Online appendix to accompany "Exchange rates, uncovered interest parity and time-varying Fama regressions"

Bowen Fu[a] *Mengheng Li*[b,d*] *Qazi Haque*[c,d]

[a] *Center for Economics, Finance and Management Studies, Hunan University, China*
[b] *Economics Discipline Group, University of Technology Sydney, Australia*
[c] *School of Economics and Public Policy, The University of Adelaide, Australia*
[d] *Centre for Applied Macroeconomic Analysis, Australia*

This online appendix provides supplementary materials for the main text "Exchange rates, uncovered interest parity and time-varying Fama regressions". Notations are defined in the main text, unless defined otherwise. Section 1 replicates the conflicting results in Ismailov and Rossi (2018) and Engel et al. (2022). Section 2 shows more details of model-implied observation weights. Section 3 gives more empirical results. Section 4 provides additional details on the Bayesian sampling scheme. Section 5 gives estimation results from a SV-in-mean model.

Keywords: *Uncovered interest rate parity, forward premium puzzle, excess volatility puzzle, Bayesian time-varying regression, and stochastic volatility*
JEL Classification: C11, C32, F31, F37

*Corresponding author: M. Li, University of Technology Sydney, UTS Business School, Ultimo NSW 2007, Australia. mengheng.li@uts.edu.au

A. Rolling-window regressions and choice of the window size

For the Fama regression estimated using RW, a window size must be chosen. Using similar dataset, RW-Fama regression coefficients from Ismailov and Rossi (2018) and Engel et al. (2022) show different patterns and thus lead to different conclusions. Figure 1 and 2 are taken from the two studies directly. Of interest is the UIP relation between US dollar and British pounds, the slope coefficient in particular. The scales of the y-axes are different due to different variable scaling used in the two studies. As Ismailov and Rossi (2018) use a shorter time period, the corresponding sample period for the results in Engel et al. (2022) is indicated by a black box.

Both studies suggest an increasing trend in β_t prior to 2009. Major difference happens after the GFC. Ismailov and Rossi (2018) find that the increase continues, whereas Engel et al. (2022) document a nosedive. Consequently, the former would suggest a persistent violation of UIP, but the latter apparently points to the very opposite.

Two possibilities are behind this conflicting result: (1) RW regressions are not robust to different forms of parameter instability; (2) RW regressions are not robust to different choices of window size. To address the first possibility, we use the SCL method of Giraitis et al. (2014) and Petrova (2019). This method has been theoretically and empirically shown to be robust to various forms of parameter evolution: structural breaks, gradual changes, and jumps. But, it involves a bandwidth tuning parameter that essentially serves as the same role of window size.

Using our dataset, which is the same as Ismailov and Rossi, 2018, we estimate the Fama regression using the SCL method but under two different bandwidths. The left and the right panels of Figure 3 largely replicate the results from Ismailov and Rossi (2018) and Engel et al. (2022), respectively.

It is unclear how one can make a consistent conclusion about the UIP relation around 2010 according to the two estimates. As the SCL approach rules out the aforementioned first possibility, the conflicting results are mostly attributable to the ad-hoc choice of the window size (or bandwidth in SCL). An estimation scheme such as the TVP-SV-Fama regression proposed in the main text that does not depend on any tuning parameter seems more suitable for inves-

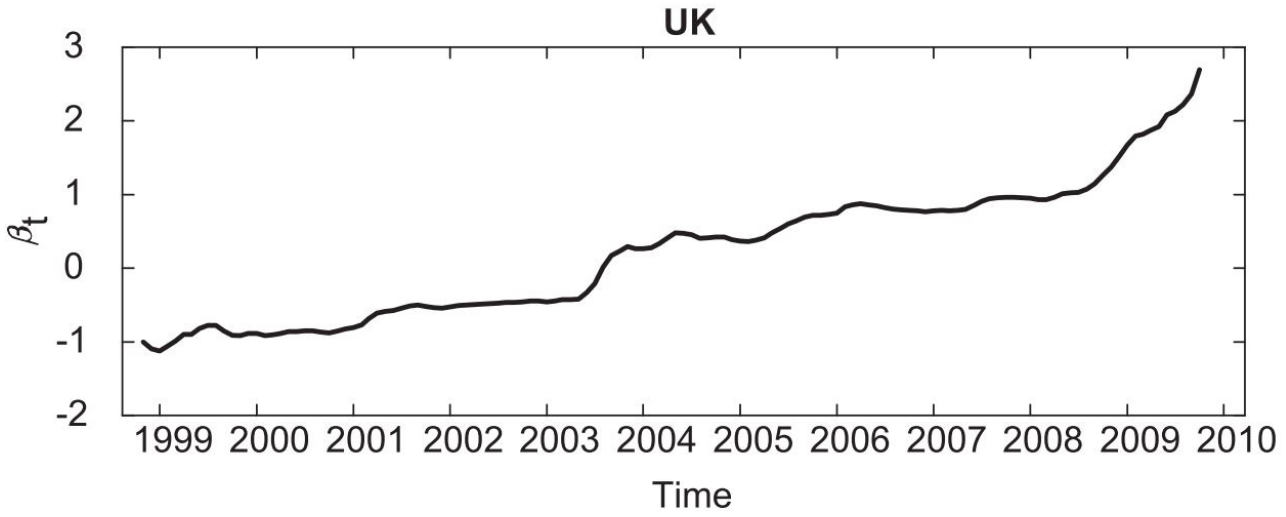

Figure 1: Slope coefficient of the Fama regression estimated by Ismailov and Rossi (2018). The Fama regression $s_{t+1} - s_t = \alpha + \tilde{\beta}(i_t - i_t^*) + \epsilon_t$ is estimated by RW.

