Can Risk Premium Explain the Uncovered Interest Parity Puzzle?  
A Nonparametric Approach  

Jing Li* and Christopher J. Adams†  

Miami University  

Abstract  

Fama (1984) suggests that the uncovered interest parity puzzle may be explained by the omitted variable bias caused by the unobserved risk premium. This paper applies the Hodrick-Prescott filter to the forecast error based on the uncovered interest rate parity, and directly estimates the risk premium as the slow-evolving smooth component. Furthermore, the Fourier Flexible Form with a small number of low frequency components is employed to obtain a similar estimate of the risk premium. We find strong evidence that the risk premium and interest differential are negatively correlated, and weaker evidence that the covariance is less than negative variance of the interest differential.

Keywords: Uncovered Interest Parity, Risk Premium, Fama’s Beta, Hodrick-Prescott Filter, Fourier Flexible Form  

JEL Classification: F31, F37  

*Corresponding Author. Department of Economics, Miami University, 800 E. High Street, Oxford, OH 45056. Phone: 513-529-4393. Fax: 513-529-6992. Email: lij14@miamioh.edu.  
†Department of Economics, Miami University, 800 E. High Street, Oxford, OH 45056. Email: adamsjc2@miamioh.edu.
1 Introduction

As an extension of the law of one price in international finance, the uncovered interest parity (UIP) hypothesis states that foreign and domestic assets of comparable quality should yield same expected returns after being adjusted for movement in exchange rates. However, historical data usually support the success of the strategy of investing in foreign assets when foreign interest rates rise and buying domestic assets vice versa, regardless of the exchange rate. This empirical anomaly is called uncovered interest parity puzzle. In terms of the regression-based test for UIP, this puzzle amounts to the coefficient of interest differential being less than the unity, and sometimes becoming even negative, in the Fama (1984) regression\(^1\).

One way to reconcile the Fama regression with the UIP is to recognize the bias in the estimated beta coefficient caused by omitted variables—the one receiving most attention in the literature is the risk premium. Fama (1984) shows that (I) a negative covariance between the risk premium and interest differential can lead to a downward bias; (II) to explain a negative slope coefficient, that covariance has to be less than the negative variance of interest differential. In other words, in the event of rising domestic interest rate, investors would treat foreign assets as safer alternative requiring less risk premium, so foreign currency can see appreciation in the next period smaller than the UIP would imply. Moreover, if the risk premium falls significantly, the foreign currency may even depreciate against the home currency.

The main challenge to empirically verify Fama’s theory is the lack of observable risk premium. By nature, the risk premium is unobserved since it depends on personal perceived risk and attitude toward the risk. To see this explicitly, consider a “structural model” of the

\(^1\)A review of the UIP puzzle literature is Froot and Thaler (1990).
risk premium\(^2\) denoted by \(\theta_t\):

\[
E (\theta_t|\Omega_t) = \frac{\text{cov} \left(\frac{u'(c_{t+1})}{p_{t+1}} \frac{p_t}{u'(c_t)} ; \frac{S_{t+1} - S_t}{S_t} \mid \Omega_t\right)}{-E \left(\frac{u'(c_{t+1})}{p_{t+1}} \frac{p_t}{u'(c_t)} \mid \Omega_t\right)}
\]

(1)

where \(u'\) represents the first order derivative of the utility function, \(c_t\) is consumption at time \(t\), \(p_t\) is the domestic currency price of the consumption good, \(S_t\) is the domestic currency price of one unit of foreign currency (spot exchange rate), and \(\Omega_t\) is the information set for conditional covariance and conditional mean. Basically, Equation (1) is a simple version of the first order condition for the generalized Lucas (1978) intertemporal asset pricing model.

For our purpose, it suffices to emphasize that restrictive assumptions about the form of utility function, the information set, and the whole structural model need to be made\(^3\) in order to obtain a “parametric” estimate of the risk premium based on (1). For the sake of robustness and flexibility, this paper instead adopts a simple “nonparametric” or “reduced-form” methodology of applying the Hodrick-Prescott (HP) filter to the Fama regression forecast error with the UIP imposed, and extracting the slow-evolving trend component as the estimate for the risk premium\(^4\). In short, our nonparametric approach makes much weaker assumptions relative to the structural-form estimate.

