

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

COSMIN L. ILUT[†]

First draft: October 2008

This draft: March 2009

Abstract

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

Key Words: uncovered interest rate parity, ambiguity aversion, robust filtering.

JEL Classification: D8, E4, F3, G1.

*I am grateful to the members of my dissertation committee Lawrence Christiano, Martin Eichenbaum, Giorgio Primiceri and Sergio Rebelo for their continuous support and advice. I would also like to thank Gadi Barlevy, Peter Benczur, Eddie Dekel, Lars Hansen, Nicolas Lehmann-Ziebarth, Ricardo Massolo, Jonathan Parker, Tom Sargent, Tomasz Strzalecki and seminar participants at the Board of Governors, Chicago Fed, Duke, ECB, New York Fed, NYU, Northwestern, Philadelphia Fed, UC Davis, UC Santa Cruz and Univ. of Virginia for helpful discussions and comments.

[†]Department of Economics, Duke University. E-mail: cosmin.ilut@duke.edu

1 Introduction

According to uncovered interest rate parity (UIP), periods when the domestic interest rate is higher than the foreign interest rate should on average be followed by periods of domestic currency depreciation. An implication of UIP is that a regression of realized exchange rate changes on interest rate differentials should produce a coefficient of 1. This implication is strongly counterfactual. In practice, UIP regressions (Hansen and Hodrick (1980), Fama (1984)) 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. This approach 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 complete knowledge of the true data generating process (DGP) and that they confront this uncertainty with ambiguity aversion. I model ambiguity aversion along the lines of the maxmin expected utility (or multiple priors) preferences as in Gilboa and Schmeidler (1989).

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

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

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

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

⁴See Lewis (1995) and Engel (1996) for surveys on this research. See Burnside et al. (2008) for a critical review of recent risk-based explanations. These criticisms are by no means definitive as there is a recent risk-based theoretical literature, including for example Verdelhan (2006), Bansal and Shaliastovich (2007), Alvarez et al. (2008) and Farhi and Gabaix (2008) that argues that the typical empirical exercises are unable by construction to capture the underlying time-variation in risk.

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

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

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

⁶The agents in my model resemble those in Bacchetta and van Wincoop (2008), except that there they investigate rational inattention and I assume ambiguity aversion.

while decreases as reflecting more persistent shocks (high precision of signals). The UIP condition holds ex-ante under these endogenously pessimistic beliefs.⁷

Because the agent underestimates, compared to the true DGP, the persistent component of the investment differential she is on average surprised next period by observing a higher investment differential than expected. Under her subjective beliefs these innovations are unexpected good news that increase the estimate of the hidden state. This updating effect creates the possibility that next period the agent finds it optimal to invest even more in the investment currency because this higher estimate raises the present value of the future payoffs of investing in the higher interest rate bond. The increased demand will drive up the value of the investment currency contributing to a possible appreciation of the investment currency. Thus, an investment currency could see a subsequent equilibrium appreciation instead of a depreciation as UIP predicts.

The main result of this paper is that such a model of exchange rate determination has the potential to resolve the UIP puzzle. Indeed, for the benchmark calibration, numerical simulations show that in large samples the UIP regression coefficient is negative and statistically significant while in small samples it is mostly negative and statistically not different from zero. The model is 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 parameterizations, 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.

The gradual incorporation of good news implied by this model can directly account also for the delayed overshooting puzzle. This is an empirically documented impulse response⁸ in which following a positive shock to the domestic interest rate the domestic currency experiences a gradual appreciation for several periods instead of an immediate appreciation and then a path of depreciation as UIP implies. For such an experiment, the ambiguity-

⁷In fact, the equilibrium condition is a risk-adjusted version of UIP which incorporates a risk-premium. However, as detailed in the model, this risk-correction is extremely small.

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

averse agent invests in equilibrium in the domestic currency and thus is worried about its future depreciation. The equilibrium beliefs then imply that the agent tends to overweigh, compared to the true DGP, the possibility that the observed increase in the interest rate reflects the temporary shock. This underestimation generates the gradual incorporation of the initial shock into the estimate and the demand of the ambiguity-averse agent.

The intuition for the model's ability to explain the UIP puzzle is related to Gourinchas and Tornell (2004) who show that if, for some unspecified reason, the agent systematically underreacts to signals about the time-varying hidden-state of the interest rate differential this can address the UIP and the delayed overshooting puzzle. The main difference is that here I investigate a model which addresses the origin and optimality of such beliefs. This model generates endogenous underreaction only to good news, with the agent in fact overreacting to bad news.

Related to the work presented here, Li and Tornell (2008) show that if the agent only cares about the mean square error of the estimate of the hidden state and she is concerned only about uncertainty in the observation equation then the robust Kalman gain is lower than in the reference model, thus implying underreaction to news. As an alternative model to generate an endogenous slow response to news, Bacchetta and van Wincoop (2008) use ideas from the rational inattention literature. In their setup, since information is costly to acquire and to process, some investors optimally choose to be inattentive and revise their portfolios infrequently. Their model implies that agents respond symmetrically to news.

The explanation for the UIP puzzle proposed in this paper relies on placing structure on the type of uncertainty that the agent is concerned about. The agent receives signals of uncertain precision about a time-varying hidden state but otherwise she trusts the other elements of her representation of the DGP. Because of the structured uncertainty, the equilibrium distorted belief is not equivalent to the belief generated by simply increasing the risk aversion and using the rational expectations assumption.⁹

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

⁹This 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.

Second, in the model hedged positions can deliver positive mean payoffs. Empirically, Burnside et al. (2008) and Jurek (2008) find significant evidence of this type of profitability. 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 the model this negative payoff does not completely offset the positive payoff of the unhedged carry trade. The reason is related to the significantly more frequent occurrence of good states for the carry trade strategy under the objective probability distribution than under the equilibrium distorted beliefs. This is in contrast to models in which peso events are associated with large losses for the unhedged carry trade strategy that do not occur in the sample but otherwise, for the non-peso events, the subjective and the probability distributions coincide.¹⁰ The theory presented in this paper is also consistent with recent empirical findings documented in Jurek (2008) about the conditional time-variation of risk-neutral moments for currency trading.

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. In my model, an increase in the high-interest-rate compared to the market's expectation produces, relatively to rational expectations, a slower appreciation of the investment currency since agents underreact to this type of innovations. However, a decrease in this rate generates a relatively sudden depreciation because agents respond quickly to that type of news.¹¹ 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 presents the model implications for exchange rate determination and discusses alternative specifications. Section 6 concludes. In the Appendix I provide details on some of the model's equations and statements.

¹⁰For such models see for example Engel and Hamilton (1989), Lewis (1989), Kaminsky (1993) and Evans and Lewis (1995). See Burnside et al. (2008) for a discussion on the possibility that peso events explain the UIP puzzle given the profitability of the hedged carry trade.

¹¹Brunnermeier et al. (2008) argue that the data suggests that the realized skewness is related to the rapid unwinding of currency positions, a feature that is replicated by my model. They propose shocks to funding liquidity as a mechanism for this endogeneity. See Burnside (2008) for a comment on this possibility.

