r_t is the observable interest rate differential, ρ_t is the hidden parameter driving its persistence. The shocks u_t and v_t are independent and distributed normally with standard deviations σ_U and time-varying $\sigma_{V,t}$ respectively. The agent entertains the possibility that these realizations are draws from the set $\Upsilon = \{\sigma_V^L, \sigma_V, \sigma_V^H\}$.

Consider a similar framework for the two country general equilibrium model but for simplicity assume risk neutrality for the per-period felicity function. The UIP condition then states: $s_t = E_t^{\tilde{P}} s_{t+1} - r_t$ where \tilde{P} is the optimal subjective belief. Solving forward the UIP condition implies that without time-variation or ambiguity the solution would simply be $s_t = \frac{1}{\rho-1} r_t$. With time-variation the forward iteration is significantly more complicated if one is taking into account that the agent updates her future beliefs about the parameter. However, resorting to a model of anticipated utility (as in Sargent (1999)) eliminates this problem because then agents in fact ignore such future updates in forecasting. Then the solution is

$$s_t = (\hat{\rho}_t - 1)^{-1} r_t \quad (47)$$

where $\hat{\rho}_t = E_t^{\tilde{P}}(\rho_t|I_t)$. This has a similar interpretation as with the temporary versus hidden components model analyzed in the paper so far. Suppose that there is a higher observed domestic interest rate. Agents would like to invest in the domestic currency but they are worried about a domestic currency depreciation. In equilibrium this depreciation occurs if next period agents demand less of the domestic bond. Under the anticipated utility assumption, agents do not forecast the updating of beliefs and are only worried that the future observed domestic interest rate use is low. Given that they observe a high interest rate today, they are in effect concerned that this realization is coming from a temporary shock and that the hidden persistence parameter is low. Then the agent would tend to misperceive news in the same as in the benchmark model: positive innovations are given low precision (high noise) while negative innovations are seen as reflecting high precision.

For that note that $E_t^{\tilde{P}} s_{t+1} = (\hat{\rho}_t - 1)^{-1} E_t^{\tilde{P}} r_{t+1}$ and $E_t^{\tilde{P}} r_{t+1} = \hat{\rho}_t r_t$. In equilibrium the agent takes as given s_t so the expected returns from the carry trade strategy are: $q_{t+1}^e = r_t - E_t^{\tilde{P}} s_{t+1} + s_t$. This can be rewritten as: $q_{t+1}^e = r_t - (\hat{\rho}_t - 1)^{-1} \hat{\rho}_t r_t + s_t$. The expected returns are decreasing in $\hat{\rho}_t$ which confirms the intuition above. For a deterministic sequence σ_V^t , the Kalman filter formulas for the time-varying parameters are:

$$\begin{aligned} \hat{\rho}_t &= \hat{\rho}_{t-1} + K_t(r_t - \hat{\rho}_{t-1}r_{t-1}) \\ K_t &= \frac{\Sigma_{t-1}r_{t-1}}{r_{t-1}^2\Sigma_{t-1} + \sigma_{V,t}^2} \\ \Sigma_t &= \Sigma_{t-1} - K_t\Sigma_{t-1}r_{t-1} + \sigma_U^2 \end{aligned}$$

The agent guards herself against the worse outcomes and chooses high values for $\sigma_{V,t}^2$ when $r_{t-1}(r_t - \hat{\rho}_{t-1}r_{t-1}) > 0$ and low values when the opposite happens. This is very much similar with the decision rule in (26). Similar to the benchmark model, the true DGP is a constant volatility sequence of σ_V . Because there is underestimation of the hidden parameter on average next period the observed interest rate differential is higher than expected. That will be perceived as positive news for the agents that invest in the domestic bonds because it raises the present value of the payoffs on the bond. This will create a gradual reallocation towards the domestic bonds and the possibility for an appreciation following an increase in the domestic rate.

References

- Alvarez, Fernando, Andrew Atkeson, and Patrick J. Kehoe (2008), “Time-varying risk, interest rates, and exchange rates in general equilibrium.” Staff Report 371, Federal Reserve Bank of Minneapolis.
- Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Clara Vega (2003), “Micro effects of macro announcements: Real-time price discovery in foreign exchange.” *American Economic Review*, 93, 38–62.
- Anderson, Evan W., Lars Peter Hansen, and Thomas 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, Philippe, Elmar Mertens, and Eric van Wincoop (2008), “Predictability in financial markets: What do survey expectations tell us?” *Journal of International Money and Finance*. Forthcoming.
- Bacchetta, Philippe and Eric van Wincoop (2008), “Infrequent portfolio decisions: A solution to the forward discount puzzle.” Mimeo.
- Backus, David K., Allan W. Gregory, and Chris I. Telmer (1993), “Accounting for forward rates in markets for foreign currency.” *Journal of Finance*, 48, 1887–1908.
- Bansal, Ravi and Ivan Shaliastovich (2007), “Risk and return in bond, currency and equity markets.” Mimeo.
- Barillas, Francisco, Lars Peter Hansen, and Thomas J. Sargent (2008), “Doubts or variability?” Mimeo.
- Barro, Robert J. (2006), “Rare disasters and asset markets in the Twentieth Century.” *Quarterly Journal of Economics*, 121, 823–866.
- Bekaert, Geert (1996), “The time variation of risk and return in foreign exchange markets: A general equilibrium perspective.” *Review of Financial Studies*, 9, 427–70.
- Benigno, Pierpaolo (2007), “Portfolio choices with near rational agents: A solution for some international-finance puzzles.” NBER Working Papers 13173.
- Bernanke, Ben S. and Jean Boivin (2003), “Monetary policy in a data-rich environment.” *Journal of Monetary Economics*, 50, 525–546.
- Brock, William A., Steven N. Durlauf, and Kenneth D. West (2007), “Model uncertainty and policy evaluation: Some theory and empirics.” *Journal of Econometrics*, 136, 629–664.
- Brunnermeier, Markus, Stefan Nagel, and Lasse Pedersen (2008), “Carry trades and currency crashes.” *NBER Macroeconomics Annual*. Forthcoming.