This identification strategy is built upon the smoothness of risk premium, which comes from two sources according to (1). First, it can be rationalized by the consumption smoothing\(^5\) \(c_t \approx c_{t+1}\), sticky price \(p_t \approx p_{t+1}\), and near time-invariant utility function in the structural model (1). Second, even if there may be heterogeneity in personal utility and information set,

\(^2\)See Equation (6) in Mark (1985) for the derivation.

\(^3\)For instance, the Constant Relative Risk Aversion (CRRA) utility function is commonly used; information set is typically assumed to contain all relevant information; the economic agent is supposedly rational trying to maximize expected discounted utility, etc.

\(^4\)Similar methodology is used in Ball and Mankiw (2002) where the HP filter is applied to estimate the slow-evolving natural unemployment rate.

\(^5\)The literature on consumption smoothing and sticky prices is vast, see Friedman (1956) and Taylor (1980) for instance.
at the aggregate level, the central limit theorem is able to add another layer of smoothness.

As a robustness check, this paper also employs the Fourier Flexible Form (FFF) to obtain an alternative estimate of the risk premium\(^6\). In theory the Fourier series with trigonometric terms is able to approximate any absolutely integrable function including risk premia. In that regard, the Fourier method is even more flexible than the HP filter as the smoothness assumption is not needed. Despite that difference, we show that the risk premium estimated by the FFF with a small number of frequencies is very similar to that generated by the HP filter. This finding provides support for the smoothness assumption. Moreover, it indicates that the HP trend is unlikely to be spurious, thus a major concern raised by Hamilton (2018) is resolved\(^7\).

Our estimates of risk premium contrast with Domowitz and Hakkio (1985), where the authors consider the ARCH-in-Mean model of Engle et al. (1987), assuming that the risk premium is function of conditional variance of market forecast errors. In our view that method is limited since it is hard to believe why other moments such as kurtosis are irrelevant for the risk premium. With equal variances, the assets with higher probabilities of extreme values are more risky, so investors would require higher expected returns. By assuming only smoothness, our finding can be seen as a complement to Domowitz and Hakkio (1985).

The second contribution of this paper is to use the estimated risk premium to directly verify the explanation for the UIP puzzle suggested by Fama (1984). Overall, our results strongly support the first part of the explanation, and the evidence for the second part is mixed, depending on whether the interest differential or its smooth component is used to compute the covariance.

\(^6\)FFF is discussed in Gallant (1981) and is used to approximate unknown function forms such as structural breaks and trend functions. See, for instance, Becker et al. (2004), Enders and Lee (2012), and Enders and Li (2015).

\(^7\)Hamilton’s main concerns about the HP filter are largely irrelevant for this study since our primary goal is to acquire smoothness, not to extract a business cycle that has to be stationary.
2 Methodology

2.1 The UIP Forecast Error Decomposition

Let $ID_t$ denote the home minus foreign nominal interest rate differential, and $\Delta s_{t+1} = \log S_{t+1} - \log S_t$ be the one-period change in log spot exchange rate (dollars per one fx). We are interested in the Fama regression

$$\Delta s_{t+1} = \beta_0 + \beta_1 ID_t + u_t$$

(2)

and testing the UIP hypothesis

$$H_0 : \beta_1 = 1$$

(3)

which states that, ceteris paribus, the interest differential is an unbiased predictor for the change in spot exchange rate. The interest differential is just one factor driving the exchange rate, thus we expect that the error term $u_t$ contains omitted or confounding variables that are potentially correlated with the interest differential, and it is the omitted variable that results in biased OLS estimation, i.e., the historical data tend to reject (3) even if UIP is true.

Previous empirical studies often find that the OLS coefficient less than unity $\hat{\beta}_1 < 1$. That is, the currency of a higher interest rate country does not depreciate as much as the UIP would imply. Moreover, it is not uncommon to report a negative $\hat{\beta}_1$, meaning that the high-interest-rate currency actually appreciates in the period immediately after its interest rate rises. This finding is at odds with UIP, so is called the UIP puzzle.