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 is the state-space model:

$$\begin{aligned} r_t &= H'x_t + \sigma_V v_t \\ x_t &= Fx_{t-1} + \sigma_U u_t \end{aligned} \tag{2.1}$$

The shocks u_t and v_t are Gaussian white noises. Thus, at time t the observable differential r_t 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 . Their maxmin expected utility at time t is:

$$V_t = \max_{b_t} \min_{\tilde{P} \in \Lambda} E_t^{\tilde{P}}[-\exp(-\gamma W_{t+1})|I_t] \tag{2.2}$$

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 $t + 1$. At time $t + 1$ the agent has to repay the interest rate bearing borrowed amount of $b_t \exp(i_t)$. Thus, the agent has to exchange back the time $t + 1$ proceeds from FCU into USD and obtains $b_t \frac{S_{t+1}}{S_t} \exp(i_t^*)$. The net end-of-life 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 0. Thus, the market clearing condition is:

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

where B is the steady state amount of Foreign bonds. Following Bacchetta and van Wincoop (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 (2.3) around steady state I get the market clearing condition¹²:

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

¹²Bacchetta 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.

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 (2.1). As in the model of multiple priors (or MaxMin Expected Utility) of Gilboa and Schmeidler (1989), the agent chooses beliefs about the stochastic process that induce the lowest expected utility under that subjective probability distribution. The minimization is constrained by a particular set of possible distortions because otherwise the agent would select infinitely pessimistic probability distributions.¹³ 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. The set Λ dictates how I constrain the problem of choosing an optimally distorted model. The type of uncertainty that I investigate is similar to Epstein and Schneider (2007, 2008), except that here I consider time-varying hidden states, while their model analyzes a constant hidden parameter. The agent believes that the standard deviation of the temporary shock is potentially time-varying and is drawn every period from a set Υ . Typical of ambiguity aversion frameworks, the agent's uncertainty manifests in her cautious approach of not placing probabilities on this set. Every period she thinks that any draw can be made out of this set. The agent trusts the remaining elements of the representation in (2.1).

Thus the agent uses the following state-space representation:

$$\begin{aligned} r_t &= H'x_t + \sigma_{V,t}v_t \\ x_t &= Fx_{t-1} + \sigma_Uu_t \end{aligned} \tag{2.5}$$

where v_t and u_t are Gaussian 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 $\sigma_{V,s}$

¹³The maxmin expected utility 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 for example Klibanoff et al. (2005), in which the agent does not choose the minimum of the set but rather weighs more the worse distributions.

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

This minimization then becomes selecting a sequence of

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

in the product space $\Upsilon^t : \Upsilon \times \Upsilon \dots \Upsilon$. As in Epstein and Schneider (2007), the agent interprets this sequence as a “theory” of how the data was generated.¹⁴

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 .¹⁵ 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.¹⁶

The type of structured uncertainty I consider implies that the minimization in (2.2) is reduced to selecting a distorted sequence of the form (2.6). The optimization in (2.2) then becomes:

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

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 .

Whether I assume uncertainty about the realizations for the variances of the temporary

¹⁴Note that the distorted model is not a constant volatility model with a different value for the standard deviation of the shocks than the reference model. Although this possibility is implicitly nested in the setup, the optimal choice will likely be different because sequences with time variation will induce a lower utility for the agent.

¹⁵This does not necessarily imply that σ_V is a priori known. If the agent uses maximum likelihood for a constant volatility model, her point estimate would be asymptotically σ_V .

¹⁶A 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.

shock or the persistent shock is intuitively innocuous. The driving force in the agent's evaluation of the expected 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 equation in which I assume uncertainty, the observation or state equation, that matters but the relative strength in the information contained in them.

The fact that the expected returns are influenced by what the agent perceives as the realized variances of the unobserved shocks is important. Typical models that deal with robustness against misspecification in the context of an estimation have mostly considered a setup with commitment to previous distortions in which the agent wants to minimize the estimation mean square error. In that case, the robust estimator features different qualitative properties. Li and Tornell (2008) study such a problem and show that if the agent is concerned only about the uncertainty of the temporary shock then she will act as if the variance of temporary shock is higher. That generates a steady state robust Kalman gain that is lower than the one implied by the reference model. Such a concern for misspecification generates the type of underreaction to news that has been proposed in the literature as a mechanism to explain the UIP puzzle. However, as Li and Tornell (2008) point out, when uncertainty about the persistent shock is added the concern for misspecification leads to a higher robust Kalman gain than in the reference model.¹⁷ That implies an overreaction to news and a UIP regression coefficient that is higher than 1, thus moving away from explaining the puzzle.

As Hansen and Sargent (2007) 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 their setup, without further structure on the type of uncertainty that the agent is concerned about, in my model simply invoking robustness against the hidden state would produce an equilibrium that is equivalent to the one under rational expectations and an increased risk aversion.¹⁸ In Appendix B 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

¹⁷In this case, as discussed in Basar and Bernhard (1995) and Hansen and Sargent (2008, Ch.17), the robust filter flattens the decomposition of variances across frequencies by accepting higher variances at higher frequencies in exchange for lower variances for lower frequencies. Intuitively, the agent is more concerned about the model being misspecified at low-frequencies.

¹⁸As discussed in Hansen and Sargent (2007) this equivalence is not a general result.

is not suited in this model for addressing the empirical findings.

2.3 Statistical constraint on possible distortions

An important question that arises in this setup is how easy it is to distinguish statistically the optimal distorted sequence from the reference one. The robust control literature approaches this problem by using the multiplier preferences in which the distorted model is effectively constrained by a measure of relative entropy to be in some distance of the reference model.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 so that $L^{Dist}(r^t)$ is lower than $L^{DGP}(r^t)$ every period by some fixed amount, and allow the

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 Anderson et al. (2003). The difference here is that I only consider the error probability when the reference model is the true DGP.

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 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 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.

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 $\{1, 2, 3\}$.

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).

2.4 Equilibrium concept

I consider an equilibrium concept analogous to a fully revealing rational expectations equilibrium, in which the price reveals all the information available to agents. Let $\{r^t\}$ denote the history of observed interest rate differentials up to time t , $\{r_s\}_{s=0,\dots,t}$. 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 (2.7) is the optimization problem faced by the agent that involves both a maximizing choice over bonds and minimizing solution for the distorted model.

Definition 1 *An equilibrium will consist of a conjecture $f(r^{t+1})$, an exchange rate function $s(r^t)$, a bond demand function, $b(r^t)$ and an optimal distorted sequence $\sigma_V^*(r^t)$ for $\{r^t\}$, $t = 0, 1, \dots, \infty$ 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 (2.5) to form a subjective probability distribution over $r^{t+1} = \{r^t, r_{t+1}\}$ and $f(r^{t+1})$ and satisfy the following equilibrium conditions:*

1. *Optimality: given $s(r^t)$, $\sigma_V^*(r^t)$ and $f(r^{t+1})$, the demand for bonds $b(r^t)$ is the optimal solution for the max problem in (2.7).*
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 (2.7).*
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 (2.4).*
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 (2.6). 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 The Rational Expectations Model Solution

Before presenting the solution to the model, I first solve the rational expectations version which will serve as a contrast for the ambiguity aversion model. By definition, in the rational expectations case the subjective and the objective probability distributions coincide, i.e. $P = \tilde{P}$. 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 (2.1). The optimization problem is

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

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 A shows that the FOC is

$$b_t = \frac{E_t(q_{t+1})}{\gamma \text{Var}_t(q_{t+1})} \quad (3.1)$$

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:

$$s_t = \frac{E_t(s_{t+1} - r_t)}{1 + .5\gamma \text{Var}_t(s_{t+1})} \quad (3.2)$$

I call (3.2) 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 which, given the utility function, comes from the conditional 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. To form expectations agents use the Kalman Filter which given the Gaussian and linear setup is the optimal filter for the state-space in (2.1).