- Brunnermeier, Markus K. and Lasse Heje Pedersen (2008), “Market liquidity and funding liquidity.” *Review of Financial Studies*. Forthcoming.
- Burnside, A. Craig, Martin S. Eichenbaum, Isaac Kleshchelski, and Sergio Rebelo (2008), “Do peso problems explain the returns to the carry trade?” Mimeo.
- Burnside, Craig (2007), “The cross-section of foreign currency risk premia and consumption growth risk: A comment.” NBER Working Papers 13129.
- Burnside, Craig, Martin S. Eichenbaum, and Sergio Rebelo (2007), “Understanding the forward premium puzzle: A microstructure approach.” NBER Working Papers 13278.
- Canova, Fabio and Jane Mairrinan (1993), “Profits, risk, and uncertainty in foreign exchange markets.” *Journal of Monetary Economics*, 32, 259–286.
- Chakraborty, Avik and George W. Evans (2008), “Can perpetual learning explain the forward-premium puzzle?” *Journal of Monetary Economics*, 55, 477–490.
- Chinn, Menzie D. and Jeffrey 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* (W. Allen and D. Dickinson, eds.), 145–167, London, Routledge.
- Chinn, Menzie D. and Guy Meredith (2005), “Testing uncovered interest parity at short and long horizons during the Post-Bretton Woods era.” NBER Working Papers 11077.
- Clarida, Richard, Jordi Gali, and Mark Gertler (2000), “Monetary policy rules and macroeconomic stability: Evidence and some theory.” *The Quarterly Journal of Economics*, 115, 147–180.
- Cogley, Timothy and Thomas J. Sargent (2005), “The conquest of U.S. inflation: Learning and robustness to model uncertainty.” *Review of Economic Dynamics*, 8, 528–563.
- Dornbusch, Rudiger (1976), “Expectations and exchange rate dynamics.” *Journal of Political Economy*, 84, 1161–76.
- Eichenbaum, Martin and Charles 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, Charles (1996), “The forward discount anomaly and the risk premium: A survey of recent evidence.” *Journal of Empirical Finance*, 3, 123–192.
- Engel, Charles and James D. Hamilton (1989), “Long swings in the exchange rate: Are they in the data and do markets know it?” *American Economic Review*, 80, 689–713.

- Engel, Charles, Nelson C. Mark, and Kenneth D. West (2007), “Exchange rate models are not as bad as you think.” NBER Working Papers 13318.
- Engel, Charles and Kenneth 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, Larry G. and Martin Schneider (2007), “Learning under ambiguity.” *Review of Economic Studies*, 74, 1275–1303.
- Erceg, Christopher J. and Andrew T. Levin (2003), “Imperfect credibility and inflation persistence.” *Journal of Monetary Economics*, 50, 915–944.
- Evans, George W. and Seppo Honkapohja (2001), *Learning and Expectations in Macroeconomics*. Princeton University Press, Princeton, NJ.
- Evans, Martin D. D. and Karen K. Lewis (1995), “Do long-term swings in the dollar affect estimates of the risk premia?” *Review of Financial Studies*, 8, 709–42.
- Fama, Eugene F. (1984), “Forward and spot exchange rates.” *Journal of Monetary Economics*, 14, 319–338.
- Farhi, Emmanuel and Xavier Gabaix (2008), “Rare disasters and exchange rates.” NBER Working Papers 13805.
- Faust, Jon and John H. Rogers (2003), “Monetary policy’s role in exchange rate behavior.” *Journal of Monetary Economics*, 50, 1403–1424.
- Flood, Robert P and Andrew K Rose (1996), “Fixes: Of the forward discount puzzle.” *The Review of Economics and Statistics*, 78, 748–52.
- Frankel, Jeffrey A and Kenneth A Froot (1987), “Using survey data to test standard propositions regarding exchange rate expectations.” *American Economic Review*, 77, 133–53.
- Froot, Kenneth A and Jeffrey A Frankel (1989), “Forward discount bias: Is it an exchange risk premium?” *The Quarterly Journal of Economics*, 104, 139–61.
- Froot, Kenneth A and Richard H Thaler (1990), “Anomalies: Foreign exchange.” *Journal of Economic Perspectives*, 4, 179–92.
- Gabaix, Xavier (2008), “Variable rare disasters: An exactly solved framework for ten puzzles in macro-finance.” NBER Working Papers 13724.
- Giannoni, Marc P. (2007), “Robust optimal monetary policy in a forward-looking model with parameter and shock uncertainty.” *Journal of Applied Econometrics*, 22, 179–213.

- Gilboa, Itzhak and David Schmeidler (1989), “Maxmin expected utility with non-unique prior.” *Journal of Mathematical Economics*, 18, 141–153.
- Gourinchas, Pierre-Olivier and Aaron Tornell (2004), “Exchange rate puzzles and distorted beliefs.” *Journal of International Economics*, 64, 303–333.
- Gourio, Francois (2008), “Disasters and recoveries.” *American Economic Review*, 98, 68–73.
- Grilli, Vittorio and Nouriel Roubini (1996), “Liquidity models in open economies: Theory and empirical evidence.” *European Economic Review*, 40, 847–859.
- Hansen, Lars Peter and Robert J Hodrick (1980), “Forward exchange rates as optimal predictors of future spot rates: An econometric analysis.” *Journal of Political Economy*, 88, 829–53.
- Hansen, Lars Peter and Thomas J. Sargent (2007a), “Fragile beliefs and the price of mofel uncertainty.” Mimeo.
- Hansen, Lars Peter and Thomas J. Sargent (2007b), “Recursive robust estimation and control without commitment.” *Journal of Economic Theory*, 136, 1–27.
- Hansen, Lars Peter and Thomas J. Sargent (2008), *Robustness*. Princeton University Press.
- Jurek, Jakub W. (2008), “Crash-neutral currency carry trades.” Mimeo.
- Kaminsky, Graciela (1993), “Is there a peso problem? Evidence from the Dollar/Pound exchange rate, 1976-1987.” *American Economic Review*, 83, 450–72.
- Kleshchelski, Isaac and Nicolas Vincent (2007), “Robust equilibrium yield curves.” Mimeo.
- Klibanoff, Peter, Massimo Marinacci, and Sujoy Mukerji (2005), “A smooth model of decision making under ambiguity.” *Econometrica*, 73, 1849–1892.
- Levin, Andrew T. and John C. Williams (2003), “Robust monetary policy with competing reference models.” *Journal of Monetary Economics*, 50, 945–975.
- Lewis, Karen K (1989), “Changing beliefs and systematic rational forecast errors with evidence from foreign exchange.” *American Economic Review*, 79, 621–36.
- Lewis, Karen K. (1995), “Puzzles in international finance markets.” In *Handbook of International Economics* (Gene G. Grossman and Kenneth Rogoff, eds.), 1913–1971, Amsterdam, Elsevier Science.
- Li, Ming and Aaron Tornell (2007), “Exchange rates under robustness: The forward premium puzzle and momentum.” Mimeo.
- Lustig, Hanno and Adrien Verdelhan (2007), “The cross section of foreign currency risk premia and consumption growth risk.” *American Economic Review*, 97, 89–117.