To explain the UIP puzzle econometrically, consider the well-known asymptotic relation-

---

8Rigorously speaking the UIP also implies $\beta_0 = 0$. We follow the literature by focusing on $\beta_1$ only.

9See Froot and Thaler (1990) for a literature review.
ship between $\beta_1$ and its OLS estimate $\hat{\beta}_1$:

$$\hat{\beta}_1 \rightarrow^p \beta_1 + \frac{\text{cov}(ID_t, u_t)}{\text{var}(ID_t)}$$

(4)

where $\rightarrow^p$ denotes the convergence in probability. This is basically the law of large number applied in the regression setting. According to (4), the OLS estimator is generally inconsistent if some omitted variables contained in $u_t$ are correlated with the regressor $ID_t$, and the bias is proportional to the covariance $\text{cov}(ID_t, u_t)$. In particular, a downward bias $\hat{\beta}_1 < \beta_1$ necessitates a negative covariance:

$$\hat{\beta}_1 < \beta_1 \Rightarrow \text{cov}(ID_t, u_t) < 0$$

(5)

Explaining the negative $\hat{\beta}_1$ found in literature requires conditions more restrictive than (5)—the covariance has to be sufficiently negative. More explicitly, we can show

$$\hat{\beta}_1 < 0 \Rightarrow \beta_1 + \frac{\text{cov}(ID_t, u_t)}{\text{var}(ID_t)} < 0$$

$$\Rightarrow 1 + \frac{\text{cov}(ID_t, u_t)}{\text{var}(ID_t)} < 0$$

(6)

$$\Rightarrow \text{cov}(ID_t, u_t) < -\text{var}(ID_t)$$

(7)

where (6) imposes the UIP hypothesis (3). The following proposition, first shown by Fama (1984), summarizes those reasoning

**Proposition 1** *An estimated Fama’s beta coefficient less than the value hypothesized by UIP reflects negative covariance between the error term and interest differential (P1); a negative Fama’s beta coefficient reflects the covariance being less than negative variance of interest rate differential (P2).*

The main challenge to verify Proposition 1 (or show that the omitted variable indeed has a
bearing on the UIP puzzle) is that the error term $u_t$ is unobserved. Among the many potential confounders\textsuperscript{10}, the one receiving most discussion in literature is the risk premium\textsuperscript{11}.

One contribution of this paper is to propose a nonparametric estimate of the risk premium, which does not assume a specific structural model or function form. We only assume smoothness of risk premium, and even that assumption can be relaxed by using the Fourier Flexible Form. Toward that end we first treat the UIP as a forecast error decomposition of the change in log exchange rate

$$\Delta s_{t+1} = ID_t + u_t$$  \hspace{1cm} (8)

$$u_t = \theta_t + e_t$$  \hspace{1cm} (9)

Decomposition (8) in effect imposes the UIP restriction $\beta_0 = 0, \beta_1 = 1$ on the Fama regression (2), where the interest differential is the part of exchange rate change explained by the UIP, whereas the error term is the unexplained part or forecast error. Moreover, (9) divides the composite forecast error $u_t$ into the risk premium $\theta_t$ and an idiosyncratic error term $e_t$.

Jointly, the decomposition of (8) and (9) just says that the movement of the exchange rate is driven by the interest rate differential, risk premium and other unobserved factors. Holding constant $\theta_t$ and $e_t$, there will be one-to-one correspondence between $\Delta s_{t+1}$ and $ID_t$.