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

recursion:

$$\hat{x}_{t,t} = F\hat{x}_{t-1,t-1} + K_t(y_t - H'F\hat{x}_{t-1,t-1}) \quad (3.3)$$

$$K_t = (F\Sigma_{t-1,t-1}F' + \sigma_U\sigma_U')H[H'(F\Sigma_{t-1,t-1}F' + \sigma_U\sigma_U')H + \sigma_V^2]^{-1} \quad (3.4)$$

$$\Sigma_{t,t} = (I - K_tH')(F\Sigma_{t-1,t-1}F' + \sigma_U\sigma_U') \quad (3.5)$$

where K_t is the Kalman gain.

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

$$s_t = \Gamma\hat{x}_{t,t} + \delta r_t \quad (3.6)$$

For simplicity, I assume convergence on the Kalman gain and the variance matrix $\Sigma_{t,t}$. Thus, I have $\Sigma_{t,t} \equiv \Sigma$ and $K_t^{RE} = K$ for all t . Then, as detailed in Appendix A and denoting the time-invariant conditional variance $Var(s_{t+1}|I_t)$ by σ^2 , the solution is

$$\delta = -\frac{1}{1 + .5\gamma\sigma^2} \quad (3.7)$$

$$\Gamma = -\frac{1}{1 + .5\gamma\sigma^2}H'F[(1 + .5\gamma\sigma^2)I - F]^{-1} \quad (3.8)$$

$$\sigma^2 = (\Gamma K + \delta)(\Gamma K + \delta)'Var(r_{t+1}|I_t) \quad (3.9)$$

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

$$s_{t+1} - s_t = \Gamma(\rho - 1)\hat{x}_{t,t} + \Gamma K(r_{t+1} - \rho\hat{x}_{t,t}) + \delta(\rho\hat{x}_{t,t} + \rho\xi_t + \sigma_U u_{t+1} + \sigma_V v_{t+1} - r_t) \quad (3.10)$$

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 (3.10) and using that $E(\varepsilon_{t+1}|I_t) = 0$ I get

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

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

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

Because $c > 1$, to get a lower bound on $\hat{\beta}$ (denoted by $\hat{\beta}^L$) 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 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 $std_t(s_{t+1}) = 0.029$. I use these parameters and substitute them in (3.7),(3.8),(3.9). Table 1 reports the model implied exchange rate volatility and $\hat{\beta}^L$ which is obtained by setting K to 1. The conclusion that emerges from Table 1 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.²⁴ With a low risk aversion, the model implied $\hat{\beta}$ is smaller than 1, but very close to it. Although the model implied $\hat{\beta}$ decreases with γ , even with a huge degree of absolute risk aversion the UIP regression coefficient is still positive and large. For example when $\gamma = 500$, the model implied $\hat{\beta}^L$ 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.

Table 1: Rational expectations model

Risk aversion	δ	Γ	$std_t(s_{t+1})$	$\hat{\beta}^L$
$\gamma = 2$	-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

²⁴This latter point has been made by Engel and West (2004) and Engel et al. (2007) who also show that when ρ is close to 1 these models are characterized by a low forecasting power for the interest rate differential in predicting the exchange rate change.

This discussion highlights why a model of unstructured uncertainty discussed in Section 2.2 and analyzed in Appendix B, does not fare well in this setup. That type of model is equivalent to a rational expectations framework but with higher risk aversion. Driving the coefficient to 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 (2.7), the subjective state space representation (2.5) and the market clearing condition (2.4). As in the rational expectations (RE) 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:

$$E_t^{\tilde{P}}[q_{t+1} \exp(-\gamma b_t q_{t+1})] = 0 \quad (4.1)$$

The FOC (4.1) can be rewritten as

$$s_t = E_t^{\tilde{P}} \left[s_{t+1} \frac{\mu_{t+1}}{E_t^{\tilde{P}} \mu_{t+1}} \right] - r_t \quad (4.2)$$

$$\mu_{t+1} = \exp(-\gamma b(r^t) q_{t+1}) \quad (4.3)$$

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 (4.2) 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 (4.2) 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 (2.7) the usual recursive Kalman Filter applies. Thus, *after* this sequence has been optimally chosen by the agent at date t , the recursive filter uses the data from 0 to t to form estimates of the hidden state and their MSE. As shown in Hamilton (1992) the estimates are updated according to the recursion in (3.3), (3.5). The difference with the constant volatility case is that the Kalman gain now incorporates the time-varying volatilities $\sigma_{V,t}^2$:

$$K_t = (F\Sigma_{t-1,t-1}F' + \sigma_U\sigma_U')H[H'(F\Sigma_{t-1,t-1}F' + \sigma_U\sigma_U')H + \sigma_{V,t}^2]^{-1} \quad (4.4)$$

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 is different at dates $t-1$ and t . This can be interpreted as an update, although not Bayesian in nature.

To keep track of this notation 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,

$$\begin{aligned}
\widehat{x}_{i,j}^t &= E(x_i | I_j^t) \\
\Sigma_{i,j}^t &= E[(x_i^t - \widehat{x}_{i,j}^t)(x_i^t - \widehat{x}_{i,j}^t)' | I_j^t] \\
K_i^t &= (F \Sigma_{i-1,i-1}^t F' + \sigma_U \sigma_U') H [H' (F \Sigma_{i-1,i-1}^t F' + \sigma_U \sigma_U') H + \sigma_{V,(t),i}^2]^{-1} \quad (4.5)
\end{aligned}$$

Thus $\widehat{x}_{i,j}^t$ 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 (2.7) and (4.1) 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

$$s_{t+1} = \Gamma \widehat{x}_{t+1,t+1}^{t+1} + \delta r_{t+1} \quad (4.6)$$

For the minimization in (2.7) 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 (4.6).

To understand the solution we could imagine an approximation of the per period felicity function in (2.7) to $U = -\exp(-\gamma E_t^{\bar{P}} W_{t+1} + \gamma^2 \text{Var}_t^{\bar{P}} W_{t+1})$. As done in the solution to the portfolio choice problem, up to a second order, $W_{t+1} = b_t q_{t+1}$.

Thus, $E_t^{\bar{P}} W_{t+1} = b_t (E_t^{\bar{P}}(s_{t+1}) - s_t - r_t)$ and $\text{Var}_t^{\bar{P}} W_{t+1} = b_t^2 \text{Var}_t^{\bar{P}} s_{t+1}$. Given the guess in (4.6) and the Kalman filtering formulas

$$s_{t+1} = \Gamma (I - K_{t+1}^{t+1} H') F \widehat{x}_{t,t}^{t+1} + (\Gamma K_{t+1}^{t+1} + \delta) r_{t+1} \quad (4.7)$$

The sequence $\sigma_V(r^t)$ does not affect the estimates $\widehat{x}_{t,t}^{t+1}$, 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 the precision of signals on expected excess returns

Using (4.7), 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 temporary shock $\sigma_{V,(t),t}^2$. Intuitively, a larger variance of the temporary shock implies less information for updating the estimate of the hidden persistent state. Combining these two monotonicity results, the estimate $\hat{x}_{t,t}^t$ 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 C.