- Lustig, Hanno and Adrien Verdelhan (2008), “Note on the cross-section of foreign currency risk premia and consumption growth risk.” NBER Working Papers 13812.
- Lyons, Richard (2001), *The Microstructure Approach to Exchange Rates*. MIT Press:Cambridge, MA.
- Maccheroni, Fabio, Massimo Marinacci, and Aldo Rustichini (2006), “Ambiguity aversion, robustness, and the variational representation of preferences.” *Econometrica*, 74, 1447–1498.
- Mark, Nelson C and Yangru Wu (1998), “Rethinking deviations from uncovered interest parity: The role of covariance risk and noise.” *Economic Journal*, 108, 1686–1706.
- Plantin, Guillaume and Shin S. Shin (2008), “Carry trades and speculative dynamics.” Mimeo.
- Sargent, Thomas (1999), *The Conquest of American Inflation*. Princeton University Press, NJ.
- Sarno, Lucio (2005), “Towards a solution to the puzzles in exchange rate economics: Where do we stand?” *Canadian Journal of Economics*, 38, 673–708.
- Scholl, Almuth and Harald Uhlig (2006), “New evidence on the puzzles: Monetary policy and exchange rates.” Working Paper 5, Society for Computational Economics.
- Sims, Christopher A. and Tao Zha (2006), “Were there regime switches in U.S. monetary policy?” *American Economic Review*, 96, 54–81.
- Strzalecki, Tomasz (2007), “Axiomatic foundations of multiplier preferences.” Mimeo.
- Tornell, Aaron (2003), “Exchange rate anomalies under model misspecification: A mixed optimal/robust approach.” UCLA Economics Online Papers 266, UCLA Department of Economics.
- Van Nieuwerburgh, Stijn and Laura Veldkamp (2006), “Learning asymmetries in real business cycles.” *Journal of Monetary Economics*, 53, 753–772.
- Verdelhan, Adrien (2006), “A habit-based explanation of the exchange rate risk premium.” Mimeo.
- Veronesi, Pietro (1999), “Stock market overreaction to bad news in good times: A rational expectations equilibrium model.” *Review of Financial Studies*, 12, 975–1007.
- Woodford, Michael (2006), “An example of robustly optimal monetary policy with near-rational expectations.” *Journal of the European Economic Association*, 4, 386–395.

Table 4. Empirical UIP regression and carry trade returns

Country	UIP regression			Carry trade returns				
	α	β	R^2	Mean	Standard Deviation	Sharpe Ratio	Skewness	Kurtosis
Belgium [†]	0.000 (0.002)	-0.593 (0.612)	0.002	0.0067 (0.0022)	0.033 (0.002)	0.203 (0.061)	0.003 (0.243)	3.94 (0.46)
Canada	0.004 (0.002)	-0.631 (0.49)	0.004	0.0019 (0.0009)	0.016 (0.001)	0.125 (0.056)	-0.500 (0.169)	4.38 (0.50)
France [†]	0.000 (0.003)	0.092 (0.705)	0.000	0.0051 (0.0019)	0.032 (0.001)	0.160 (0.062)	-0.035 (0.155)	3.45 (0.31)
Germany [†]	0.003 (0.002)	-0.657 (0.832)	0.003	0.0012 (0.0022)	0.033 (0.002)	0.036 (0.065)	-0.184 (0.128)	3.54 (0.32)
Italy [†]	-0.000 (0.001)	0.196 (0.388)	0.001	0.0023 (0.0021)	0.031 (0.002)	0.074 (0.068)	-0.297 (0.223)	4.08 (0.50)
Japan [‡]	0.010 (0.003)	-2.405 (0.667)	0.026	0.0025 (0.0020)	0.034 (0.002)	0.073 (0.058)	-0.669 (0.246)	4.64 (0.87)
Switzerland	0.007 (0.003)	-1.407 (0.692)	0.014	0.0010 (0.0020)	0.035 (0.003)	0.029 (0.057)	-0.229 (0.210)	3.74 (0.46)
UK	-0.002 (0.002)	-1.533 (0.86)	0.014	0.0053 (0.0015)	0.030 (0.003)	0.177 (0.052)	-0.029 (0.376)	4.91 (0.91)
Average	0.003	-0.85	0.008	0.0033	0.0306	0.11	-0.242	4.08

Notes: The first 3 columns report estimates of the regression of time $t + 1$ exchange rate difference on the time t forward premium: $S_{t+1}/S_t - 1 = \alpha + \beta(F_t/S_t - 1) + \varepsilon_{t+1}$. Both F_t and S_t are USD per FCU. Heteroskedasticity-robust standard errors are in parentheses. This table reports summary statistics for returns from implementing the standard carry trade involving the USD and the pairing foreign currency. Returns are measured in USD, per dollar bet. Monthly data is used for the sample M1:1976 to M12:2006, except for Euro legacy countries ([†]) for which the data ends in M12:1998 and Japan for which data begins on M7:1978.