2.2 The HP Filter

Following Ball and Mankiw (2002), we apply the HP filter of Hodrick and Prescott (1997) to the observable composite forecast error after rearranging (8)

$$u_t \equiv \Delta s_{t+1} - ID_t$$  \hspace{1cm} (10)

\textsuperscript{10}For example, Lewis (1989) considers the rational exchange rate forecast error; Pippenger (2011) suggests two new omitted variables.

and extract the slow-evolving smooth component, which serves as the estimate for the unobservable risk premium. In other words, the HP filter enables us to isolate the risk premium as the signal, while the idiosyncratic error is left as noise. More explicitly, the HP filter aims to solve this optimization problem:

\[
\min_{\hat{\theta}^\text{HP}_t} \sum_{t=1}^n \left( u_t - \hat{\theta}^\text{HP}_t \right)^2 + \lambda \sum_{t=2}^{n-1} \left[ \left( \hat{\theta}^\text{HP}_{t+1} - \hat{\theta}^\text{HP}_t \right) - \left( \hat{\theta}^\text{HP}_t - \hat{\theta}^\text{HP}_{t-1} \right) \right]^2
\]

(11)

where \( \hat{\theta}^\text{HP}_t \) is the HP estimate of the risk premium (trend component). Note that the HP filter generates a smooth trend with gradual changes in the slope. A tuning parameter \( \lambda \), which is strictly positive, is employed to ensure the smoothness of the trend by minimizing the sum of squared errors between the actual series and the trend, with greater smoothness of the trend component series being achieved with higher values of \( \lambda \). See Hodrick and Prescott (1997) for more details.

In order to justify the usage of HP filter, we can state the identification assumptions as

**Assumption 1** (A) Risk premium \( \theta_t \) is the slow-evolving trend component of the composite forecast error \( u_t \). (B) Idiosyncratic error \( e_t \) is uncorrelated with the interest differential \( \text{cov}(\text{ID}_t, e_t) = 0 \)

The smoothness assumption (A) is innocuous—it is based on the observation that most people adjust their perception of risk in a gradual manner. Even though there can be individual heterogeneity, the central limit theorem indicates that the average or combined risk premium should be smooth. Empirical evidence for the smoothness can be found in previous studies such as Giovanini and Jorion (1989) and the result section of this paper. Assumption (B) is needed for interpretational purpose, under which

\[
\text{cov}(\text{ID}_t, u_t) = \text{cov}(\text{ID}_t, \theta_t)
\]

(12)
Basically Assumption (B) implies that the only indirect channel through which the interest differential affects the exchange rate is via the risk premium.

**2.3 The Fourier Flexible Form**

Hamilton (2018) voices the concern that the HP filter might generate misleading artifact regarding the true data-generating process. In order to guard against the possibly spurious signal from the HP filter, this section adopts the Fourier Flexible Form of Gallant (1981), which is able to provide satisfactory approximation to any bounded function. More explicitly, consider replacing the risk premium \( \theta_t \) with its Fourier approximation \( \text{FFF}_t \) in the forecast error decomposition (9) and get

\[
\begin{align*}
  u_t &= \text{FFF}_t + \hat{\epsilon}_t \\
  \text{FFF}_t &= \mu + \sum_{k=1}^{p} \alpha_k \cos\left(2\pi kt/n\right) + \sum_{k=1}^{p} \beta_k \sin\left(2\pi kt/n\right)
\end{align*}
\]  

where \( k \) is the index for frequency, \( p \) is the number of frequencies, and \( n \) is the sample size. When \( p = 0 \), the risk premium is assumed to be a constant. With rising \( p \) we can allow for increasingly complex patterns in the risk premium. When \( p = n/2 \), the Fourier approximation is perfect for any absolutely integrable function, see Gallant (1981) for details.

Given the parsimony principle, this paper uses \( p = 4 \), and the resulting Fourier approximation is able to mimic the trend component of the HP filter to a large extent.

The unknown coefficients of trigonometric terms \( \alpha_k \) and \( \beta_k \) can be easily estimated after regressing \( u_t \) onto the sine and cosine regressors. Then the estimated risk premium based on the FFF is nothing but the fitted value:

\[
\hat{\theta}^{\text{FFF}}_t = \hat{\mu} + \sum_{k=1}^{p} \hat{\alpha}_k \cos\left(2\pi kt/n\right) + \sum_{k=1}^{p} \hat{\beta}_k \sin\left(2\pi kt/n\right)
\]

Becker et al. (2004), Enders and Lee (2012) and Enders and Li (2015) provide evidence that
for a variety of function forms the FFF approximation performs reasonably well even with just one or two frequencies.