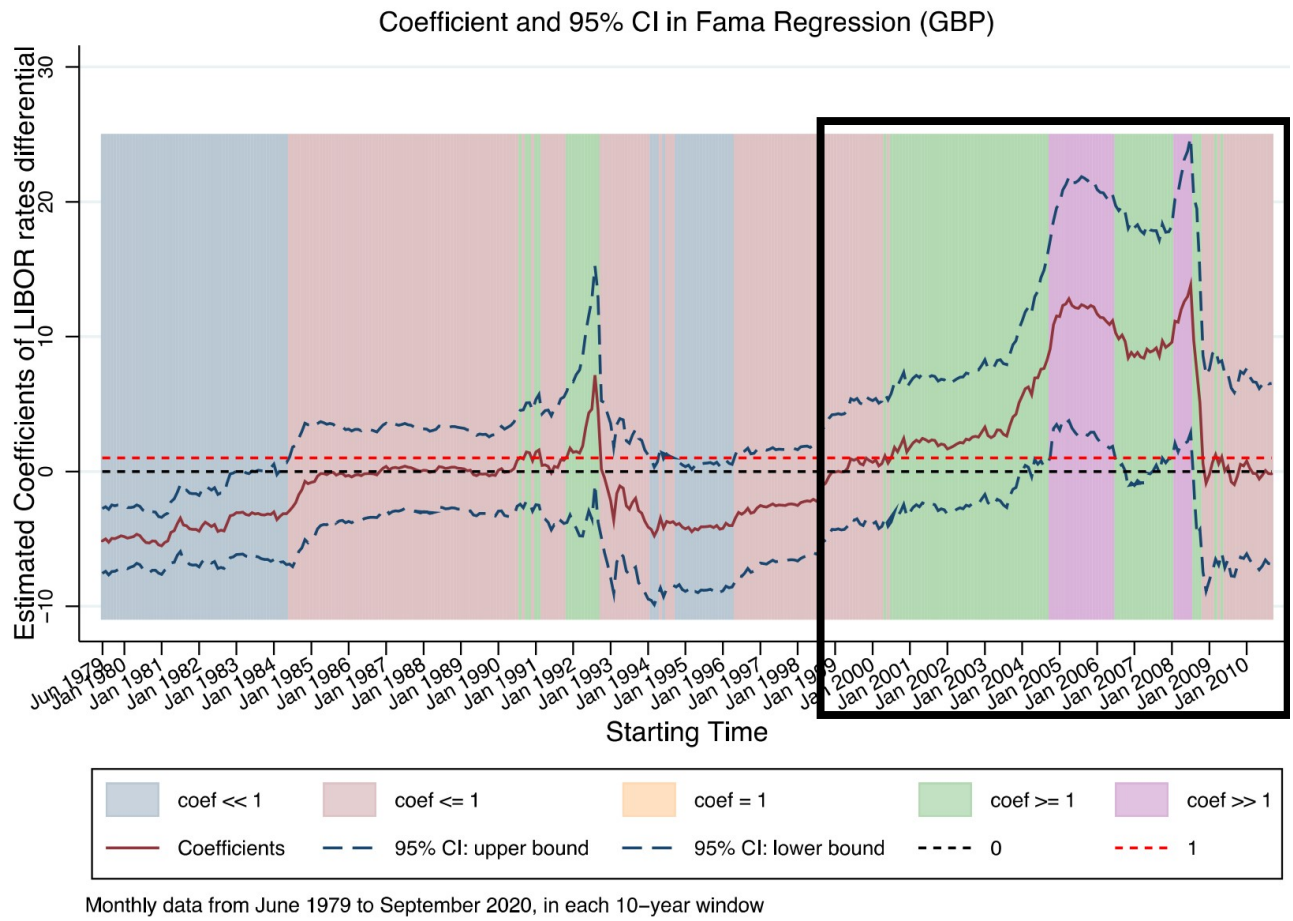

Figure 2: Slope coefficient of the Fama regression estimated by Engel et al. (2022). The Fama regression $s_{t+1} - s_t = \alpha + \tilde{\beta}(i_t - i_t^*) + \epsilon_t$ (Equation (2) in the main text) is estimated by RW. Black box indicates the sample period between 1999 and 2010.

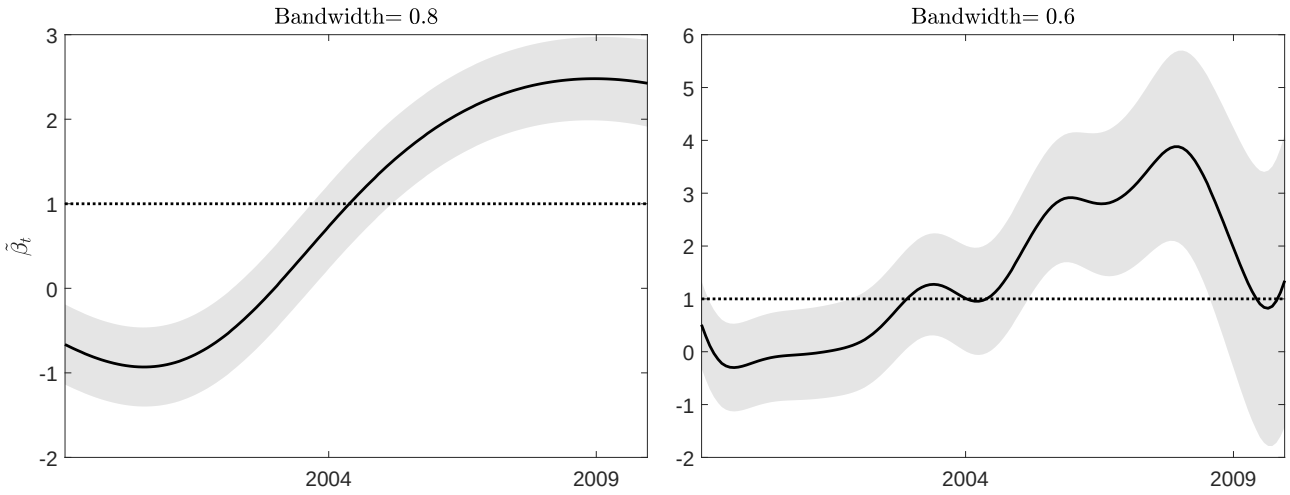

Figure 3: Slope coefficient of the Fama regression estimated by the SCL method. The left and the right panel shows estimates using two different bandwidths.

tigating time-variation in UIP coefficients.

B. Derivation of observation weights

Conditional on $\boldsymbol{\sigma}$, *e.g.* an MCMC draw or a posterior estimate, Kalman filter and smoother can be used to compute the observation weights of the TVP-SV-Fama regression model (**??**), with respect to the UIP coefficient β_t. For simplicity, we assume $\alpha_t = 0$ for all t, which is in line with our empirical results. The model becomes

$$\frac{\rho_{t+1}}{i_t^* - i_t} = \beta_t + \frac{\sigma_t}{i_t^* - i_t}\epsilon_t,$$
$$\beta_{t+1} = \beta_t + \sigma_\beta \eta_{\beta,t}.$$

Suppose the system were time-invariant with error variance σ_ϵ^2. Kalman filter outputs reach their steady state quickly (Durbin and Koopman, 2012, Chapter 4). Let a_t and P_t denote the filtering expectation $E(\beta_t|\{\rho_{s+1}, i_s^* - i_s\}_{s=1}^{t-1})$ and variance $\mathrm{Var}(\beta_t|\{\rho_{s+1}, i_s^* - i_s\}_{s=1}^{t-1})$, respectively. The Kalman filter iterates forward over

$$a_{t+1} = a_t + \frac{P_t}{P_t + \sigma_\epsilon^2}\left(\frac{\rho_{t+1}}{i_t^* - i_t} - a_t\right), \quad P_{t+1} = \frac{P_t \sigma_\epsilon^2}{P_t + \sigma_\epsilon^2} + \sigma_\beta^2.$$