Table 5. ML estimates of a state space representation with constant volatilities

Country pair	Interest rate differential			Forward Discount		
	$\sum \rho$	σ_V	σ_U	$\sum \rho$	σ_V	σ_U
Belgium [†]	0.965 (0.01)	0.19 (0.06)	0.56 (0.14)	0.89 (0.04)	0.77 (0.15)	1.23 (0.6)
Canada	0.955 (0.01)	0.28 (0.05)	0.53 (0.05)	0.99 (0.06)	0.33 (0.032)	0.41 (0.06)
France [†]	0.972 (0.03)	0.003 (0.18)	0.43 (0.11)	0.945 (0.02)	1.31 (0.2)	1.48 (0.340)
Germany [†]	0.979 (0.02)	0.001 (0.005)	0.45 (0.025)	0.987 (0.02)	0.12 (0.32)	1.2 (0.41)
Italy [†]	0.969 (0.01)	0.08 (0.05)	0.23 (0.09)	0.972 (0.03)	0.21 (0.11)	0.35 (0.11)
Japan	0.986 (0.06)	0.031 (0.088)	0.45 (0.058)	0.983 (0.03)	0.21 (0.061)	0.65 (0.095)
Switzerland	0.978 (0.08)	0.0056 (0.065)	0.52 (0.022)	0.962 (0.03)	0.28 (0.08)	0.63 (0.17)
UK	0.974 (0.02)	0.29 (0.02)	0.31 (0.04)	0.969 (0.05)	0.38 (0.03)	0.35 (0.06)

Notes: The state-space representation is described in (32). The sample for the interest rate differentials data is M1 1981-M12 2006; for the forward discount is M1 1976-M12 2006, except for Japan M7 1978-M12 2006. 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. The long run autocorrelation for the hidden state is denoted by $\sum \rho$ and is defined as the sum of the AR coefficients in the transition equation.

Table 6. Model implied $\hat{\beta}$ for the UIP regression

		(1)		(2)		(3)		(4)		(5)	
		$\hat{\beta}$	$t_{\hat{\beta}}$	$\hat{\beta}$	$t_{\hat{\beta}}$	$\hat{\beta}$	$t_{\hat{\beta}}$	$\hat{\beta}$	$t_{\hat{\beta}}$	$\hat{\beta}$	$t_{\hat{\beta}}$
T=300	Mean	-0.25	-0.81	0.37	1.27	0.34	1.1	-0.57	-1.1	-0.19	-0.68
	Median	-0.22	-0.76	0.32	1.24	0.28	0.95	-0.55	-0.93	-0.17	-0.65
	St.dev.	0.14	0.3	0.22	0.51	0.21	0.36	0.24	0.47	0.11	0.28
T=3000	Mean	-0.28	-2.14	0.54	1.75	0.48	1.58	-0.66	-2.58	-0.21	-1.71
	Median	-0.28	-2.14	0.52	1.72	0.47	1.55	-0.67	-2.59	-0.21	-1.71
	St.dev.	0.07	0.27	0.11	0.32	0.09	0.2	0.11	0.27	0.06	0.25

Note: Table 6 reports statistics for the model implied UIP regression defined in (33) $s_{t+1} - s_t = \beta(i_t - i_t^*) + \varepsilon_{t+1}$. The results are extremely similar when a constant is included in the regression. The estimated coefficient is $\hat{\beta}$ and its t-statistic in the sample is denoted by $t_{\hat{\beta}}$. The reported values are statistics across $N = 1000$ simulations for $T = 300$ and $N = 500$ for $T=3000$. Column (1) presents the results for the benchmark model. Column (2) for the case when the true noise/signal ratio is $\sigma_V^2/\sigma_U^2 = 10$. Column (3) is the case in which the long-run autocorrelation in the state equation is $\sum \rho = 0.7$. Column (4) is the case in which there is no restriction on the number of periods in which the agent can distort the sequence of variances, i.e. $n = t$. Column (5) maintains the benchmark restriction of $n = 2$ but allows $m = 10$ so that the agent distorts 2 out of the last 10 periods. See Section 5.1 for a discussion.

Table 7. A. Model implied statistics for the unhedged carry trade returns

	Mean	Standard Deviation	Sharpe Ratio	Skewness	Kurtosis
$T = 300$	0.0017 (0.0005)	0.0106 (0.007)	0.16 (0.047)	-0.421 (0.288)	5.37 (0.91)
$T = 3000$	0.0018 (0.0001)	0.0115 (0.00045)	0.1565 (0.01)	-0.19 (0.15)	6.58 (0.68)

Table 7.B. Model implied statistics for the hedged carry trade returns

	Mean	Standard Deviation	Sharpe Ratio	Skewness	Kurtosis
$T = 300$	0.0012 (0.0004)	0.0067 (0.0006)	0.179 (0.038)	1.46 (0.38)	7.22 (1.81)
$T = 3000$	0.0012 (0.0001)	0.0073 (0.00037)	0.164 (0.0135)	2.63 (0.17)	12.01 (1.66)

Table 7.C. Model implied statistics for the “crash risk” regressions

$T = 300$				$T = 3000$			
$\hat{\beta}_2$	$t_{\hat{\beta}_2}$	$\hat{\beta}_3$	$t_{\hat{\beta}_3}$	$\hat{\beta}_2$	$t_{\hat{\beta}_2}$	$\hat{\beta}_3$	$t_{\hat{\beta}_3}$
-3.1	-8.61	-0.72	-10.21	-2.96	-7.83	-0.83	-12.41
(1.21)	(2.44)	(0.29)	(2.1)	(0.89)	(1.48)	(0.22)	(1.9)

Note. Results obtained for the benchmark version. The unhedged carry trade is defined in (30). The hedged carry trade is defined in (31). The excess returns at time $t+1$ are defined as $ex_{t+1} = r_t - (s_{t+1} - s_t)$. The first regression in Table 7.C is $Skew_{t+1} = \beta_2 r_t + \xi_{2,t+1}$ where $Skew_{t,t+1}$ is the time t model implied realized skewness of time $t+1$ excess returns. The second regression is $Skew_{t,t+1}^{\tilde{P}} = \beta_3 r_t + \xi_{3,t+1}$, where $Skew_{t,t+1}^{\tilde{P}}$ is the time t model implied risk-neutral skewness of time $t+1$ returns. The reported values are statistics across $N = 1000$ simulations for $T = 300$ and $N = 500$ for $T=3000$. Standard errors of the statistics across the simulations are in parentheses.