Notice that the FFF method is really a “data-driven” or “endogenous” way of estimating the unknown risk premium. No assumption is made about the function form or the generating process of the risk premium. As long as the risk premium is bounded, the Fourier series is able to yield a global approximation.

It is noteworthy that the FFF approximation provides adequate approximation for not just smooth functions, but also functions with jumps or kinks. See Panels A, B, C, D of Figure 1 in Enders and Li (2015) for an illustration. Basically, those authors show that the FFF works well even in the absence of smoothness in the target function. Consequently we have more faith in the trend of HP filter if it shows similarity to the Fourier approximation. Fortunately, it is the case here.

Finally, readers are cautioned that the extracted trend component from the HP filter or the Fourier series is actually a “catch-all” variable that may include not just the risk premium, but a combination of all omitted variables that evolve slowly over time. Unless further assumptions are made regarding the structure of forecast error $u_t$, there is no way to disentangle the true risk premium from other smooth confounders. So we might conservatively interpret our estimates as a proxy for the risk premium\footnote{Alternatively, we might redefine the risk premium as a term representing all slow-evolving factors.}. Note that the term risk premium in Fama (1984) implicitly is a “catch-all” variable as well.

3 Results

3.1 Descriptive Statistics

Monthly observations for interest rates and exchange rates from January 1971 to June 2016 are downloaded at FRED Economic Data\footnote{The website is https://fred.stlouisfed.org/}. This sample is chosen to maximize the data
availability for all series. The nominal interest rates are the treasury bills rates (percent per annum) for UK, Canada and US. The spot exchange rates are US dollars per one British Pound, and US dollars per one Canadian dollar. None of the series are seasonally adjusted. Descriptive statistics of sample size, sample mean, standard deviation, maximum and minimum are reported in Table 1. On average the US interest rate is the lowest. Moreover, one British Pound on average is worth more than one US dollar, while one Canadian dollar is worth less than one US dollar.

3.2 Fama Beta Coefficient

To duplicate the stylized facts reported in the literature\(^\text{14}\), Figure 1 plots the \(\hat{\beta}_1\) in the Fama regression (2) estimated by OLS using rolling windows—each window contains 60 (five years) observations, and every time we move the window forward by one month. From Figure 1 it is clear that for both UK Pound and Canadian Dollar the UIP hypothesis is violated most of the time since majority of \(\hat{\beta}_1\) are less than unity. In fact we can see a similar pattern in \(\hat{\beta}_1\) for both currencies—before 1990, \(\hat{\beta}_1\) was mostly negative; it became positive very briefly in early 1990s; then it turned negative until 2005.

Note that \(\hat{\beta}_1\) has become very erratic since 2007. This high volatility in the OLS estimate is largely a reflection of the fact that all three countries experienced zero lower bound in interest rates during the Great Recession, which substantially reduces the variation in the interest rate differential and therefore the preciseness of the OLS estimate\(^\text{15}\). Hence it is sensible to focus on the pre-2007 subsample.

\(^{14}\text{For instance, our estimated } \hat{\beta}_1 \text{ is comparable to those reported in Table 2 of Fama (1984).}\)

\(^{15}\text{The standard error of } \hat{\beta}_1 \text{ is negatively related to the sample variance of the regressor.}\)
3.3 HP Estimate of Risk Premium

Panel A of Figure 2 plots the composite error $u_t$ (10) for UK pound. This series is choppy thanks to the volatility of exchange rates and interest rates. However, there is evidence for clustering— a large (small) composite error tends to be followed by another large (small) composite error. For instance, we see a cluster of large composite errors between 1985 and 1991, and a cluster of small composite errors between 1995 and 2000.

This clustering phenomenon may be indicative of the smoothness in the risk premium, and Panel B plots the estimated risk premium $\hat{\theta}_t^{\text{HP}}$ obtained by solving (11), i.e., it is the trend component of the composite error extracted by the HP filter. In Panel B we see that the HP estimate of risk premium is much smoother relative to the composite error. There are three periods associated with rising risk premium—the early 1970s, the late 1980s, and early 2000s.