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 $(r_t - H'F\hat{x}_{t-1,t-1}^t)$ is positive and is decreased if the innovation is negative.

4.1.2 Effect of the precision of signals on expected variance of 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 (4.7) and taking as given $\hat{x}_{t,t}^{t+1}, K_{t+1}^{t+1}$

$$Var_t^{\tilde{P}}s_{t+1} = (\Gamma K_{t+1}^{t+1} + \delta)(\Gamma K_{t+1}^{t+1} + \delta)'(Var_t^{\tilde{P}}r_{t+1}) \quad (4.8)$$

By the filtering formulas

$$Var_t^{\tilde{P}} r_{t+1} = H' F \Sigma_{t,t}^t F' H + H' \sigma_U \sigma_U' H + E_t^{\tilde{P}} (\sigma_{V,t+1}^2) \quad (4.9)$$

In Appendix C.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 (4.9) the latter is increasing in $\sigma_{V,(t),t}^2$ through the effect on $\Sigma_{t,t}^t$. For details see Appendix C.

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

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 (4.8). 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 (4.9) the $sign\left(\frac{\partial Var_t^{\tilde{P}} r_{t+1}}{\partial \sigma_{V,(t),t}^2}\right)$ 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 a 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 the domestic currency and is thus worried about a future domestic depreciation. It is helpful to think about investing in a currency as buying an asset whose

payoff is the interest differential and whose capital gain is the appreciation of the currency. In equilibrium the agent will essentially invest in the domestic currency if the domestic interest rate differential is positive. The higher the differential the larger the demand for this asset. The agent realizes then that this asset's capital loss will be higher the less agents will demand this asset in the next period. In equilibrium, the lower is the domestic differential next period the less is the demand for this asset next period. Since the agent is worried about capital losses when she considers investing in this asset, she in fact is concerned that the domestic differential will be lower next period. The expectation about this future differential is controlled by the estimate of the hidden state. Thus the agent facing different possible DGPs fears that this estimate, although positive since she invests in the asset, is lower than what the reference DGP implies.

To make the estimate of the hidden state of the domestic differential ($\hat{x}_{t,t}^t$) lower, 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 positive and low values when these innovations are negative. The decision rule for choosing a distorted $\sigma_{V,(t),s}$ is:

$$\begin{aligned}\sigma_{V,(t),s} &= \sigma_V^H \text{ if } b_t(r_s - F\hat{x}_{s-1,s-1}^t) < 0 \\ \sigma_{V,(t),s} &= \sigma_V^L \text{ if } b_t(r_s - F\hat{x}_{s-1,s-1}^t) > 0\end{aligned}\tag{4.10}$$

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

In Section 2.3 I introduced the restriction that the agent only considers n dates out of the last m to be different from the reference model. Out of these possible sequences, the intuition given by (4.10) 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 (4.10) 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 (4.1). To solve that problem the agent needs to form forecasts about the next period exchange rate. The conjecture in (4.6) is that $s_{t+1} = \Gamma \widehat{x}_{t+1,t+1}^{t+1} + \delta r_{t+1}$ so that forecasts of $\widehat{x}_{t+1,t+1}^{t+1}$ and r_{t+1} are needed. At time t the agent knows the equilibrium updating rule for time $t + 1$:

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

The agent will use the decision rule for $\sigma_V(r^{t+1})$ given by (4.10). 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 $\widehat{x}_{t+1,t+1}^{t+1}$ there will be a s_{t+1} based on the conjecture (4.6). This distribution of s_{t+1} will imply the distribution for $q_{t+1} = s_{t+1} - s_t - r_t$ in (4.1). 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 (4.6). 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 (4.10).
2. Use (4.10) and call the resulting optimal sequence $\sigma_V^*(r^t)$. Use the Kalman filter based on the sequence $\sigma_V^*(r^t)$ to form an estimate for $\widehat{x}_{t,t}^t$ and $\Sigma_{t,t}^t$.
3. Draw realizations for r_{t+1} from $N(H' F \widehat{x}_{t,t}^t, \text{Var}_t^{\tilde{P}} r_{t+1})$, where $\text{Var}_t^{\tilde{P}} r_{t+1}$ is defined in (4.9). 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 $\widehat{x}_{t+1,t+1}^{t+1}$ and use the conjecture in (4.6) 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 (4.1) to solve for s_t^* .
6. If $\text{sign}(s_t^*) = \text{sign}(b_t)$ 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 (2.7). 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 (4.6). The consistency of beliefs in the definition of equilibrium (see Section 2.4) 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. 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 (4.6) 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. Realizations for r_{t+1} that would result under the rational expectations model in switching investment 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 $\Gamma\hat{x}_{t,t}^t + \delta r_t$. The parameters Γ and δ that minimize the distance between the two objects even in subsamples are characterized by a smaller response of s_t to the estimate $\hat{x}_{t,t}^t$. These parameters would correspond to the rational expectations model solution if there would be less persistence in the hidden state.

The market clearing condition states that $b_t = .5s_t$. In the RE 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 (3.1). 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

$$s_t = \frac{E_t^{\tilde{P}}(s_{t+1} - r_t)}{1 + .5\gamma Var_t^{\tilde{P}}(s_{t+1})} \quad (4.11)$$

I call (4.11) 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 (4.2): $E_t^{\tilde{P}^N}[q_{t+1}] = E_t^{\tilde{P}}\left[\frac{d\tilde{P}^N}{d\tilde{P}}q_{t+1}\right]$.

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. 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 is²⁵:

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

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:

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

where μ_{t+1} is defined in (4.3).

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 $s_t < 0$.²⁶ The payoff on a dollar bet for the unhedged carry trade

²⁵The 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 $z_{t+1}^C(k_t) = \max(0, s_{t+1} - k_t) - C(k_t)$.

²⁶In simulations, the coefficient of correlation between s_t and r_t is around -0.97 . The correlation is not -1 because the exchange rate depends mostly on the sign of the hidden state estimate and this can be different from the sign of the interest rate. Moreover, due to the ‘‘inaction’’ effect in some cases agents choose not to invest at all if r_t is too close to 0. In these few case I assume that the payoff to the carry trade is 0.

strategy is:

$$\begin{aligned} z_{t+1} &= r_t - (s_{t+1} - s_t) \text{ if } b_t < 0 \\ z_{t+1} &= (s_{t+1} - s_t) - r_t \text{ if } b_t > 0 \end{aligned} \tag{4.12}$$

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 :

$$\begin{aligned} z_{t+1}^H(k_t) &= z_{t+1} + z_{t+1}^C(k_t) \text{ if } b_t < 0 \\ z_{t+1}^H(k_t) &= z_{t+1} + z_{t+1}^P(k_t) \text{ if } b_t > 0 \end{aligned} \tag{4.13}$$

where z_{t+1} is the payoff to the unhedged carry trade defined in (4.12).

4.4 Parameterization

In the benchmark case, the reference model is a state space representation with constant volatilities as in (2.1). I estimate the relevant parameters by using Maximum Likelihood on data on interest rate differential. 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. 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 10.