Figure 1: Model generated exchange rate path

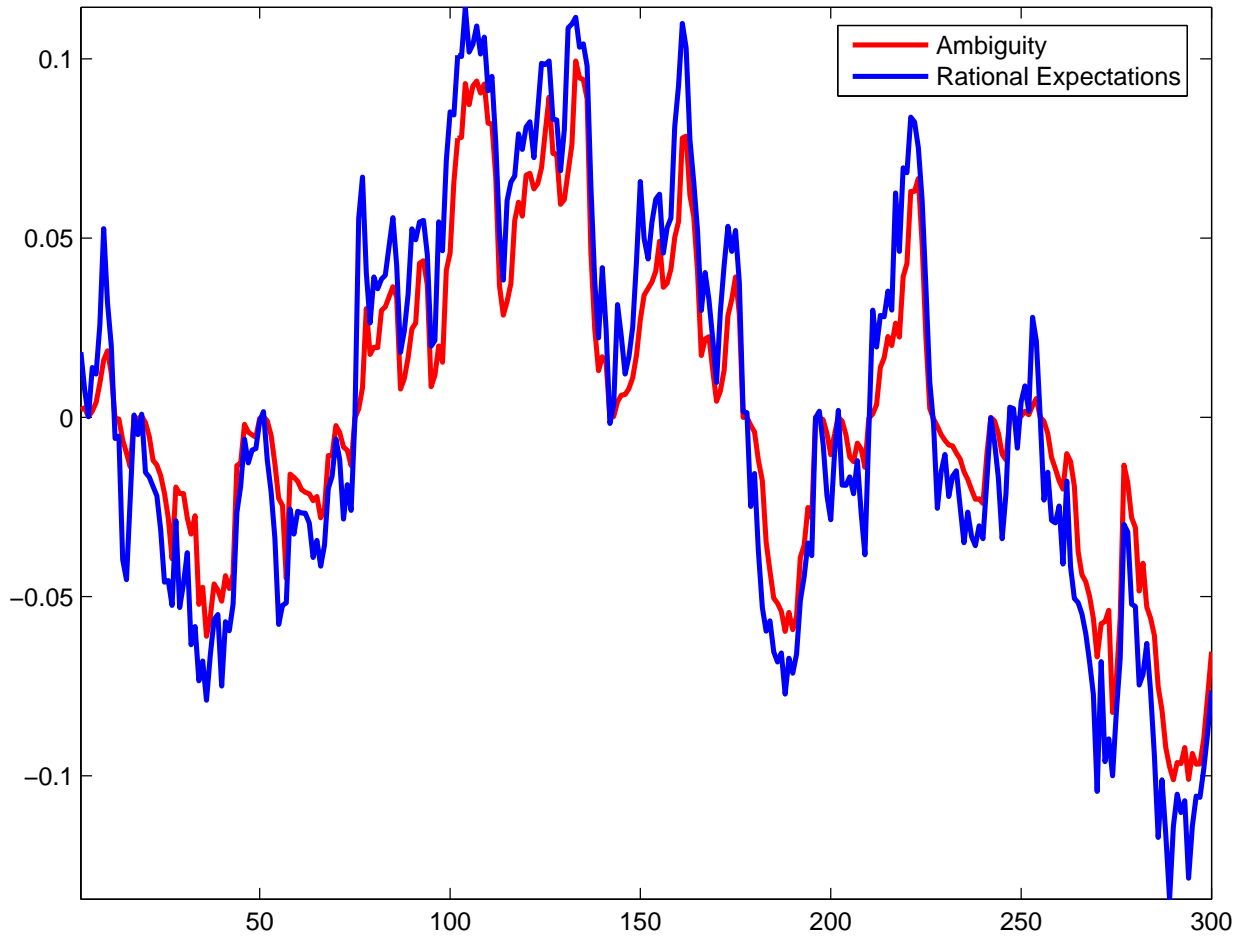


Table 8 Correlations exchange rates

(1)	(2)	(3)	(4)
0.96	0.26	0.164	0.356
(0.02)	(0.04)	(0.04)	(0.05)

Note. Figure 1 plots the exchange rate path for a simulation of $T=300$. Table 8 computes the correlation between the exchange rate under the distorted expectations (s_{t+1}) and under rational expectations (s_{t+1}^{RE}) for a sample of $T = 3000$. The first and second column are the unconditional correlation of the exchange rate levels ($Corr(s_{t+1}, s_{t+1}^{RE})$) and first difference ($Corr(\Delta s_{t+1}, \Delta s_{t+1}^{RE})$) respectively. The third and fourth column are the conditional correlations between the exchange rate differences in states in which the product $(\Delta s_{t+1}^{RE} s_{t+1}^{RE}) > 0$ and $(\Delta s_{t+1}^{RE} s_{t+1}^{RE}) < 0$ respectively. Standard errors are computed across $N=500$ simulations of samples of size T . Standard errors of the statistics across the simulations are in parentheses. See Section 5.1