Panels C and D in Figure 2 are for Canadian Dollar. We see similar clustering behavior in the composite error, and similar smoothness in the estimated risk premium. Furthermore, there seems to be co-movement or synchronization between the two risk premium series—there is a peak in the year 1991, followed by another peak in the year 2004, in both Panels B and D.

To check the robustness of the estimated risk premium to the tuning parameter of the HP filter, Figure 3 compares the HP estimates of risk premium using the default value of tuning parameter for monthly data $\lambda = 144000$ and the value for quarterly data $\lambda = 1600$. Obviously, such a big change in $\lambda$ has little effect on the risk premium since the extracted trends mostly overlap as shown by Figure 3. As expected a greater $\lambda$ leads to a smoother trend. Following the suggestion of Ravn and Uhlig (2002) we also try the improved tuning value for monthly data $\lambda = 129600$, and no qualitative difference is found.
3.4 FFF Estimate of Risk Premium

To address the concern of Hamilton (2018) that the HP filter may generate spurious artifact, Figure 4 displays the Fourier Flexible Form estimate of the risk premium $\hat{\rho}_t^{\text{FFF}}$ given by (15) using four frequencies $p = 4$. To ease the comparison, Figure 4 also shows the HP estimate of risk premium $\hat{\rho}_t^{\text{HP}}$. The key message from Figure 4 is that the FFF estimate of risk premium is largely consistent with the HP estimate: the magnitudes of the peak and trough are similar, and most importantly, the locations of turning points match very well. That resemblance provides support for the smoothness assumption made for the HP filter.

This tight co-movement in the two estimates is crucial because it implies that the misleading phase shift is unlikely to be produced by the HP filter in our case. Thus, the primary concern of Hamilton (2018) is resolved. We also try five frequencies when obtaining the FFF estimate but the improvement is marginal.

3.5 Moving-Average Estimate of Risk Premium

More robustness check is given by Figure 5, in which the HP estimate of risk premium is compared to an ad hoc 12-point centered flat moving average (MA) filter applied to the composite error $u_t$. Keep in mind that both HP and MA filters aim to isolate the smooth trend or signal.

Figure 5 shows that overall the HP estimate of risk premium moves closely with the MA filter—they reach the peaks and troughs at roughly the same time. This synchronization can be seen as another support for the HP estimate. Nevertheless, compared to Figure 4, the gap between the HP and MA estimates is more noticeable than the gap between the HP and FFF estimates, in large part because the MA filter is choppier than the HP filter. This finding is expected given that the HP filter puts more emphasis on the smoothness of trend component than the MA filter.
3.6 ARCH-in-Mean Estimate of Risk Premium

Figure 6 compares the standardized HP estimate of risk premium to that of Domowitz and Hakkio (1985) where the authors assume a proxy of risk premium is the conditional standard deviation of forecasting error obtained from the ARCH-in-Mean model of Engle et al. (1987). More explicitly, consider augmenting the Fama regression with the conditional heteroskedasticity equation

\[
\Delta s_{t+1} = \beta_0 + \beta_1 ID_t + \gamma \sigma_t + \epsilon_t \\
\sigma_t^2 = c_0 + c_1 \epsilon_{t-1}^2 + c_2 \sigma_{t-1}^2
\]

(16) (17)

Basically, in regression (16) the conditional standard deviation \( \sigma_t \) serves as a proxy for the risk premium\(^\text{16}\), and (17) is the standard GARCH(1,1) model. Maximum Likelihood method is used to estimate unknown parameters including \( \gamma \), which measures the partial effect of risk premium on the change of exchange rates.