Table 11 reports results for the Maximum Likelihood estimation of (2.1) on interest rate differential data. First, I find a very high degree of persistence in the state evolution. The table reports values for 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.²⁷ The benchmark parameterization is a version of (2.1) which averages across these heterogeneity. 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.

²⁷Performing 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 .

For the state space defined in (2.1) $H' = [1 \ 0 \ 0 \ 0]$, $\sigma_U = [\sigma_U^{DGP} \ 0 \ 0 \ 0]$ and

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

Table 2 reports the benchmark parameterization.

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

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.²⁸

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. As introduced in Section 2.3 for computational reasons I also make a restrictive assumption about the distorted sequence so that at time t the agent considers the n points of possible distortions only for dates $t - m + 1$ through t . 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 ($L^{Dist}(r^t)$).

Table 3 reports some statistics for the likelihood comparison $L^{Dist}(r^t) - L^{DGP}(r^t)$, computed for a sample of $T = 300$, for various cases. The Table reports the mean average and the standard deviation of this difference and the percent of times for which this difference is positive. The standard deviation of these statistics is then computed across the $N = 1000$ simulations. Column (1) reports the results for the benchmark parameterization in which $m = n = 2$. It also shows that for about 17% of cases $L^{DGP}(r^t)$

²⁸Also 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.

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 -

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 across 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 (3.7), (3.8) and (3.9) 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 equation. 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. Figures 1 and 2 illustrate for one simulation of the model the exchange rate path and the estimate of the hidden state of the domestic interest rate differential under the distorted expectations version and the rational expectations (RE) version. For comparison, the same exogenous driving process is used but the estimates are computed under the different model solutions. When the agent is investing in the foreign bond, i.e. $b_t > 0$ and $s_t > 0$, she is concerned that the negative estimate of the hidden state is in fact less negative than what the reference model would imply. Her investment position will reflect her pessimistic assessment of the future distribution and she will invest less compared to the RE model. Thus $s_t < s_t^{RE}$ when $s_t > 0$ and $s_t > s_t^{RE}$ when $s_t < 0$ where s_t^{RE} denotes the exchange rate under RE.

Figure 1: Model generated exchange rate path

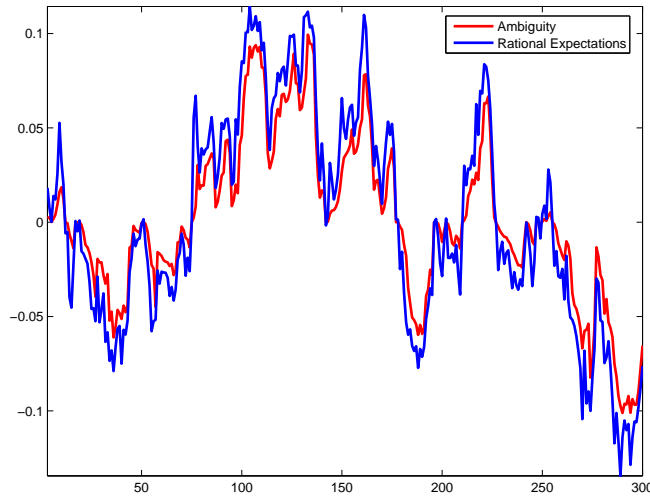


Table 4: Correlations exchange rates

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

Figure 1 shows 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)). To investigate this asymmetric response further, Table 4 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.

For the same simulation that generated the exchange rate path in Figure 1, Figure 2 illustrates the estimate of the hidden state of the domestic differential. The same intuition regarding the asymmetric response applies here. Table 5 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.

Figure 2: Model generated estimates of hidden state

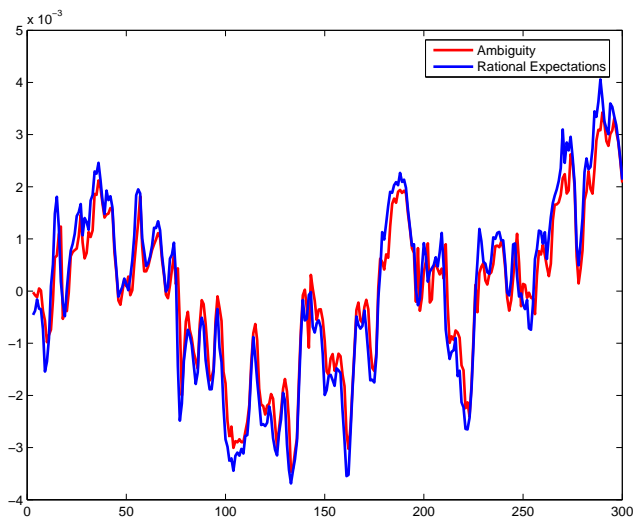


Table 5: Correlations estimate of hidden state

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

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 (4.2)). The UIP regression is:

$$s_{t+1} - s_t = \beta r_t + \varepsilon_{t+1} \quad (5.1)$$

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

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

The message of the time-varying risk premia story is that the term $\frac{\text{cov}(\text{cov}(s_{t+1}\lambda_{t+1}), r_t)}{\text{var}(r_t)}$ 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

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

As explained at length in the RE case, the conditional variance $\text{Var}_t^{\tilde{P}}(s_{t+1})$ 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 $\text{Var}_t^{\tilde{P}}(s_{t+1})$ 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

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

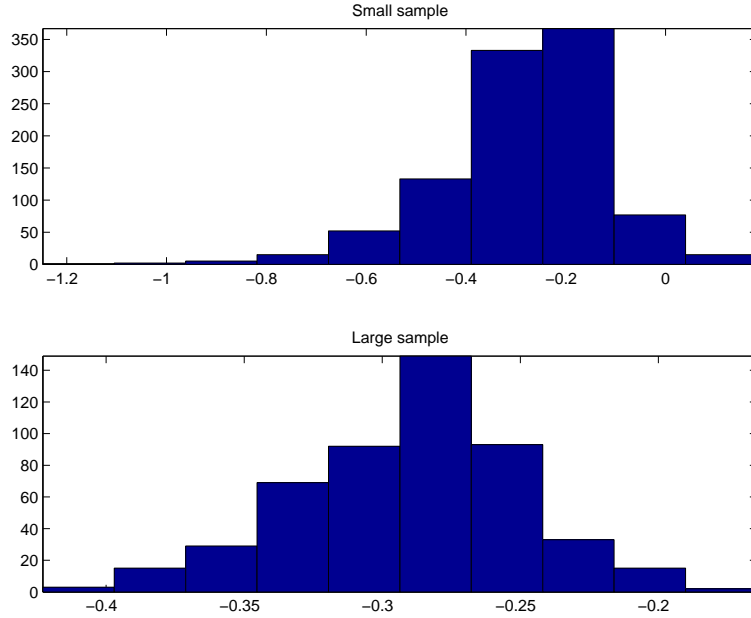
results I will make use of the fact that the $1 + .5\gamma Var_t^{\tilde{P}}(s_{t+1})$ is approximately equal to 1. In a richer environment the ambiguity aversion theory presented in this paper can be incorporated into models that take more seriously the risk premia corrections. With the covariances becoming effectively zero the risk neutral UIP implies that if the true DGP is the distorted model then $\hat{\beta} = 1$ 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 (5.1) 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 is the histogram of the UIP coefficients across many repeated samples. The top panel plots the histogram of the estimated UIP coefficients on $N = 1000$ samples of $T = 300$. It shows that the vast majority of the estimates are negative. For the large sample of $T = 3000$ the bottom panel of Figure 3 indicates the distribution of the estimates and shows that there are no positive values obtained.