The steady state is given by the fixed-point solution to the second equation, and it is $\bar{P} = (q + \sqrt{q^2 + 4q})/2$ where $q = \sigma_\beta^2/\sigma_\epsilon^2$ is the signal-to-noise ratio. Let b_t denote the smoothed expectation $E(\beta_t|\{\rho_{s+1}, i_s^* - i_s\}_{s=1}^T)$. The Kalman smoother iterates backward from $r_T = 0$ and $\text{Var}(r_t) = N_t$ over

$$r_{t-1} = \frac{\rho_{t+1}/(i_t^* - i_t) - a_t}{P_t + \sigma_\epsilon^2} + \frac{\sigma_\epsilon^2}{P_t + \sigma_\epsilon^2} r_t, \quad b_t = a_t + P_t r_{t-1}, \quad N_{t-1} = \frac{1}{P_t + \sigma_\epsilon^2} + \left(\frac{\sigma_\epsilon^2}{P_t + \sigma_\epsilon^2}\right)^2 N_t.$$

Because $\frac{\sigma_\epsilon^2}{P_t + \sigma_\epsilon^2} < 1$, the steady state of N_t exists and equals to $\bar{N} = (\bar{P} + \sigma_\epsilon^2)/(\bar{P}^2 + 2\bar{P}\sigma_\epsilon^2)$.

Suppose we can write $b_t = \sum_{j=1}^T \omega_{jt}[\rho_{j+1}/(i_j^* - i_j)]$ with weight ω_{jt} associated with the j-th observation corresponding to t-th smoothed estimate. Then we have $E(b_t\epsilon_j) = \omega_{jt}E([\rho_{j+1}/(i_j^* - i_j)]\epsilon_j) = \omega_{jt}$, but also

$$E(b_t\epsilon_j) = \begin{cases} -\text{Cov}(\epsilon_j - E(\epsilon_j|\{\rho_{s+1}, i_s^* - i_s\}_{s=1}^T), \beta_t - b_t), & \text{for } j < t; \\ -\text{Cov}(\beta_t - b_t, \epsilon_j - E(\epsilon_j|\{\rho_{s+1}, i_s^* - i_s\}_{s=1}^T)), & \text{for } j \geq t. \end{cases}$$

After minor algebraic manipulation, the steady state gives for $j < t$

$$\text{Cov}(\epsilon_j - E(\epsilon_j|\{\rho_{s+1}, i_s^* - i_s\}_{s=1}^T), \beta_t - b_t) = E(\epsilon_j(\beta_t - b_t)) = -\frac{\bar{P}}{\bar{P} + \sigma_\epsilon^2}\left(\frac{\sigma_\epsilon^2}{\bar{P} + \sigma_\epsilon^2}\right)^{t-j} \frac{\bar{P}\sigma_\epsilon^2}{\bar{P}^2 + 2\bar{P}\sigma_\epsilon^2}.$$

Similarly, for $j \geq t$ we have

$$\text{Cov}(\beta_t - b_t, \epsilon_j - E(\epsilon_j|\{\rho_{s+1}, i_s^* - i_s\}_{s=1}^T)) = -\frac{\bar{P}}{\bar{P} + \sigma_\epsilon^2}\left(\frac{\sigma_\epsilon^2}{\bar{P} + \sigma_\epsilon^2}\right)^{j-t} \frac{\bar{P}\sigma_\epsilon^2}{\bar{P}^2 + 2\bar{P}\sigma_\epsilon^2}.$$

Therefore, as j moves away from t, observation $\rho_{j+1}/(i_j^* - i_j)$ receives exponentially declining weight proportional to

$$\omega_{jt} \propto \left(\frac{\sigma_\epsilon^2}{\bar{P} + \sigma_\epsilon^2}\right)^{|t-j|} = \left(\frac{2\sigma_\epsilon^4}{\sigma_\beta^2 + \sqrt{\sigma_\beta^4 + 4\sigma_\beta^2\sigma_\epsilon^2} + 2\sigma_\epsilon^4}\right)^{|t-j|}.$$

When SV is present, the first-order approximation of the weighting function can be computed

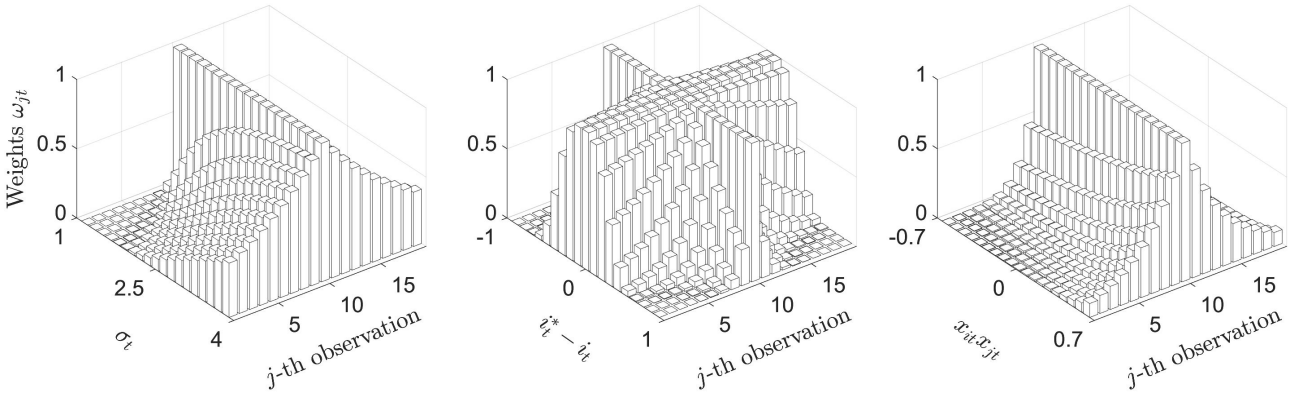

Figure 4: Weighting functions implied by the TVP-SVX-Fama regression model. The left, middle and right figure show the effect of volatility σ_t, interest rate differential $i_t^* - i_t$, and cross-sectional correlation in x_t on ω_{jt} for $j = 1, ..., 19$ and $t = 10$.

by replacing σ_ϵ^2 with $\sigma_t^2/(i_t^* - i_t)^2$.[1] So if σ_t increases or the interest rate differential $i_t^* - i_t$ decreases, ω_{jt} becomes larger. This means that during volatile times, the estimation of the slope β_t relies on more backward and forward information, and vice versa.

In section 4.3 of the main text, we conduct a robustness check where the state dynamics is augmented with regression effects from US economic shocks. In this case, σ_β^2 in the above observation weight is replaced by $\sigma_\beta^2 + \gamma_\beta' x_t x_t' \gamma_\beta$, meaning that the signal-to-noise ratio, and thus the weight, is also affected by the variation of shocks. In Figure 4, we plot the weighting function for estimating β_{10} with $T = 19$ under increasing volatility, changing interest rate differential and changing cross-sectional dependence in x_t.