The ARCH-in-Mean model is applied to our data with partial success —the absolute t value of \( \hat{\gamma} \) is less than 1.96 for both UK Pound and Canadian Dollar. Taken literally, that means controlling for risk premium (or its proxy) does not improve the in-sample fitting. Because of the inconsequential \( \hat{\gamma} \), the Fama estimate \( \hat{\beta}_1 \) is still significantly less than the UIP value \( \beta = 1 \) for both countries. Despite that failure in the first moment regression, the estimated ARCH coefficient \( \hat{c}_1 \) and GARCH coefficient \( \hat{c}_2 \) in the second moment regression are both significant at the 5% level. Therefore it is still meaningful to use Figure 6 to contrast the standardized HP estimate of risk premium with the standardized conditional standard deviation \( \sigma_t \).

Volatility clustering, or the tendency that a big \( \sigma_t \) is followed by another big one, is evident in Figure 6. More importantly, it is remarkable that there is some agreement between the

\(^{16}\)There is no qualitative change in results if \( \log(\sigma_t) \) is used on the right hand side of (16).
two estimates. Take UK Pound. We see both rising HP risk premium and high volatility between 1985 and 1992. On the other hand, we see coexistence of low risk premium and low volatility between 1995 and 2005. After 2008 it seems that one estimate starts deviating from the other. But this divergence is an artifact and should be downplayed since the key regressor \( ID_t \) is close to zero during the zero-lower-bound period.

There are at least two ways to explain why the HP risk premium differs from the ARCH-in-Mean estimate. First, the former aims to achieve smoothness while the latter does not. Second, Engle et al. (1987) assume normality of data and constant absolute risk aversion (CARA) utility function to justify only conditional variance matters for the risk premium, see their equation (3). We think the ARCH-in-Mean model is prone to misspecification since either the normality or CRRA can fail in practice\(^{17}\). As a quick check, we run a generalized “Kurtosis-in-Mean” model by including the fourth moment of \( \hat{\varepsilon}_t \) on the right hand side of (16), and the t-value of the kurtosis term is significant for Canada. So at least for Canadian Dollar, the likelihood of extreme values matters and the ARCH-in-Mean model suffers the omitted variable bias.

In short, given the findings from Figure 6, our HP and FFF estimates are able to shed new light in the risk premium literature.

### 3.7 Verifying Proposition 1

After the risk premium is estimated by the HP filter and Fourier Series Approximation, it is straightforward to evaluate (P1) and (P2) in Proposition 1—we only need to compare the covariance between the estimated risk premium and interest differential to zero, and the negative variance of interest differential. Equivalently, we can run the following two auxiliary

\(^{17}\)For instance, Fama (1965) rejects the normality and supports heavier-tailed distributions.
regressions using the HP and FFF estimates of risk premium as the dependent variables:

\[
\hat{\theta}_t^{HP} = \phi_{10} + \phi_{11}ID_t + \eta_{1t} \quad (18)
\]
\[
\hat{\theta}_t^{FFF} = \phi_{20} + \phi_{21}ID_t + \eta_{2t} \quad (19)
\]

where \(\phi_{11} = \frac{\text{cov}(\hat{\theta}_t^{HP}, ID_t)}{\text{var}(ID_t)}\) and \(\phi_{21} = \frac{\text{cov}(\hat{\theta}_t^{FFF}, ID_t)}{\text{var}(ID_t)}\). It follows that these two slope coefficients are negative if the covariances in the numerators are negative; while they are less than \(-1\) if the covariances are less than the negative variances.

Figure 7 plots \(\hat{\phi}_{11}\) (marked by square) and \(\hat{\phi}_{21}\) (marked by diamond) from regressions (18) and (19) with the same rolling windows used for Figure 1. We find that majority of \(\hat{\phi}_{11}\) and \(\hat{\phi}_{21}\) are negative, supporting (P1). Nevertheless, the evidence for (P2) is elusive. For instance, Panel A in Figure 1 shows that the Fama beta is mostly negative between 1975 and 1990. If (P2) were true, we would see that \(\hat{\phi}_{11}\) and \(\hat{\phi}_{21}\) are less than \(-1\) during that period. Instead, none of the \(\hat{\phi}_{11}\) and \(\hat{\phi}_{21}\) in Panel A of Figure 7 are less than \(-1\) between 1975 and 1990. In fact, the only convincing finding that favors (P2) is in 1998, when the Fama beta for Canadian Dollar is negative and the \(\hat{\phi}_{11}\) is less than \(-1\) at the same time.