Figure 2: Model generated estimates of hidden state

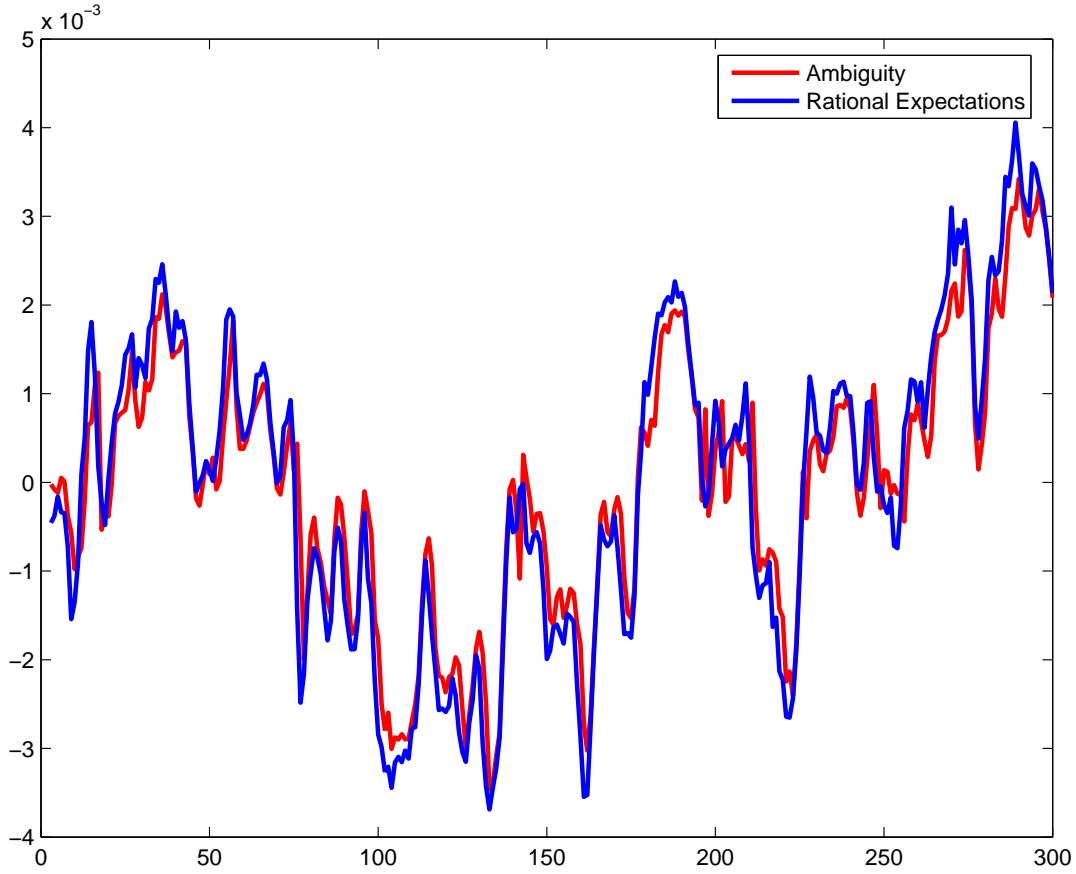
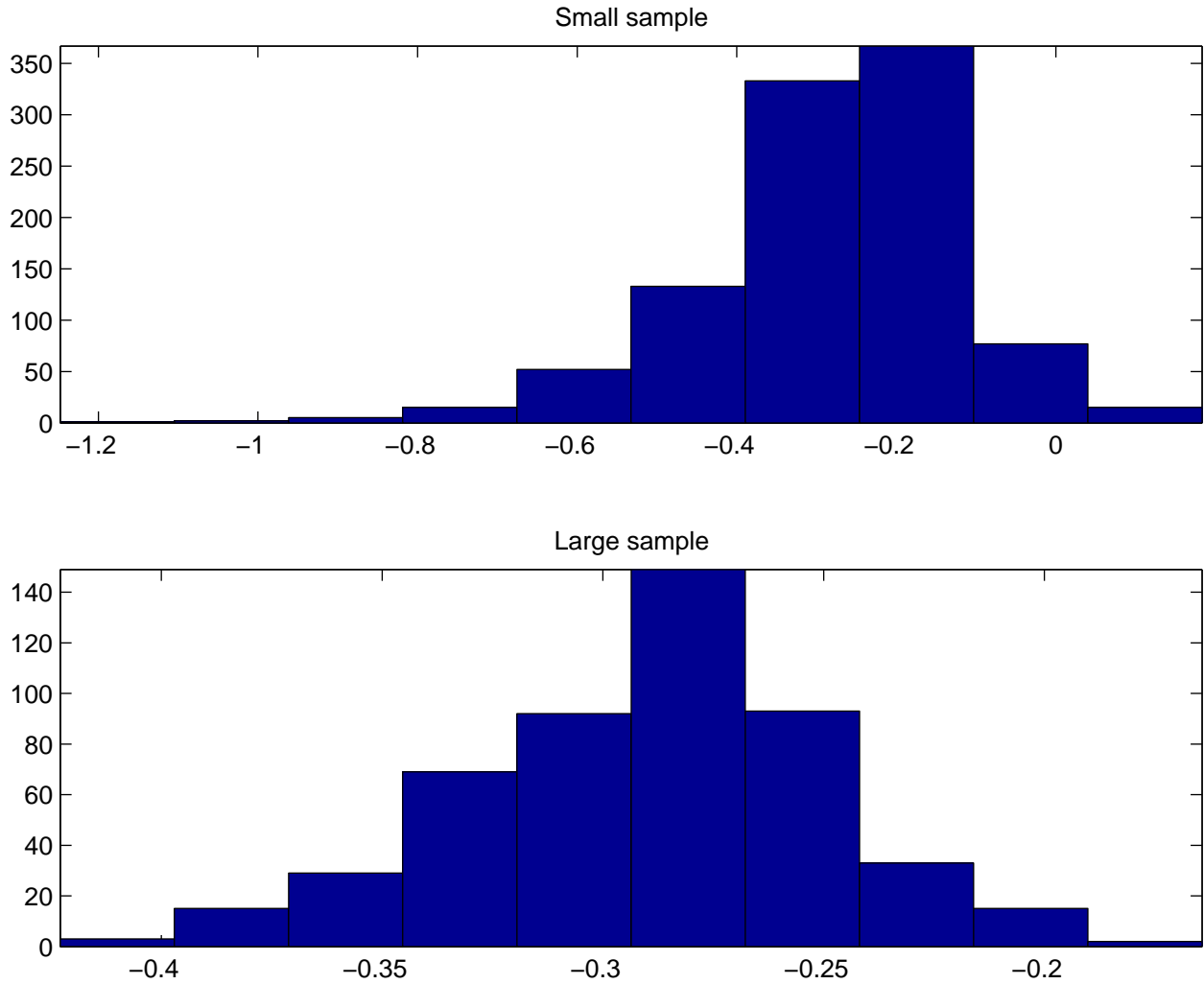


Table 9 Correlations estimate of hidden state

(1)	(2)	(3)	(4)
0.97	0.49	0.35	0.69
(0.02)	(0.05)	(0.04)	(0.06)