Column (2) of Table 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

Figure 3: Model implied UIP coefficients



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 Column (2) the gain under the reference model is closer to the one implied by σ_V^H , so the underreaction effect is less active. That is the reason why the estimated UIP coefficient is positive.

Column (3) of Table 6 shows that when there is significantly less persistence in the state evolution the model cannot account for the UIP puzzle. For that experiment 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 puzzle.

These exercises show that the two main features of the true DGP that are required for the theory to succeed are relatively small temporary shocks and large persistence of the hidden state. The benchmark parameterization is characterized by these conditions because the data strongly suggests such a calibration. It is important to note though that Columns (2) and (3) of Table 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 (4.6). 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 $m = 2$.

Figure 4: Model generated carry trade payoffs

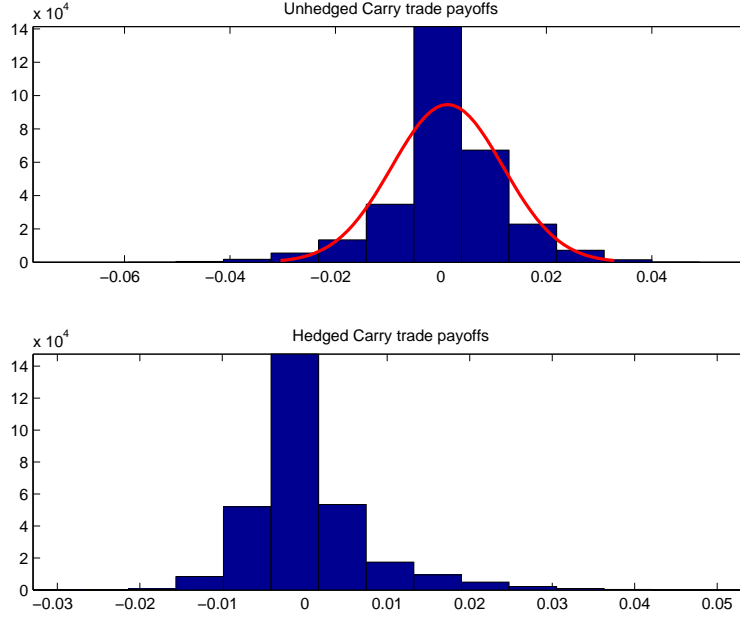


Table 7: Model implied statistics for the unhedged carry trade returns

	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)

The payoff on a dollar bet for the unhedged carry trade strategy is defined in (4.12). The model implied non-annualized monthly carry trade returns are described in Table 7. 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 is 0.0017. For the data analyzed in Table 10 the average mean payoff for the carry trade strategy is 0.0033. This does not take into account transactions costs. Burnside et al. (2008) (henceforth 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 7 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 10 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 4 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 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 10. The model implies a negative skewness of -0.42 while the average for the countries in Table 10 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 Brunnermeier et al. (2008). It involves checking whether periods (countries in Brunnermeier et al. (2008)) characterized by a higher domestic currency also experience a negative skewness in the excess 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

Table 8: Model implied statistics for the “crash risk” regressions

$T = 300$				$T = 3000$			
$\widehat{\beta}_2$	$t_{\widehat{\beta}_2}$	$\widehat{\beta}_3$	$t_{\widehat{\beta}_3}$	$\widehat{\beta}_2$	$t_{\widehat{\beta}_2}$	$\widehat{\beta}_3$	$t_{\widehat{\beta}_3}$
-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)

$N = 20000$ 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 8 reports the results of the regression

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

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

Consistent with the data, 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.

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 8 also reports results for the regression

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

Table 9: 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)

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

To measure the hedged carry trade I use the definition in (4.13) and as in BEKR I use at-the-money options by setting $k_t = s_t$. Table 9 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).²⁹ BEKR report values that imply that the average individual country pair loss

²⁹Jurek (2008) uses a more complicated version of hedging that also takes into account the delta exposure

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.

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.³⁰ 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 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 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 small probability rare disaster event that does not materialize in the sample under the latter distribution. The mechanism presented here delivers a realized distribution for the exchange rates that can have a different shape than the subjective. Unlike the peso problem, there is no reasons to expect that any property of these distributions is the same. Their mean are not equal 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

of the option.

³⁰This explanation has been put forward by recent papers such as Farhi and Gabaix (2008). In their model there is a rare disaster that can happen with a small, time-varying probability.

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) find that these results are sensitive to the recursive identification assumptions and that the peak of the exchange rate response is imprecisely estimated. Scholl and Uhlig (2006) use sign restrictions and also find evidence for delayed overshooting. For the country pairs they analyze the estimated peak occurs within a year or two. Although they differ in their estimates of the length of the delayed overshooting effect all these identifying approaches reach a robust conclusion: following a monetary policy shock there are significant deviations from the UIP.

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

To generate the impulse response of the exchange rate to a shock to the interest rate differential, assume that the economy starts in steady state. Thus, $r_{t-1} = 0$, $b_{t-1} = 0$, $E_{t-1}^{\tilde{P}} s_t = E_{t-1}^P s_t = 0$. At time t there is a positive shock to the interest rate differential. 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

$$\begin{aligned} r_t &= x_t + \sigma_V v_t \\ x_t &= \rho x_{t-1} + \sigma_U u_t \end{aligned} \tag{5.2}$$

where $\rho = 0.9718$. The values for σ_V and σ_U are the same as in Section 4.4.

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 (RE) 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 RE and $\hat{x}_{t,t}^t$ is the estimate under the ambiguity aversion model. A feature of this experiment is that $\hat{x}_{t,t}^{RE} = x_t$ while $\hat{x}_{t,t}^t \neq x_t$.

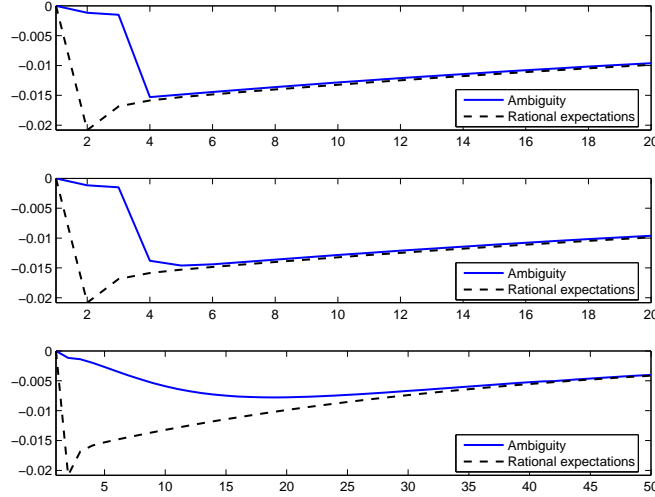
Figure 5 plots the dynamic response of the exchange rate to the observed increase in r_t . The RE version features the Dornbusch (1976) overshooting result in which the peak of the impact is at time t . Consider the decision at time t . The agent sees the increase in the interest rate differential but she is worried about a significant depreciation at time $t + 1$. In equilibrium 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:

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

In this setting $\hat{x}_{t+1,t+1}^{t+1}$ is higher than $\hat{x}_{t,t}^t$ and a further appreciation occurs. Because $r_t - \rho \hat{x}_{t-1,t-1} > 0$ 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 future periods. Eventually, the estimate of the hidden state converges to the RE case.