Regressions (18) and (19) can be improved by replacing the interest differential \(ID_t\) with its smooth component such as the HP trend. That is, it makes more sense to do an apple-to-apple comparison by removing high-frequency fluctuation from both the composite forecast error and the interest differential. Panels C and D of Figure 7 plot new \(\hat{\phi}_{11}\) and \(\hat{\phi}_{21}\) when the HP trend component of interest differential is used as the regressor in (18) and (19). Now we see more supporting evidence for (P2) such as \(\hat{\phi}_{11}\) being less than \(-1\) in particular.

4 Conclusion

It is important to understand the risk premium, which plays a critical role in a variety of topics in finance such as asset pricing and term structure of interest rates. This paper focuses
on one theory that explains the uncovered interest parity puzzle—the prediction for future exchange rate based on the interest differential is biased due to the unobserved risk premium.

The main contribution of this study is advocating two nonparametric approaches of estimating the risk premium. The first estimate from the Hodrick-Prescott filter is built upon the smoothness assumption motivated by consumption smoothing, sticky prices and central limit theorem in the intertemporal asset pricing framework. The second estimate is more flexible, free of typical assumptions about the structural form, and entails the Fourier series approximation to underlying risk structure. The two approaches generate similar estimates of the risk premium, which contrast with the one produced by the ARCH-in-Mean model. We argue that the conditional variance of forecast error may be an insufficient proxy for risk premium, and therefore the proposed nonparametric estimates can serve as an important complement.

Finally, we use the estimated risk premium to directly test the theory of Fama (1984) that the risk premium can resolve the uncovered interest parity puzzle. The covariance between the interest differential and estimated risk premium is found to be negative, which explains why the Fama coefficient is usually less than unity. However, there is mixed finding about the covariance being less than negative variance of interest differential, which is needed to explain why the Fama coefficient is sometimes negative.
References


<table>
<thead>
<tr>
<th>Series</th>
<th>( n )</th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>( i_{UK} )</td>
<td>546</td>
<td>6.904</td>
<td>4.229</td>
<td>0.224</td>
<td>16.181</td>
</tr>
<tr>
<td>( i_{CA} )</td>
<td>546</td>
<td>6.044</td>
<td>4.198</td>
<td>0.170</td>
<td>20.820</td>
</tr>
<tr>
<td>( i_{US} )</td>
<td>546</td>
<td>4.876</td>
<td>3.439</td>
<td>0.010</td>
<td>16.290</td>
</tr>
<tr>
<td>( S_{UK} )</td>
<td>546</td>
<td>1.757</td>
<td>0.308</td>
<td>1.093</td>
<td>2.618</td>
</tr>
<tr>
<td>( S_{CA} )</td>
<td>546</td>
<td>0.839</td>
<td>0.115</td>
<td>0.625</td>
<td>1.047</td>
</tr>
</tbody>
</table>

Note: \( i \) is the treasury bill rate, \( S \) is the spot exchange rate; UK, CA, US stand for United Kingdom, Canada, and United States, respectively.
Figure 1: Time-Varying Beta1

Panel A: UK

Panel B: Canada
Figure 2: Composite Forecast Error and HP Risk Premium

Panel A: UK, u

Panel B: UK, thetha

Panel C: Canada, u

Panel D: Canada, thetha
Figure 3: HP Risk Premium---Tuning Parameter

Panel A: UK
Lambda=144000
Lambda=1600

Panel B: Canada
Lambda=144000
Lambda=1600
Figure 4: FFF and HP Risk Premia

Panel A: UK

Panel B: Canada
Figure 5: Risk Premium---Moving Average Filter

Panel A: UK

Panel B: Canada
Figure 6: HP Risk Premium and Conditional Standard Deviation

Panel A: UK

Panel B: Canada
Figure 7: Verifying Proposition 1

Panel A: UK
HP Filter Flexible Fourier Form

Panel B: Canada

Panel C: UK

Panel D: Canada