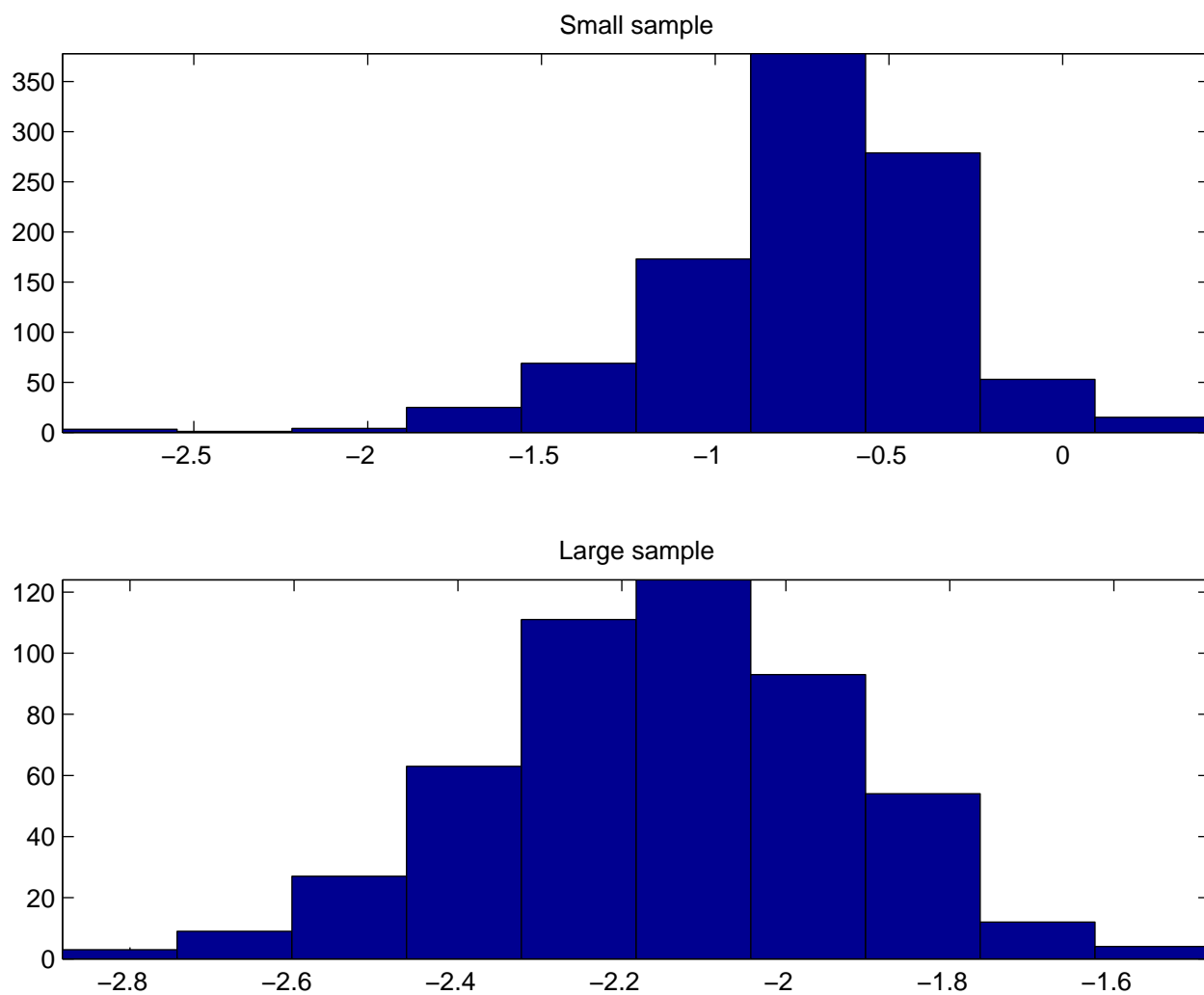
Note. Figure 2 plots the estimate of the hidden state for the same simulation of $T=300$ that generated Figure 1. Table 9 computes the correlation between the time t estimate of the time t hidden state under the distorted expectations ($\hat{x}_{t,t}^t$) and under rational expectations ($\hat{x}_{t,t}^{RE}$) for a sample of $T = 3000$. The first and second column are the unconditional correlation of the estimate levels ($Corr(\hat{x}_{t,t}^t, \hat{x}_{t,t}^{RE})$) and first difference ($Corr(\Delta\hat{x}_{t,t}^t, \Delta\hat{x}_{t,t}^{RE})$) respectively. The third and fourth column are the conditional correlations between the first difference in the estimates in states in which the product $(\Delta\hat{x}_{t,t}^{RE} s_t^{RE}) < 0$ and $(\Delta\hat{x}_{t,t}^{RE} s_t^{RE}) > 0$ respectively. Standard errors are computed across $N=500$ simulations of samples of size T . Standard errors of the statistics across the simulations are in parentheses.

Figure 3: Model implied UIP coefficients



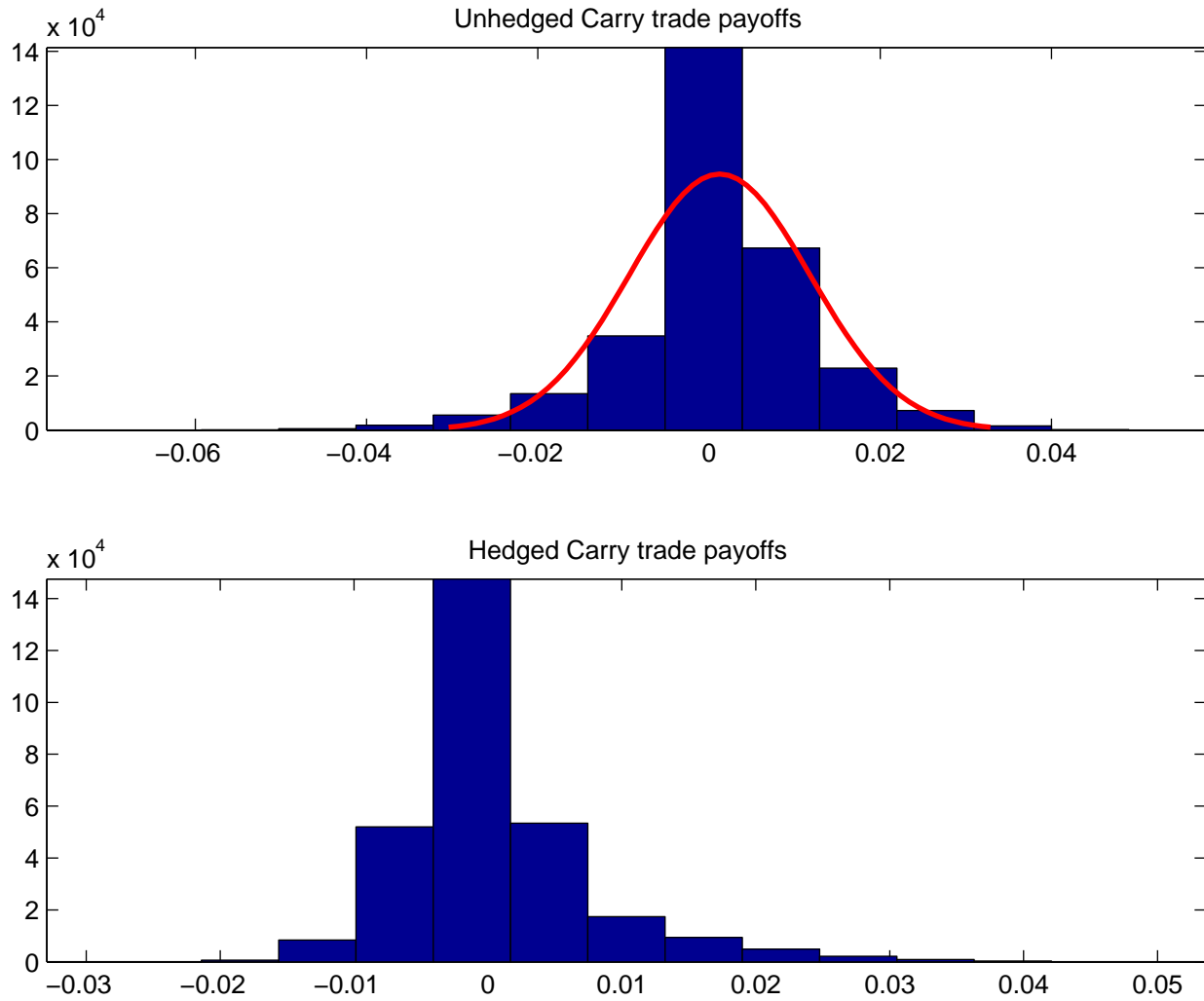
Note: The top plot is the histogram for the model implied distribution of estimated UIP coefficients for a small sample of $T=300$ based on $N=1000$ simulations. The bottom plot is the histogram for a large sample of $T=3000$ based on $N=500$ simulations.