In sum, the data-driven feature of the model considers more observations when estimating β_t if the signal-to-noise ratio is low. In particular, remote observations (*i.e.* j is away from t) becomes more important if period t is associated with (1) high uncertainty or volatility; (2) small interest rate gap; (3) low variation in non-US shocks; and (4) low variation in US shocks. This is intuitive because low signal-to-noise ratio leads to estimation uncertainty, which should be mitigated by using more observations.

[1] Derivations of filtering weights for a simpler local level model with SV can be found in Shephard (2015).

C. Additional empirical results

C.1. Break date determination via marginal data likelihood

In section 4.1, we estimate the SV-Fama model with a break in the slope coefficient. To determine the break data, we consider a sequence of models, with each defined by a break data. Suppose the i-th model indicates the model with a structural break in slope occuring at time $10+i$, $i=1,...,T-21$. We then compute the marginal data likelihood for all $T-21$ models via

$$p(\boldsymbol{\rho}|\boldsymbol{X}_i) = \int p(\boldsymbol{\rho}|\boldsymbol{X}_i,\boldsymbol{\theta})p_0(\boldsymbol{\theta})d\boldsymbol{\theta} \approx \frac{1}{J}\sum_{j=1}^{J} p(\boldsymbol{\rho}|\boldsymbol{X}_i,\boldsymbol{\theta}^{(j)}),$$

where $\boldsymbol{\theta}^{(j)}$, $j=1,...,J$, is a draw from $p_0(\boldsymbol{\theta})$, the prior distribution of the vector of static parameters $\boldsymbol{\theta}$. This means the middle term in the above is the integrated likelihood, and the right term is an unbiased estimate of it. $\boldsymbol{X}_i$ is a $T\times 3$ design matrix with the t-th row $(1, i_t^* - i_t, 0)$ if $t<i$ and $(1, i_t^*-i_t, i_t^*-i_t)$ if $t\geq i$. We can partition $\boldsymbol{\theta}=(\boldsymbol{\beta}', v_h)'$, where $\boldsymbol{\beta}$ is the 3×1 vector of regression coefficients (*i.e.* the pre-break slope is β_2 and the post-break slope is $\beta_2+\beta_3$), and v_h is the innovation variance of the random walk log variance $h_t=\log\sigma_t^2$ in the SV-Fama regression.

Drawing from the prior is trivial. We focus on the evaluation of $p(\boldsymbol{\rho}|\boldsymbol{X}_i,\boldsymbol{\theta}^{(j)})$, which is an integrated conditional likelihood:

$$p(\boldsymbol{\rho}|\boldsymbol{X}_i,\boldsymbol{\theta}^{(j)}) = \int p(\boldsymbol{\rho}|\boldsymbol{X}_i,\boldsymbol{h},\boldsymbol{\theta}^{(j)})p(\boldsymbol{h})d\boldsymbol{h}, \tag{1}$$

where $\boldsymbol{h}=(h_1,...,h_T)'$. In the above, the right term involves high-dimensional integration, and thus cannot be evaluated by Monte Carlo integration feasibly. One can use a particle filter to integrate out the SV sequentially (Chib et al., 2002). Considering the sheer number of model evaluations, we adopt the numerically accelerated importance sampling (NAIS) method developed by Koopman et al. (2015) and Li and Koopman (2021). NAIS is specifically designed for SV models and performs way faster than popular sequential Monte Carlo methods.

For each j, $\boldsymbol{X}_i$ and $\boldsymbol{\theta}^{(j)}$ are fixed, so we suppress the dependence and denote $\boldsymbol{\epsilon} = \boldsymbol{\rho} - \boldsymbol{X}_i\boldsymbol{\beta}$. It follows from (1) that $p(\boldsymbol{\rho}) = p(\boldsymbol{\epsilon})$ and

$$p(\boldsymbol{\epsilon}) = g(\boldsymbol{\epsilon}) \int \frac{p(\boldsymbol{\epsilon}|\boldsymbol{h})p(\boldsymbol{h})}{g(\boldsymbol{\epsilon}|\boldsymbol{h})g(\boldsymbol{h})} g(\boldsymbol{h}|\boldsymbol{\epsilon}) d\boldsymbol{h} \approx g(\boldsymbol{\epsilon}) \frac{1}{K} \sum_{k=1}^{K} \boldsymbol{\omega}^{(k)}, \tag{2}$$

where $g(.)$ is the density implied by an importance model which enables evaluation of $g(\boldsymbol{\epsilon})$ and drawing $\boldsymbol{h}^{(k)}$, $k = 1, ..., K$, from $g(\boldsymbol{h}|\boldsymbol{\epsilon})$ to easily construct the unbiased estimate expressed by the importance weights $\boldsymbol{\omega}^{(k)}$. The importance model is chosen to be linear and Gaussian

$$\begin{aligned} \frac{b_t}{c_t} &= h_t + \eta_t, \quad \eta_t \sim N\left(0, \frac{1}{c_t}\right), \\ h_t &= h_{t-1} + \sqrt{v_h}\epsilon_{h,t}, \quad \epsilon_{h,t} \sim N(0,1), \end{aligned} \tag{3}$$

where (b_t, c_t), $t = 1, ..., T$, are importance parameters to be determined and implicit functions of data. In the above, the state transition is identical to the SV-Fama regression model, so $p(\boldsymbol{h}) = g(\boldsymbol{h})$ and the importance weight is given by

$$\boldsymbol{\omega}^{(k)} = \frac{p(\boldsymbol{\epsilon}|\boldsymbol{h}^{(k)})p(\boldsymbol{h}^{(k)})}{g(\boldsymbol{\epsilon}|\boldsymbol{h}^{(k)})g(\boldsymbol{p}^{(k)})} = \prod_{t=1}^{T} \frac{p(\epsilon_t|h_t^{(k)})}{g(\epsilon_t|h_t^{(k)})},$$

where $p(\epsilon_t|h_t^{(k)}) \sim N(0, \exp(h_t^{(k)}))$ given by the SV-Fama model, and it is easy to show that $g(\epsilon_t|h_t^{(k)}) = g(b_t/c_t|h_t^{(k)}) \sim N(h_t^{(k)}, 1/c_t)$ such that

$$\log g(\epsilon_t|h_t) = a_t + b_t h_t - \frac{1}{2} h_t^2 c_t,$$

where the constant $a_t = -\frac{1}{2}(\log 2\pi - \log c_t + b_t^2 c_t)$ does not dependent on h_t. That (b_t, c_t) enters $\log g(\epsilon_t|h_t)$ linearly means that the optimal importance parameters can be found in a least squares problem:

$$(b_t, c_t)^{\text{new}} = \arg\min_{b,c} \sum_{s=1}^{S} \left(\log p(\epsilon_t|h_t^{(s)}) - a - h_t^{(s)} b + \frac{1}{2}(h_t^{(s)})^2 c \right)^2,$$

where $(1, h_t^{(s)}, -(h_t^{(s)})^2/2)'$, is the vector of regressors. With $s = 1, ..., S$, they form a set of "data". One can draw $h_t^{(s)}$ from $g(\boldsymbol{h}|\boldsymbol{\epsilon})$ implied by the importance model (3) under $(b_t, c_t)^{\text{old}}$, $t = 1, ..., T$. But NAIS method makes this process simulation free, by using Gauss-Hermite quadrature. This means we simply set $h_t^{(s)} = E_g(h_t|\boldsymbol{\epsilon}) + \sqrt{\text{Var}_g(h_t|\boldsymbol{\epsilon})} z_s$, where $E_g(h_t|\boldsymbol{\epsilon})$ and $\text{Var}_g(h_t|\boldsymbol{\epsilon})$ are smoothed mean and variance under the importance model (3) which are easily obtained by Kalman smoother, or even more efficiently by the precision sampler of Chan and Jeliazkov (2009), and where z_s is one of the 10 Gauss-Hermite nodes associated with weight q_s. This leads to a weighted least square problem:

$$(b_t, c_t)^{\text{new}} = \arg\min_{b,c} \sum_{s=1}^{10} q_s \left(\log p(\epsilon_t|h_t^{(s)}) - a - h_t^{(s)} b + \frac{1}{2}(h_t^{(s)})^2 c \right)^2,$$

After finding the optimal importance parameters, $g(\boldsymbol{\epsilon})$ can be computed from Kalman filter applied to (3), or even more efficiently from the precision sampler of Chan and Jeliazkov (2009). This gives us all ingredients for computing the integrated likelihood (2). NAIS method requires more coding effort than particle filtering, but is much more efficient and accurate for SV models (Li and Koopman, 2021). Its speed turns out to be instrumental for break date determination which involves otherwise a prohibitively large number of model evaluations. In our exercise, we choose $J = 10,000$ (*i.e.* 10K draws of static parameters are made from its prior) for each possible break dates, totaling 2.82×10^6 marginal likelihood evaluations. For each currency pair, it takes around 8 minutes on a personal laptop without parallelization.

Figure 5 shows the marginal data likelihood for the SV-Fama regression with a break in the slope, at each possible break date and for all currency pairs. The break data is chosen to be the one that maximizes the data likelihood.

C.2. More on mixing properties

As a monotonic transformation of the inefficiency factor used in the main text, Table 1 presents the effective sample size computed by the method given in Roy (2020), divided by 5,000. When computing, we use Bartlett weights. Exponential weights generate similar results.

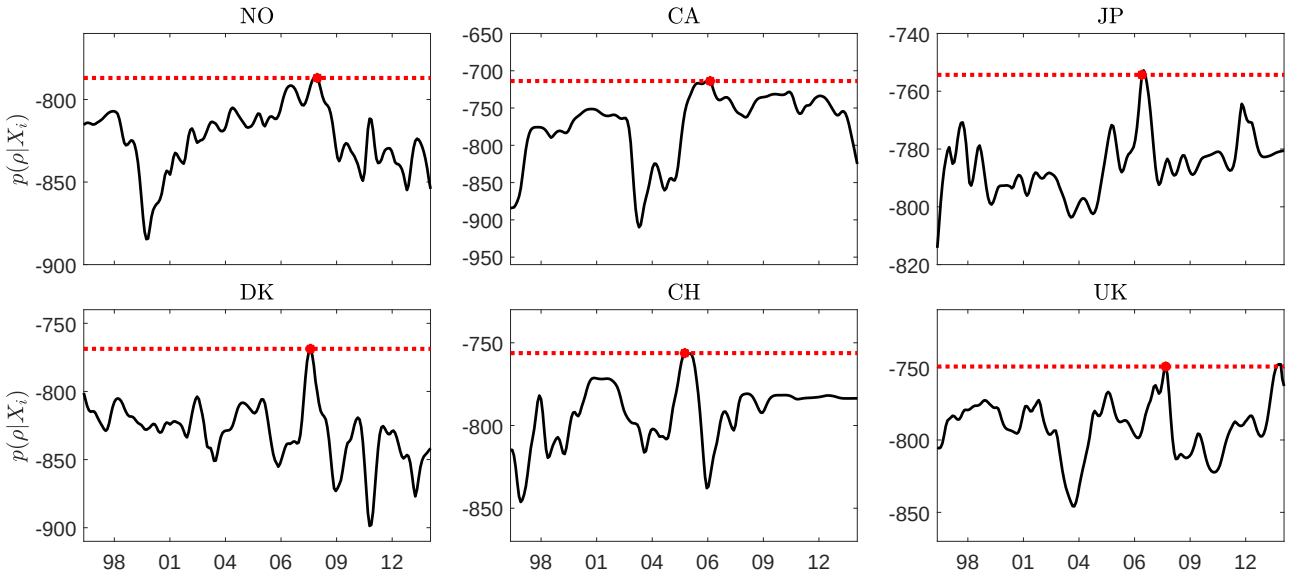

Figure 5: Marginal data likelihood of the SV-Fama regression model. Each possible break date gives a model which is evaluated via the marginal data likelihood. Candidate break dates cover 1995 - 2016. Red indicates the date that corresponds to the highest data likelihood, pinning down the break date.

Table 1: Effective sample size of TVP-SV-Fama model parameters

	NO	CA	JP	DK	CH	UK
v_α	3.26	1.73	2.01	1.04	1.67	1.35
v_β	1.85	1.70	2.22	1.16	1.41	0.76
$v_{\log \sigma^2}$	0.82	1.06	0.81	0.68	0.65	0.63
$\{\alpha_t\}$	1.5 - 4.6	2.3 - 4.8	1.3 - 2.4	1.1 - 6.3	0.8 - 4.3	1.4 - 3.4
$\{\beta_t\}$	0.8 - 2.7	0.9 - 3.0	0.8 - 3.2	0.9 - 1.9	1.2 - 3.8	0.6 - 1.8
$\{\log \sigma_t^2\}$	0.4 - 0.9	0.4 - 1.0	0.3 - 5.0	0.4 - 1.4	0.3 - 0.6	0.5 - 0.8

The table summarizes effective sample size of the TVP-SV-Fama model parameters, divided by 5,000. For TVPs, the minimum and the maximum of effective sample sizes across all t are reported. Results are based on every 5-th draw from the 50,000 MCMC samples after the burn-in period.

