# Out-of-Sample Forecasting of Foreign Exchange Rates: The Band Spectral Regression and LASSO *

Tatsuma Wada†

**Keio University**

This Version: August 17, 2020

**Abstract**

We propose to utilize the band spectral regression for out-of-sample forecasts of exchange rates. When one period ahead forecast is considered, there is some evidence that the band spectral regression benefits us, especially when the Taylor rule fundamentals model is employed. However, when the forecasting horizon increases, the purchasing power parity (PPP) fundamentals model is found to be powerful, and we can improve the out-of-sample forecast by selecting appropriate frequency bands. Bayesian model averaging shows that placing a high weight on the business cycle frequency improves the accuracy of the out-of-sample forecasting of the PPP model (as well as the monetary fundamentals model) when a longer forecasting horizon is our focus. Without specifying the frequency bands prior to applying the regression, LASSO can provide better out-of-sample exchange rate forecasts for many cases – most patently for the PPP fundamentals model – and provide information about the dynamic relationship between forecasting variables and exchange rates.

**JEL Classification Number:** C22; C52; F31; F47.

*We would like to thank James Morley, Zhongjun Qu, the participants at the 2020 Econometrics Society World Congress, 91st Annual Conference of the Western Economic Association International, the 27th Annual SNDE Symposium, and 2019 BU pi-day econometrics conference for their helpful comments and suggestions. We also acknowledge the financial assistance provided by the Japan Society for the Promotion of Science Grant in Aid for Scientific Research No.15H06585, No.17K03709, the Murata Science Foundation Research Grant, Japan Center for Economic Research Grant, and the Okawa Foundation Research Grant. All data and programs used for this paper are available upon request. The Online Appendix can be downloaded from http://econ.sfc.keio.ac.jp/band_appendix.pdf.

†Faculty of Policy Management, Keio University, 5322 Endo, Fujisawa, Kanagawa, 252-0882, Japan (E-mail: twada@sfc.keio.ac.jp, Tel. +81-466-49-3451, Fax. +81-466-49-3451).

**Keywords:** Band Spectral Regression, Bayesian Model Averaging, Exchange Rate Models, Frequency Domain, LASSO.

## 1 Introduction

Searching for models that describe the mechanism of exchange rate fluctuations is an intriguing research topic in international finance. After three decades of advancements, not only in econometric techniques and new general equilibrium models but also in trading environments in the real world, classical studies, such as Meese and Rogoff (1983), remain as a powerful voice against the existence of such a model. However, when we look at recent developments in this area of study in more detail, we see that a few things are worth further investigation. Among others, Molodtsova and Papell (2009) find that incorporating monetary policy based on the Taylor rule into exchange rate models significantly improves the out-of-sample forecast of exchange rates.

The literature has recognized that the link between macroeconomic variables, in the same way as standard economic models suggest, emerges only within certain frequencies. For example, after utilizing the band-pass filter in the manner of Baxter and King (1995), Pakko (2002) shows that the phenomenon known as the consumption correlation puzzle appears only in business cycle frequencies as defined by Burns and Mitchell (1946). Wada (2012) claims that filtering should be applied to uncover the significant relationship between the exchange rate and the interest rate differential that is implied by the interest parity condition. This is because the existence of a unit root and rational expectations suggest that the Beveridge-Nelson (BN) decomposition should reveal the link between the BN cyclical components of exchange rate and the interest rate differential. Recently, similar results are discovered by Perron and Wada (2016), who apply a very flexible trend-cycle decomposition method proposed by Perron and Wada (2009). Therefore, it is meaningful to investigate whether the exchange rate models' performance, especially in their forecasting power, substantially improves once data are broken down into frequencies. To this end, we employ the band spectral regression proposed by Hannan (1965) and Engle (1974), among others.

The contribution of this paper is two fold. First, we examine the out-of-sample forecasting power of the purchasing power parity (PPP) model, the monetary model, and the Taylor rule model, such as Molodtsova and Papell (2009), by utilizing the band-spectral regression. Extending the idea of band spectral regression, our second contribution is to introduce Bayesian Model Averaging (BMA) and LASSO regression (Tibshirani, 1996) into the literature of the out-of-sample forecast of exchange rates[1]. The reason why we introduce those two elements into this study is that we wish to understand which frequencies are relatively important for out-of-sample forecasting. By regarding the band-spectral regression for a certain frequency band as a model, we are able to compare the relative importance of each frequency band in the framework of BMA. While we find that BMA with the estimated posterior model probability does not always improve the out-of-sample forecasting power, it is remarkably informative as to which frequency band should be used to make better forecasts in some cases.[2] The reason for employing LASSO is that normally we have little knowledge as to which frequency bands should be used in the band spectral regression prior to applying such regressions. In addition, the classification of low-, middle-, and high-frequencies is somewhat arbitrary. Without specifying the frequency band beforehand, LASSO regression only needs one smoothing parameter to be set. In this paper, the smoothing parameter is selected in such a way as to forecast the future exchange rate more precisely. After finding such a smoothing parameter, we estimate the relevant and irrelevant frequencies for out-of-sample forecasts.