Figure 4: Model implied t-stat significance of UIP coefficients



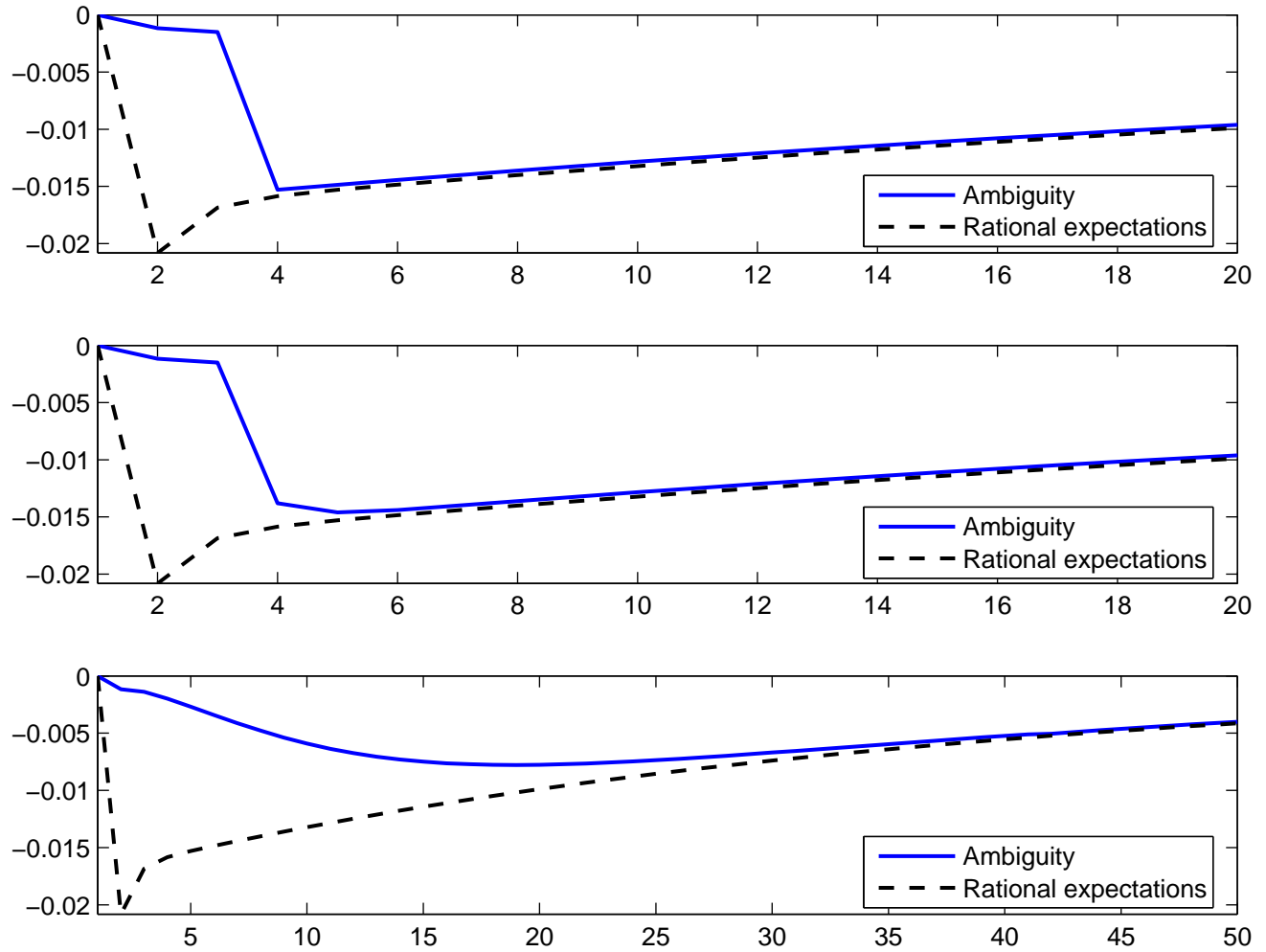
Note: These plots are histograms for the model implied distribution of the t-stat significance of the estimated UIP coefficients in Figure 3.

Figure 5: Model generated carry trade payoffs



Note: The top plot is the histogram for the model implied distribution of the unhedged carry trade payoffs across a sample size of $T=300$ and number of simulations of $N=1000$. The red line indicates the histogram implied by a normal distribution with the same mean and standard deviation as in the model implied distribution. The bottom plot is the histogram of the hedged carry trade payoffs for the same simulations as the top plot. The payoffs are computed at the monthly frequency.

Figure 6: Delayed overshooting



Note: These plots show the response of the exchange rate to an unanticipated decrease in the foreign interest rate for the ambiguity aversion and rational expectations model. The same size of distortions on the temporary shock's variance is used for all three plots. The top plot employs the benchmark restriction of allowing only the last two periods to be distorted. In the middle plot the agent distorts any two of the last ten dates. In the bottom plot there are no restrictions on the number of periods in which this distortion is optimally applied.