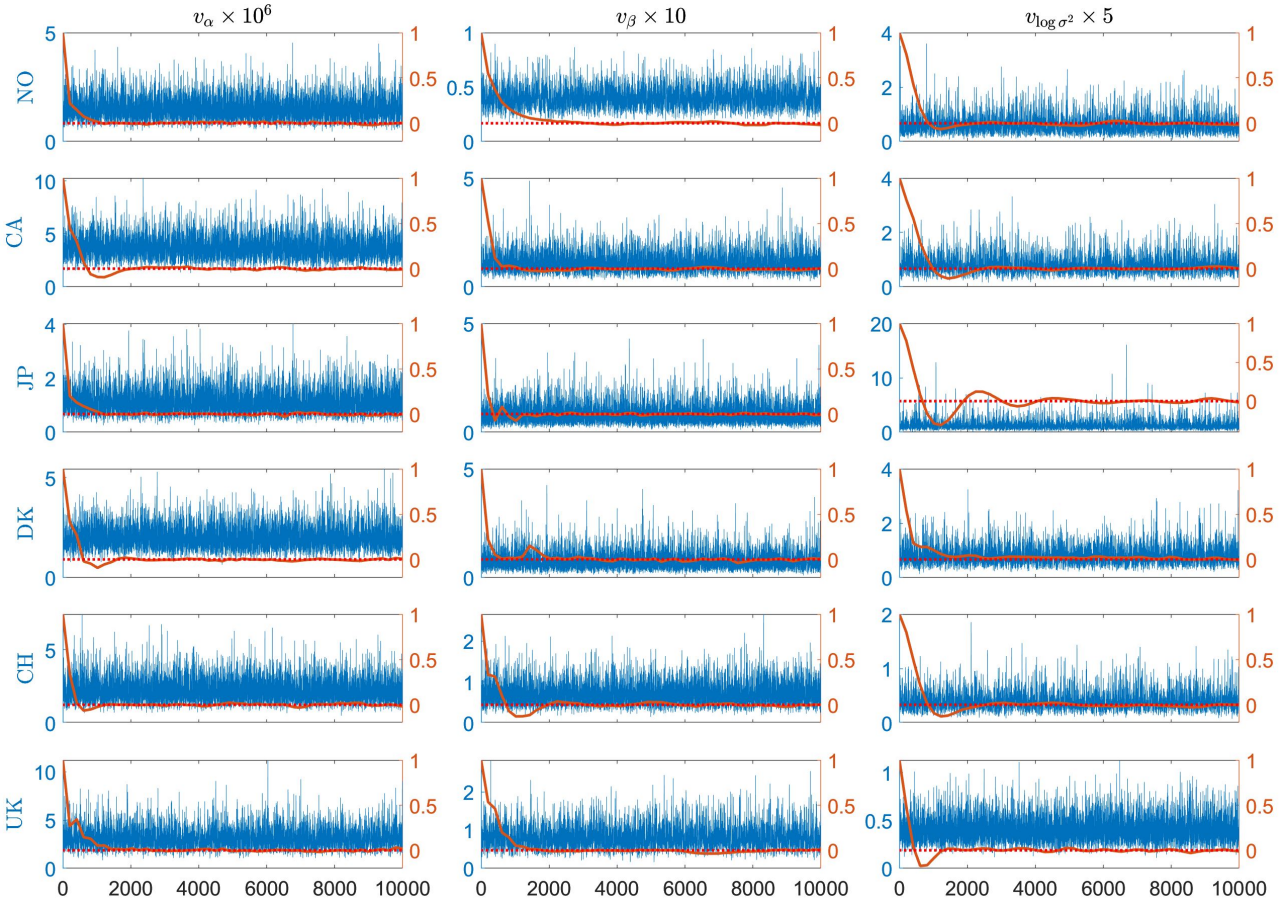

Figure 6: Posterior trace and autocorrelation function. The blue lines associated with the left y-axis indicate the scaled MCMC sample paths for the three variance parameters. The red curves associated with the right y-axis and x-axis ticks divided by 200 show the first 50 autocorrelation functions.

Among all model parameters, the smallest effective sample size is 1,366, sufficiently large for computing posterior statistics reported in the main text. To complement the results on the total absolute autocorrelation function in the main text, Figure 6 reports the posterior traces of the variance parameters and the associated first 50 autocorrelation functions for each of the variances. Convergence is clearly achieved with satisfactory mixing.

C.3. Intercept coefficients from the TVP-SVX-Fama model

Figures 7 shows estimated regression coefficients $\boldsymbol{\gamma}_\alpha$ (rescaled for readability) and summarize which economic shocks affect α_t and by how much. Comparing the the effects on β_t reported in the main text, we see more heterogeneous effects across shocks, with investment and wage

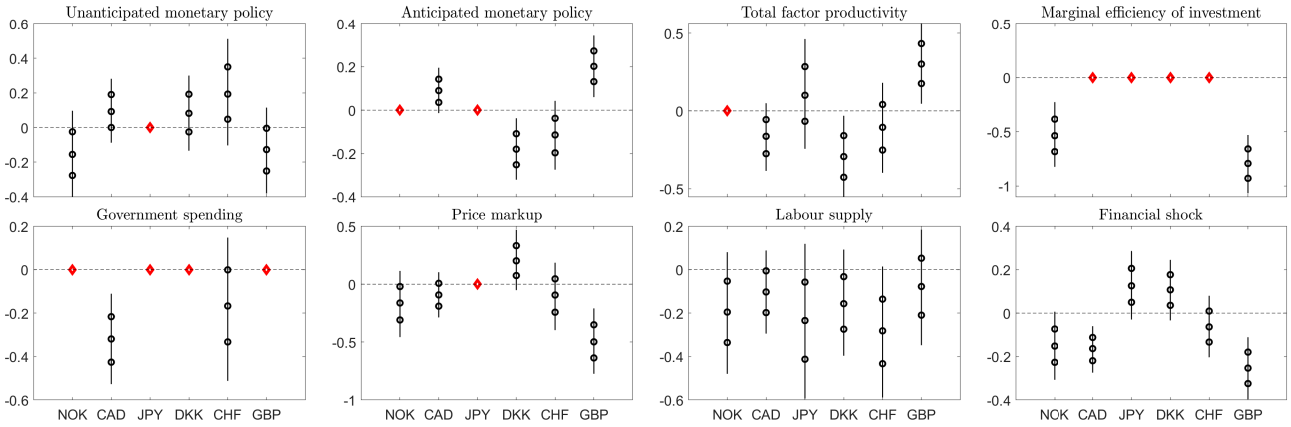

Figure 7: Effects of US shocks on time-varying intercept. Reported are estimates of $100 \times \boldsymbol{\gamma}_\alpha$ of the TVP-SVX-Fama regression for six currency pairs. Each coefficient in $\boldsymbol{\gamma}_\alpha$ measures the responsiveness of α_t to a US shock, whose range, 0.05, 0.5, and 0.95 posterior quantiles are indicated by the vertical bar, the lowest, the middle, and the highest circle, respectively. Red diamonds indicate shocks that are not selected by Bayesian Lasso.

shocks as leading examples. This suggests that structural shocks affect excess returns mostly via the covariance between expected depreciation and interest rate differentials.