Our findings from this study are the following: i) band spectral regression improves the out-of-sample forecasting of exchange rates, with varying degrees of success across countries; ii) the fundamentals included in the Taylor rule model are found to be strong forecasters for a one period ahead forecast, but the PPP fundamentals are a powerful model when forecasting horizons are longer; iii) BMA improves the accuracy of forecasting especially for the PPP fundamentals model, and it suggests the importance of business cycle frequencies especially when long-horizon out-of-sample forecasting of exchange rates is considered; and iv) without specifying the frequency bands prior to applying the regression, LASSO leads us to better out-of-sample exchange rate forecasts for many cases but most patently for the PPP fundamentals model, and LASSO can also provide information about dynamic relationships between forecasting variables and exchange rates.

The rest of this paper is organized as follows. Section 2 details the exchange rate models and forecasting procedures based on the band spectral regression. We also discuss as to why we should focus on frequency bands to provide better forecasts in that section. To further investigate which frequencies should be selected for forecasting, we employ BMA and LASSO in Section 3. Our empirical results are presented in Section

[1] There are some studies using LASSO for out-of-sample forecasting of exchanges rates in the time domain. See for example, Colombo and Pelagatti (2020).

[2] As a recent study by Diebold and Shin (2019) suggests that it is possible to improve the forecasting power by combining BMA with LASSO. However, the approach taken by Diebold and Shin (2019) does not use the frequency domain.

4. Section 5 concludes.

## 2 Forecasting Exchange Rates

### 2.1 Exchange Rate Models

Among a variety of models describing exchange rate dynamics, we consider here the exchange rate models that are frequently used for empirical studies, such as those considered by Mark (1995), Engel et al. (2007), and Molodtsova and Papell (2009).[3] Those models allow researchers to investigate as to whether the current fundamentals, such as money supply and output have statistical associations with the future exchange rate. In other words, with those models, one can examine whether the current fundamentals have a forecasting power on the future exchange rate. Let $s_t$ denote the (logarithm of) nominal, spot exchange rate at time-t. Then, a forecasting model for the h-period ahead exchange rate, $s_{t+h}$ is given by:

$$s_{t+h} - s_t = \alpha_h + \beta_h z_t + u_{t+h} \tag{1}$$

where $z_t = f_t - s_t$; and $f_t$ is (a linear combination of) fundamentals at time-t. One such set of fundamentals are those suggested by the monetary model,

$$f_t = (m_t - m_t^*) - \phi (y_t - y_t^*) \, . \tag{2}$$

where $m_t$ is the log of the money supply, $y_t$ is output, and $\phi$ is a constant that is specified later. Throughout this paper, variables with an asterisk indicate the variables for a foreign country whose currency value relative to a home currency's value is defined as the exchange rate, $s_t$. Departing from the monetary model, Molodtsova and Papell (2009) consider empirical exchange rate models including i) the interest parity condition and ii) the monetary policy based on the Taylor rule, which incorporates the real exchange rate (a la Clarida, Gali, and Gertler 1998).[4] When such departures are taken into account, future exchange rate depreciation is predicted by

$$s_{t+1} - s_t = \alpha_1 + \beta_1' w_t + v_{t+1}, \tag{3}$$

[3] Evans (2011) reviews empirical studies.

[4] Molodtsova and Papell (2009) consider the case involving interest rate smoothing, which implies the observed current interest rate is the weighted average of the target rate and the past interest rate. In such a case, the fundamentals vector $w_t$ includes the past interest rate. This paper, however, does not consider interest rate smoothing.

where the regressor $w_t$ is a vector of variables that pertain to the Taylor rule and $v_{t+1}$ is an error term. In this paper, we will focus on the following variables in our regressor to forecast future depreciations of the exchange rate. Note that equation (3) only focuses on the $h = 1$ horizon in the forecasting model (1), which was first introduced by Mark (1995). This is due to the fact that equation (3) is a consequence of the interest parity condition and the Taylor rule, whereas the forecasting equation (1) does not have a strong theoretical background for higher horizons ($h > 1$).