The top plot in Figure 5 shows the evolution of s_t in the benchmark specification in

Figure 5: Delayed overshooting



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 5 even after converging the exchange rate under ambiguity and RE 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. 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 5 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 considerations 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 applied to other settings that involve forecastability of excess returns. Bacchetta et al. (2008) use survey data to conclude that most of the predictability of excess returns in bond, stock and foreign exchange market is caused by predictability of expectational errors. Interestingly, in the stock market a similar impulse response as the delayed overshooting puzzle has been documented (as for example in Hong and Stein (1999)): stock prices tend to respond slowly to new public releases. For the bond market, Piazzesi and Schneider (2009) find that a model with adaptive learning and recursive utility is able to explain the statistical premia in the bond market. Studying a model in which the exogenously-specified adaptive learning is endogenized to the extent that it reflects the cautious behavior of an ambiguity-averse agent seems a fruitful extension. Another extension could be investigating 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). 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.

A 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 as in (3.1):

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

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 (3.3), (3.4) and (3.5). 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:

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

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

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

Denoting $\text{Var}_t(s_{t+1} | I_t)$ by σ^2 the UIP condition (3.2) 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 (3.2) and solve for the unknown coefficients Γ , δ and σ^2 to get the equations (3.7), (3.8) and (3.9).

B Multiplier preferences and unstructured uncertainty

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

$$\max_b \min_{\tilde{P} \in \Phi} E^{\tilde{P}}[U(c(b; \varepsilon))] + \theta R(\tilde{P}|P) \quad (\text{B.1})$$

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

$$R(\tilde{P}|P) = \begin{cases} \int_{\Omega} \log\left(\frac{d\tilde{P}}{dP}\right) d\tilde{P} & \text{if } \tilde{P} \text{ is absolutely continuous w.r.t } P \\ \infty & \text{otherwise} \end{cases} \quad (\text{B.2})$$

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) the problem in (B.1) is equivalent to³¹:

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

Note that now the subjective probability distribution becomes the reference model P . 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 .

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

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 (2007) consider the problem in which the agent takes the process in (2.1) as an

³¹The 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.

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

$$r_{t+1} \stackrel{\tilde{P}}{\sim} N(F\widehat{x}_{t,t}^{RE} + F\varepsilon_t + \epsilon_{t+1}^V + \epsilon_{t+1}^U, F^2\Sigma_{t,t} + \sigma_U^2 + \sigma_V^2). \quad (\text{B.3})$$

Note that under the reference model

$$r_{t+1} \stackrel{P}{\sim} N(F\widehat{x}_{t,t}^{RE}, F^2\Sigma_{t,t} + \sigma_U^2 + \sigma_V^2) \quad (\text{B.4})$$

The relative entropy defined in (B.2) for the implied distributions \tilde{P} in (B.3) and P in (B.4) 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 (B.1) imply:

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

To solve for an equilibrium in this setup use a guess and verify approach and conjecture that the solution for s_{t+1} is $s_t = \Gamma \widehat{x}_{t,t}^{RE} + \delta r_t$ 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 $E_t^{\tilde{P}}(s_{t+1}) = (\Gamma + \delta)F\hat{x}_{t,t}^{RE} + (\Gamma K + \delta)\omega_{t+1}$. Replacing $E_t^{\tilde{P}}(s_{t+1})$ in (B.5) 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 :

$$\Gamma = [1 + \frac{Var_t^P(s_{t+1})}{2\theta}]^{-1}(\Gamma + \delta)F \quad (B.6)$$

$$\delta = -[1 + \frac{Var_t^P(s_{t+1})}{2\theta}]^{-1} \quad (B.7)$$

The solution to (B.6) and (B.7) is identical to the one in (3.7) and (3.8) 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} .

C 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 (4.7)

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

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

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

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 $\sigma_V(r^t)$

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

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

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

This establishes *Proposition 1*.

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

Proof: The variance $\text{Var}_t^{\tilde{P}} W_{t+1} = b_t^2 \text{Var}_t^{\tilde{P}} s_{t+1}$. In turn, using the conjecture (4.7) and taking as given $\hat{x}_{t,t}^{t+1}, K_{t+1}^{t+1}$

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

By the Kalman filtering formulas I get

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

Use the formula in (4.7) 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) then

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

Thus,

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

This establishes *Proposition 2*.

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

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

The sign of this derivative is :

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

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

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

This total derivative equals:

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

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 $b_t < 0$.

C.1 Expected variance of future temporary shocks

In Section 2.3 I introduce the constant n 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 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 $\sigma_{V, s} \neq \sigma_V$ only for n out of t times in the sample of observable data she has.³² Thus, in forming $E_t^{\tilde{P}} \sigma_{V, t+1}$, 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. In this model, where agents dislike variance, they will choose at time 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

³²The 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.

n/t . For simplicity, I will not keep track specifically of this choice when referring to the minimization part in (2.7). When this object appears in the forecast I make use of the above argument and conclude that

$$E_t^{\tilde{P}} \sigma_{V,t+1} = \sigma_V^H \frac{n}{t} + \sigma_V \left(1 - \frac{n}{t}\right) \tag{C.2}$$

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 $E_t^{\tilde{P}} \sigma_{V,t+1} = \sigma_V$.

D Supplementary tables

Table 10: Empirical UIP regression and carry trade returns

Country	UIP regression		Carry trade returns				
	α	β	Mean	Standard Deviation	Sharpe Ratio	Skewness	Kurtosis
Belgium [†]	0.000 (0.002)	-0.593 (0.612)	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)	0.0019 (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)	0.0051 (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)	0.0012 (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)	0.0023 (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)	0.0025 (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)	0.0010 (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)	0.0053 (0.0015)	0.030 (0.003)	0.177 (0.052)	-0.029 (0.376)	4.91 (0.91)
Average	0.003	-0.85	0.0033	0.0306	0.11	-0.242	4.08

Notes: The first 2 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 carry trade returns involving the USD and the pairing FCU. 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 11: 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 (2.1). 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.