D. Model sampling details

D.1. Sampling interest rate parity coefficients

The TVP-SV-Fama model (8)-(9) in the main text gives the conditional likelihood

$$p(\boldsymbol{\delta}|\boldsymbol{\rho}, \boldsymbol{D}, \boldsymbol{\sigma}) \propto |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\boldsymbol{\rho} - \boldsymbol{Z}\boldsymbol{\delta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\rho} - \boldsymbol{Z}\boldsymbol{\delta})\right).$$

The state equation (9) in the main text implies a joint prior for $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ via

$$\boldsymbol{\delta}|\boldsymbol{\theta} \sim N\big(\boldsymbol{\mu}, (\boldsymbol{H}'\boldsymbol{\Omega}^{-1}\boldsymbol{H})^{-1}\big),$$

where $\boldsymbol{\mu} = \mathbf{0}$, $\boldsymbol{H}$ is a sparse lower-triangular band matrix with ones on its main diagonal and minus ones on its second lower diagonal, and $\boldsymbol{\Omega} = \text{diag}(10^6, 10^6, v_\alpha, v_\beta, ..., v_\alpha, v_\beta)$. So, the

conditional prior is given by

$$p_0(\boldsymbol{\delta}|\boldsymbol{\theta}) \propto |\boldsymbol{H}'\boldsymbol{\Omega}^{-1}\boldsymbol{H}|^{\frac{1}{2}} \exp\left(-\frac{1}{2}(\boldsymbol{\delta}-\boldsymbol{\mu})'\boldsymbol{H}'\boldsymbol{\Omega}^{-1}\boldsymbol{H}(\boldsymbol{\delta}-\boldsymbol{\mu})\right).$$

Combining the above with $p(\boldsymbol{\delta}|\boldsymbol{\rho}, \boldsymbol{D}, \boldsymbol{\sigma})$, we obtain the conditional posterior given in the main text.

D.2. Sampling of the TVP-SVX-Fama regression model

Let $\boldsymbol{f} = (f_1, ..., f_T)'$ and $\boldsymbol{X} = (\boldsymbol{x}_1, ..., \boldsymbol{x}_T)'$. Given the conditional likelihood $\tilde{\boldsymbol{H}}\boldsymbol{f}|\boldsymbol{X}, \boldsymbol{\gamma}_f, v_f \sim N(\boldsymbol{X}\boldsymbol{\gamma}, v_f\boldsymbol{I}_T)$ where $\boldsymbol{I}_T$ is a $T \times T$ identify matrix, it can be shown that

$$\boldsymbol{\gamma}_f|\boldsymbol{f}, v_f, \tau_1^2, ..., \tau_J^2 \sim N(\hat{\boldsymbol{\gamma}}_f, v_f\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\gamma}_f}^{-1}),$$

where $\hat{\boldsymbol{\gamma}}_f = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\gamma}_f}^{-1}\boldsymbol{X}'\tilde{\boldsymbol{H}}\boldsymbol{f}$ and $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\gamma}_f} = \boldsymbol{X}'\boldsymbol{X} + \boldsymbol{\Gamma}^{-1}$. Combining the conditional likelihood and the improper prior $p_0(v_f) = 1/v_f$, the conditional posterior for the innovation variance satisfies

$$v_f \sim IG\left(\frac{T-1}{2} + \frac{J}{2}, \frac{(\bar{\boldsymbol{H}}\boldsymbol{f} - \boldsymbol{X}\boldsymbol{\gamma}_f)'(\bar{\boldsymbol{H}}\boldsymbol{f} - \boldsymbol{X}\boldsymbol{\gamma}_f)}{2} + \frac{\boldsymbol{\gamma}_f'\boldsymbol{\Gamma}^{-1}\boldsymbol{\gamma}_f}{2}\right).$$

Combining the conditional likelihood $p_0(\boldsymbol{\gamma}_f|v_f, \tau_1^2, ..., \tau_J^2) \sim N(\boldsymbol{0}, v_f\boldsymbol{\Gamma})$ and the exponential prior $p_0(\tau_j^2) \sim Exp(\lambda/2)$, we obtain the conditional posterior of $1/\tau_j$, $j = 1, ..., J$, given by

$$p(1/\tau_j|\boldsymbol{\gamma}_f, v_f) \propto \sqrt{\lambda}\tau_j^{3/2} \exp\left(-\frac{\lambda\tau_j(1/\tau_j - \sqrt{\mu})^2}{2\mu^2}\right),$$

which is the kernel of the inverse-Gaussian density with parameter $\mu = \sqrt{\lambda v_f}/\gamma_{f,j}$ and λ. We use the algorithm in Eisenstat et al. (2016) to sample from the inverse-Gaussian distribution.

E. Extended model

We consider a TVP-SVX-in-mean model, similar to the SV-in-mean model developed in Sarantis (2006) and Chan (2017). The model is given by

$$\rho_{t+1} = \alpha_t + \beta_t(i_t^* - i_t) + \delta_t \exp(h_t) + \exp\left(\frac{h_t}{2}\right)\epsilon_t, \tag{4}$$

and α_t, β_t and δ_t are TVPs that follow autoregressive distributed lag dynamics driven by their own past, innovations, and a set of US structural shocks; that is

$$f_t = \mu_f(1-\phi_f) + \phi_f f_{t-1} + \boldsymbol{x}_t'\boldsymbol{\gamma} + \eta_{f,t}, \quad f = \{\alpha, \beta, h\}$$

In this formulation, h_t is the time-varying log variance of Fama regression error term. δ_t captures the time-varying feedback effect from the risk premium received from forward speculation. As a robustness check, we consider autoregressive dynamics, rather than a random walk, for state variables.

Figure 8 shows the estimated time-varying intercept in the extended model (4) for the US-CA pair, in comparison to that in the TVP-SVX model. With autoregressive dynamisc, the TVP-SVX-Fama estimate of α_t and $\sigma_t = \exp(h_t/2)$ follow closely with those given in the main text. This means that our resutls are robust to different specifications of state transition.

Comparing the intercept obtained from the SV-in-mean model, we see a clear difference between α_t's from two models. The intercept from the extended model shows a lower degree of time variation. In fact, its 90% credible band covers 0, the theoretical value of the intercept. This result is also obtained by Sarantis (2006) where the author uses a traded volatility index instead of model-implied volatility. The muted variation in α_t prompts one to consider two possibilities: (1) temporal variation in α_t is dominated by that of $\delta_t \exp(h_t)$, as an additive term in (4), or (2) the TVP-SVX model is incorrectly specified.