As a candidate for $w_t$, we first consider the real exchange rate to assess whether the Taylor rule model is only partially correct, in other words, whether the coefficients of the variables, other than the price differentials in the Taylor rule model, are all zero. In such a case, our regressor is

$$w_t = p_t - p_t^* - s_t. \tag{4}$$

Second, we utilize the monetary fundamentals that are often used in the studies of the forecasting model (1):

$$w_t = (m_t - m_t^*) - \phi\,(y_t - y_t^*) - s_t, \tag{5}$$

where $\phi$ is either 0 or 1.

Our third regressor is the Taylor rule equation with a particular set of parameters,

$$w_t = 0.1\,(s_t + p_t^* - p_t) + 1.5\,(\pi_t - \pi_t^*) + 0.1\left(y_t^G - y_t^{*G}\right), \tag{6}$$

where $y_t^G$ is the output gap, which measures the percentage deviation of output from its potential level.

Finally, we employ all the variables that are associated with the Taylor rule, namely, the differentials of inflation rates, output gaps, and the real exchange rate:

$$w_t = \begin{bmatrix} \pi_t & \pi_t^* & y_t^G & y_t^{*G} & p_t - p_t^* - s_t \end{bmatrix}', \tag{7}$$

which is called the asymmetric Taylor rule model with heterogeneous coefficients by Molodtsova and Papell (2009).

### 2.2 Why Does a Frequency Band Matter? Two Reasons

Our rationale for using the band spectral regression – instead of the time domain regression – stems from the following two reasons. First, aside from a group of empirical studies that reveal the importance of filtering or using specific frequencies to uncover the statistically significant link between economic variables, there is a theoretical reason for us focusing on frequency bands to improve our forecast. Consider that process $y$ is generated by

$$y_t = \sum_{j=-\infty}^{\infty} b_j^0 x_{t-j} + \varepsilon_t,$$

where $\{y_t, x_t\}$ are jointly stationary and $E\left[\varepsilon_t x_{t-j}\right] = 0$ for all $j$. Suppose we run a regression assuming

$$y_t = \sum_{j=-\infty}^{\infty} b_j^1 x_{t-j} + u_t,$$

where $b_j^1 = 0$ for some $j$ by assumption. Then, according to Sims (1972) and Sargent (1987), ordinary least squares (OLS) is equivalent to choosing $\left\{b_j^1\right\}$ by solving a minimization problem:

$$\min_{\left\{b_j^1\right\}} \int_{-\pi}^{\pi} \left|b^0\left(e^{-i\omega}\right) - b^1\left(e^{-i\omega}\right)\right|^2 g_x\left(e^{-i\omega}\right) d\omega, \tag{8}$$

where

$$b^k\left(e^{-i\omega}\right) = \sum_{j=-\infty}^{\infty} b_j^k e^{-ij\omega},$$

for $k = 0, 1$; and $g_x\left(e^{-i\omega}\right)$ is the spectral density of $x_t$. The mean squared error is the integral that is minimized in (8) plus a constant that corresponds to the variance of $\varepsilon_t$. Roughly, the (time-domain) OLS finds coefficients $b_j^1$ that minimize the distance between the true parameters $b_j^0$ and $b_j^1$ weighted by the spectral density of $x_t$ over all the frequencies, $(-\pi, \pi)$.

On the other hand, to consider the band spectral regression following Corbae et al. (2002), suppose that we allow low frequency band $[0, \omega_0]$ and high frequency band $(\omega_0, \pi]$ to have different coefficients. More specifically, we assume $\beta_A$ for the low frequency band and $\beta_{A^C}$ for the high frequency band. Then, we have

$$b^1\left(e^{-i\omega}\right) = \beta_A 1_{(|\omega|<\omega_0)} + \beta_{A^C} 1_{(|\omega|>\omega_0)},$$

where $1_{(z)}$ is an indicator function that takes 1 if $z$ is true and takes 0 otherwise.

Note that as clarified in the next section, band spectral regression also minimizes the squared residuals. If our model has the condition that $b_j^1 = 0$ for all $j \neq 0$, then it is clear that band spectral regression has two free parameters to be estimated, while the time domain regression has only one parameter. Hence, excluding the estimation errors, band spectral regression provides a better forecast than time domain regression. However, it is not immediately obvious whether band spectral regression yields a better forecast when we impose a restriction that $\beta_{A^C} = 0$; that is, we only utilize the low frequency band for our regression. Even in this case, we can show that band spectral regression beats time domain regression if, for example, $x_t$ is a white noise process (see Online Appendix 1). Intuitively, this is because the time domain regression fits the model to the data for all frequencies to find a parameter that is constant across frequencies. On the other hand, band spectral regression allows the parameter to vary, either $\beta_A$ or 0, depending on the relative importance of frequencies.