References

- ALVAREZ, F., A. ATKESON, AND P. J. KEHOE (2008): “Time-varying risk, interest rates, and exchange rates in general equilibrium,” Staff Report 371, Federal Reserve Bank of Minneapolis.
- ANDERSEN, T. G., T. BOLLERSLEV, F. X. DIEBOLD, AND C. VEGA (2003): “Micro Effects of Macro Announcements: Real-Time Price Discovery in Foreign Exchange,” *American Economic Review*, 93, 38–62.
- ANDERSON, E. W., L. P. HANSEN, AND T. J. SARGENT (2003): “A Quartet of Semigroups for Model Specification, Robustness, Prices of Risk, and Model Detection,” *Journal of the European Economic Association*, 1, 68–123.
- BACCHETTA, P., E. MERTENS, AND E. VAN WINCOOP (2008): “Predictability in Financial Markets: What Do Survey Expectations Tell Us?” *Journal of International Money and Finance*, forthcoming.
- BACCHETTA, P. AND E. VAN WINCOOP (2008): “Infrequent Portfolio Decisions: A Solution to the Forward Discount Puzzle,” *American Economic Review*, forthcoming.
- BANSAL, R. AND I. SHALIASTOVICH (2007): “Risk and return in Bond, Currency and Equity Markets,” Mimeo.
- BARILLAS, F., L. P. HANSEN, AND T. J. SARGENT (2008): “Doubts or Variability?” Mimeo.
- BASAR, T. AND P. BERNHARD (1995): *H_∞ Optimal Control and Related Minimax Design Problems*, Birkhuser, Boston.
- BENIGNO, P. (2007): “Portfolio Choices with Near Rational Agents: A Solution for Some International-Finance Puzzles,” NBER Working Papers 13173.
- BERNANKE, B. S. AND J. BOIVIN (2003): “Monetary Policy in a Data-Rich Environment,” *Journal of Monetary Economics*, 50, 525–546.
- BRUNNERMEIER, M., S. NAGEL, AND L. PEDERSEN (2008): “Carry Trades and Currency Crashes,” *NBER Macroeconomics Annual*, forthcoming.
- BURNSIDE, A. C., M. S. EICHENBAUM, I. KLESHCHELSKI, AND S. REBELO (2008): “Do Peso Problems Explain the Returns to the Carry Trade?” Mimeo.
- BURNSIDE, C. (2008): “Carry Trades and Currency Crashes: A Comment,” *NBER Macroeconomics Annual*, forthcoming.
- CHINN, M. D. AND J. A. FRANKEL (2002): “Survey Data on Exchange Rate Expectations: More Currencies, More Horizons, More Tests,” in *Monetary Policy, Capital Flows and Financial Market Development in the Era of Financial Globalization: Essays in Honor of Max Fry*, ed. by W. Allen and D. Dickinson, London, Routledge, 145–167.

- CHINN, M. D. AND G. MEREDITH (2005): "Testing Uncovered Interest Parity at Short and Long Horizons during the Post-Bretton Woods Era," NBER Working Papers 11077.
- DORNBUSCH, R. (1976): "Expectations and Exchange Rate Dynamics," *Journal of Political Economy*, 84, 1161–76.
- EICHENBAUM, M. AND C. L. EVANS (1995): "Some Empirical Evidence on the Effects of Shocks to Monetary Policy on Exchange Rates," *The Quarterly Journal of Economics*, 110, 975–1009.
- ENGEL, C. (1996): "The Forward Discount Anomaly and the Risk Premium: A Survey of Recent Evidence," *Journal of Empirical Finance*, 3, 123–192.
- ENGEL, C. AND J. 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, C., N. C. MARK, AND K. D. WEST (2007): "Exchange Rate Models Are Not as Bad as You Think," NBER Working Papers 13318.
- ENGEL, C. AND K. D. WEST (2004): "Accounting for Exchange-Rate Variability in Present-Value Models When the Discount Factor Is Near 1," *American Economic Review*, 94, 119–125.
- EPSTEIN, L. G. AND M. SCHNEIDER (2007): "Learning Under Ambiguity," *Review of Economic Studies*, 74, 1275–1303.
- EVANS, M. D. D. AND K. K. LEWIS (1995): "Do Long-Term Swings in the Dollar Affect Estimates of the Risk Premia?" *Review of Financial Studies*, 8, 709–42.
- FAMA, E. F. (1984): "Forward and Spot Exchange Rates," *Journal of Monetary Economics*, 14, 319–338.
- FARHI, E. AND X. GABAIX (2008): "Rare Disasters and Exchange Rates," NBER Working Papers 13805.
- FAUST, J. AND J. H. ROGERS (2003): "Monetary Policy's Role in Exchange Rate Behavior," *Journal of Monetary Economics*, 50, 1403–1424.
- FROOT, K. A. AND J. A. FRANKEL (1989): "Forward Discount Bias: Is It an Exchange Risk Premium?" *The Quarterly Journal of Economics*, 104, 139–61.
- FROOT, K. A. AND R. H. THALER (1990): "Anomalies: Foreign Exchange," *Journal of Economic Perspectives*, 4, 179–92.
- GILBOA, I. AND D. SCHMEIDLER (1989): "Maxmin Expected Utility with Non-unique Prior," *Journal of Mathematical Economics*, 18, 141–153.
- GOURINCHAS, P.-O. AND A. TORNELL (2004): "Exchange Rate Puzzles and Distorted Beliefs," *Journal of International Economics*, 64, 303–333.
- GRILLI, V. AND N. ROUBINI (1996): "Liquidity Models in Open Economies: Theory and empirical evidence," *European Economic Review*, 40, 847–859.

- HANSEN, L. P. AND R. J. HODRICK (1980): “Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis,” *Journal of Political Economy*, 88, 829–53.
- HANSEN, L. P. AND T. J. SARGENT (2007): “Recursive Robust Estimation and Control without Commitment,” *Journal of Economic Theory*, 136, 1–27.
- (2008): *Robustness*, Princeton University Press.
- HONG, H. AND J. C. STEIN (1999): “A Unified Theory of Underreaction, Momentum Trading and Overreaction in Asset Markets,” *Journal of Finance*, 54, 2143–2184.
- JUREK, J. W. (2008): “Crash-neutral Currency Carry Trades,” Mimeo.
- KAMINSKY, G. (1993): “Is There a Peso Problem? Evidence from the Dollar/Pound Exchange Rate, 1976-1987,” *American Economic Review*, 83, 450–72.
- KLESHCHELSKI, I. AND N. VINCENT (2007): “Robust Equilibrium Yield Curves,” Mimeo.
- KLIBANOFF, P., M. MARINACCI, AND S. MUKERJI (2005): “A Smooth Model of Decision Making under Ambiguity,” *Econometrica*, 73, 1849–1892.
- LEWIS, K. K. (1989): “Changing Beliefs and Systematic Rational Forecast Errors with Evidence from Foreign Exchange,” *American Economic Review*, 79, 621–36.
- (1995): “Puzzles in International Finance Markets,” in *Handbook of International Economics*, ed. by G. G. Grossman and K. Rogoff, Elsevier Science, 1913–1971.
- LI, M. AND A. TORNELL (2008): “Exchange Rates Under Robustness: The Forward Premium Puzzle and Momentum,” Mimeo.
- LYONS, R. (2001): *The Microstructure Approach to Exchange Rates*, MIT Press:Cambridge, MA.
- MACCHERONI, F., M. MARINACCI, AND A. RUSTICHINI (2006): “Ambiguity Aversion, Robustness, and the Variational Representation of Preferences,” *Econometrica*, 74, 1447–1498.
- PIAZZESI, M. AND M. SCHNEIDER (2009): “Trend and Cycle in Bond Premia,” Mimeo.
- SARNO, L. (2005): “Towards a Solution to the Puzzles in Exchange Rate Economics: Where Do We Stand?” *Canadian Journal of Economics*, 38, 673–708.
- SCHOLL, A. AND H. UHLIG (2006): “New Evidence on the Puzzles: Monetary Policy and Exchange Rates,” Working Paper 5, Society for Computational Economics.
- STRZALECKI, T. (2007): “Axiomatic Foundations of Multiplier Preferences,” Mimeo.
- VERDELHAN, A. (2006): “A Habit-Based Explanation of the Exchange Rate Risk Premium,” Mimeo.