As can be seen from Figure 8, the evolution of δ_t has low persistence while fluctuating around zero – an overfitting sign that δ_t captures noise rather than a signal in the dynamics of the

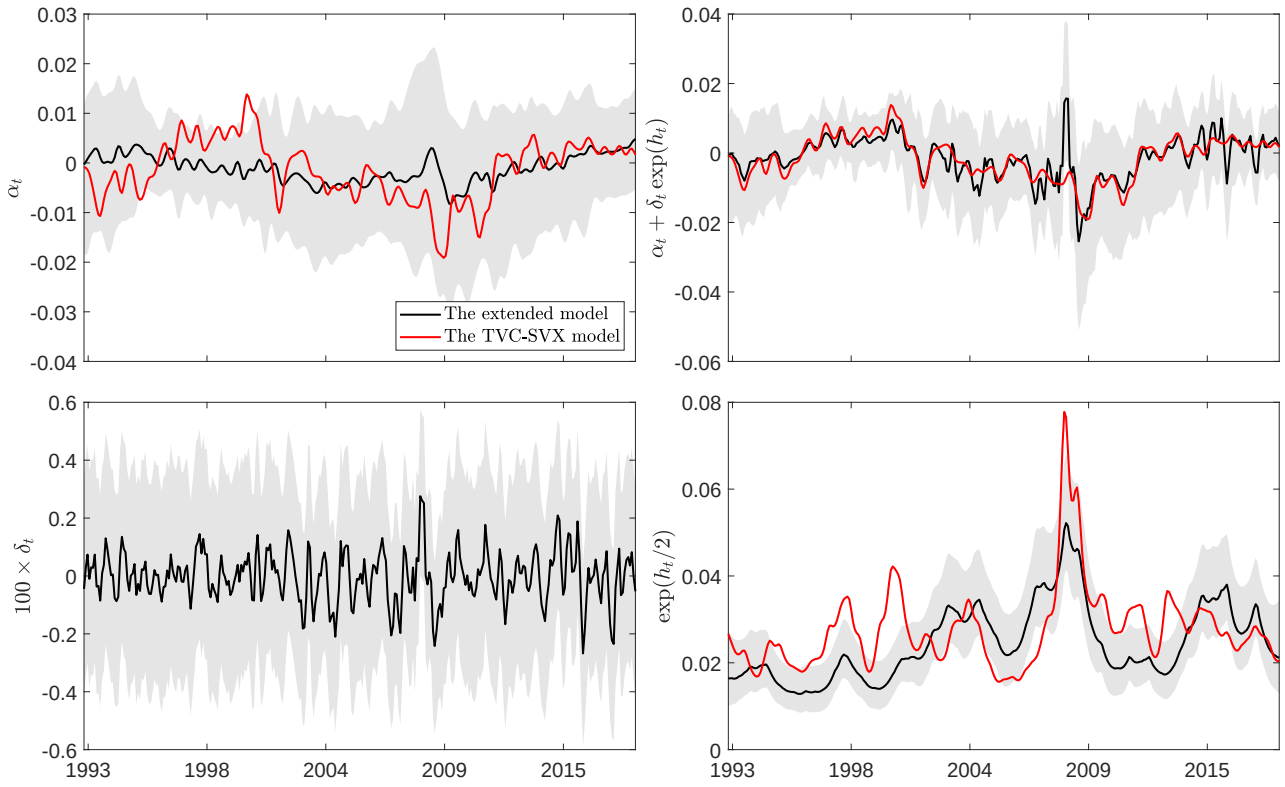

Figure 8: Estimates of time-varying coefficients for Canada under the extended model (4).

exchange rate. Interestingly, when we consider the intercept term in the extended model as $\alpha_t + \delta_t \exp(h_t)$, we observe similarity between the two models as shown in the top-right panel of the figure. Notice that this result does not conflict with δ_t is largely around zero. Because the median of $\delta_t \exp(h_t)$ does not equal to median(δ_t) exp (median(h_t)), due to the nonlinearity of the product function.

The similarity between the estimate of $\alpha_t + \delta_t \exp(h_t)$ from the extended model and α_t from the TVP-SVX model leads us to conclude that the TVP-SVX can capture the effect of risk premium on deviations from UIP in a similar way as the more involved extended model (4). This is to a large extent expected because of the inclusion of US structural shocks in the dynamics of α_t in the TVP-SVX model which span a space that covers the unobserved risk premium. Estimation results from other currency pairs robustly give us the same conclusion. Therefore, we prefer TVP-SVX model over the extended model due to its simpler form and estimation.

References

Chan, J. C. (2017). The stochastic volatility in mean model with time-varying parameters: An application to inflation modeling. *Journal of Business & Economic Statistics 35*(1), 17–28.

Chan, J. C. and I. Jeliazkov (2009). Efficient simulation and integrated likelihood estimation in state space models. *International Journal of Mathematical Modelling and Numerical Optimisation 1*(1-2), 101–120.

Chib, S., F. Nardari, and N. Shephard (2002). Markov chain Monte Carlo methods for stochastic volatility models. *Journal of Econometrics 108*(2), 281–316.

Durbin, J. and S. J. Koopman (2012). *Time series analysis by state space methods*. Oxford university press.

Eisenstat, E., J. C. Chan, and R. W. Strachan (2016). Stochastic model specification search for time-varying parameter VARs. *Econometric Reviews 35*(8-10), 1638–1665.

Engel, C., K. Kazakova, M. Wang, and N. Xiang (2022). A reconsideration of the failure of uncovered interest parity for the US dollar. *Journal of International Economics 136*, 103602.

Giraitis, L., G. Kapetanios, and T. Yates (2014). Inference on stochastic time-varying coefficient models. *Journal of Econometrics 179*(1), 46–65.

Ismailov, A. and B. Rossi (2018). Uncertainty and deviations from uncovered interest rate parity. *Journal of International Money and Finance 88*, 242–259.

Koopman, S. J., A. Lucas, and M. Scharth (2015). Numerically accelerated importance sampling for nonlinear non-gaussian state-space models. *Journal of Business & Economic Statistics 33*(1), 114–127.

Li, M. and S. J. Koopman (2021). Unobserved components with stochastic volatility: Simulation-based estimation and signal extraction. *Journal of Applied Econometrics 36*(5), 614–627.

Petrova, K. (2019). A quasi-bayesian local likelihood approach to time varying parameter VAR models. *Journal of Econometrics 212*(1), 286–306.

Roy, V. (2020). Convergence diagnostics for markov chain Monte Carlo. *Annual Review of Statistics and Its Application 7*, 387–412.

Sarantis, N. (2006). Testing the uncovered interest parity using traded volatility, a time-varying risk premium and heterogeneous expectations. *Journal of International Money and Finance 25*(7), 1168–1186.

Shephard, N. (2015). Martingale unobserved component models. In S. J. Koopman and N. Shephard (Eds.), *Unobserved Components and Time Series Econometrics*, pp. Chapter 10. Oxford University Press.