The second reason we investigate band spectral regression is that forecasting higher horizons entails filtering that amplifies low frequency components. To see this, suppose that the first difference of exchange rate $s_t$ has a Wold representation:

$$s_t - s_{t-1} = C(L)\,\eta_t,$$

where $C(L) = 1 + c_1 L + c_2 L^2 + \cdots$, with $\sum_{j=1}^{\infty} c_j^2 < \infty$, $L$ is a lag operator, and $\eta_t$ is a zero-mean white noise with the variance of $\sigma_\eta^2$. Then, the forecasting equation for $h$ periods ahead is

$$s_t - s_{t-h} = A(L)(s_t - s_{t-1}) = A(L)\,C(L)\,\eta_t, \tag{9}$$

where $A(L) = 1 + L + L^2 + \cdots + L^{h-1}$. Hence, the depreciation rate for $h$ periods is equivalent to applying a filter $A(L)$ to the exchange rate, and the filter's transfer function is

$$\left|A\left(e^{-iw}\right)\right|^2 = \left|\frac{1 - e^{-hiw}}{1 - e^{-iw}}\right|^2 = \frac{1 - \cos(hw)}{1 - \cos w}.$$

As Figure 1 clearly shows, a higher $h$ places an extremely large weight on low frequencies and 0 or near 0 weight on high frequencies. Therefore, it is reasonable to utilize only low frequencies to estimate unknown parameters rather than using all the frequencies – as the time domain regression does– because fitting data to the model with less important frequencies, that is, high frequencies, may lead to an imprecise estimation of the parameters owing to the same argument pertaining to equation (8). This fact does not necessarily mean that we should remove high frequency components from the data before running a regression or that high frequency components do not convey useful information for forecasting. However, it is sensible for us to assume that variables' forecasting power varies with frequency, especially when high forecasting horizons are considered.

### 2.3 The Band Spectral Regression

We follow the notations of Corbae et al. (2002) to describe the band spectral regression model. Let us assume that the observed exchange rate (in percentage change over time) $y_t$ follows

$$y_t = \pi_1 + \widetilde{y}_t,$$

where $\pi_1$ is an unknown parameter and $\widetilde{y}_t$ is a zero-mean stationary process. Let us also assume that the variables that represent exchange rate fundamentals are a $k \times 1$ vector $x_t$, which follows

$$x_t = \Pi_2 + \widetilde{x}_t,$$

where $\Pi_2$ is a $k \times 1$ vector and $\widetilde{x}_t$ is a zero-mean, $k \times 1$ time series. For the relationship between the vectors stacking $\widetilde{y}_t$ and $\widetilde{x}_t$, $\widetilde{y}$ and $\widetilde{x}$, respectively, we suppose that the following equations hold:

$$\begin{aligned} AW\widetilde{y} &= AW\widetilde{X}\beta_A + AW\varepsilon && (10) \\ A^c W\widetilde{y} &= A^c W\widetilde{X}\beta_{A^c} + A^c W\varepsilon, && (11) \end{aligned}$$

where $W$ is a matrix[5] for the discrete Fourier transform; $A$ is a frequency band selection matrix whose diagonal entries are either 0 or 1, and the rest of $A$ is zeros; $A^c = I - A$; and $\varepsilon$ is a vector of the error term,

[5] $W = \frac{1}{\sqrt{T}} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & e^{2\pi i/T} & e^{2\pi 2i/T} & e^{2\pi 3i/T} & \cdots & e^{2\pi(T-1)i/T} \\ 1 & e^{2\pi 2i/T} & e^{2\pi 4i/T} & e^{2\pi 6i/T} & \cdots & e^{2\pi 2(T-1)i/T} \\ 1 & e^{2\pi 3i/T} & e^{2\pi 6i/T} & e^{2\pi 9i/T} & \cdots & e^{2\pi 3(T-1)i/T} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & e^{2\pi(T-1)i/T} & e^{2\pi 2(T-1)i/T} & e^{2\pi 3(T-1)i/T} & \cdots & e^{2\pi(T-1)^2 i/T} \end{bmatrix}$

$\varepsilon_t$. Further, we assume that $\varepsilon_t$ and the variable $\widetilde{x}_t$ follow a jointly stationary vector process.[6] Equation (10) states that the coefficient vector on $\widetilde{x}_t$ is $\beta_A$ when a regression of $\widetilde{y}_t$ on $\widetilde{x}_t$ is applied only for the frequency band selected by the selection matrix $A$. For the other bands, as appears in equation (11), the coefficient vector is $\beta_{A^c}$. With the inverse Fourier transform we have the time domain representation of our regression model:

$$\begin{aligned} y &= \pi_1 + \Psi\widetilde{X}\beta_A + \Psi^c\widetilde{X}\beta_{A^c} + \varepsilon & (12) \\ &= \pi_1 - \Psi\Pi_2\beta_A - \Psi^c\Pi_2\beta_{A^c} + \Psi X\beta_A + \Psi^c X\beta_{A^c} + \varepsilon & (13) \\ &= \pi + \Psi X\beta_A + \Psi^c X\beta_{A^c} + \varepsilon & (14) \end{aligned}$$

where $\pi$ is a constant; $\Psi$ and $\Psi^c$ are matrices for the Fourier and inverse-Fourier transforms of the data. More specifically,

$$\Psi = W^+AW, \qquad \Psi^c = W^+A^cW$$

and $W^+$ is a complex conjugate matrix of $W$. It is important to note that one needs to specify the frequency band prior to applying the band spectral regression. If we assume the significant relation between $\widetilde{y}_t$ and $\widetilde{x}_t$ only in the frequency band specified by the selection matrix $A$, and no such relationship in all the other frequencies, it is reasonable to set $\beta_{A^c} = 0$.

### 2.4 Out-of-Sample Forecast and the Band Spectral Regression

As is explained later, the band spectral regression requires pre-specified frequency bands for which the regression is carried out in the frequency domain. To this end, we first specify the following three bands that are commonly used in macroeconomics, namely, high frequency, low frequency, and middle frequency. We define high frequency bands as $(\pi/2, \pi)$ and $(\pi/4, \pi)$; low frequency bands as $(0, \pi/2)$ and $(0, \pi/4)$; and middle frequency bands as $(\pi/2 - .15\pi, \pi/2 + .15\pi)$, and the business cycle frequencies as $(\pi/48, \pi/9)$. This classification of frequency bands, except for the business cycle frequencies, is in accordance with a study by Perron and Yamamoto (2013). Table 1 summarizes the frequency bands used in this study.

[6] We employ Assumption 1 of Corbae et al. (2002): $(\varepsilon_t, \widetilde{x}_t)$ is a jointly stationary time series with Wold representation; and there is no cross spectral for $\varepsilon_t$ and $\widetilde{x}_t$.

Table 1: Frequency Bands

| Name | Frequencies | BMA Model Number |
|---|---|---|
| High 1 | $(\pi/2, \pi)$ | 1 |
| High 2 | $(\pi/4, \pi)$ | 2 |
| Low 1 | $(0, \pi/2)$ | 3 |
| Low 2 | $(0, \pi/4)$ | 4 |
| Middle | $(\pi/2 - .15\pi, \pi/2 + .15\pi)$ | 5 |
| Business Cycle | $(\pi/48, \pi/9)$ | 6 |
| All | $(0, \pi)$ | 7 |

Our out-of-sample forecasting of exchange rates utilizes the rolling method: First, we divide the whole sample that has observations from $t = 1$ to $T$ into two subsamples. (If $T$ is an odd number, $(T + 1)/2$ is used as the size of the first subsample.) Then, using the first subsample, which starts at $t = 1$ and ends at $t = R$ (here, $R = T/2$ if $T$ is an even number), we estimate the coefficients for each band specified above. Let us call this first subsample the estimation sample. Once the parameters are estimated for the estimation sample, the $h-$period ahead forecast, which is meant to forecast $t = R+h+1$, is computed as in equation (12), by fitting the estimated coefficient to the estimation sample that adds the next observation ($t = R + 1$) to the original estimation sample and drops the oldest observation ($t = 1$) from it. Note that the predicted value of the dependent variable at $t = R + h + 1$ appears as the last entry of the estimated vector $\widehat{y}$ in equation (12). This is different from the method commonly used in forecasting based on a linear regression model, where the predicted value is computed by multiplying the estimated coefficients by a single observation of the regressor for $t = R + 1$. The main reason why we need to utilize the subsample rather than a single new observation to compute the predicted value is that our band spectral regression involves both the Fourier and inverse Fourier transforms, as seen in equation (12). Needless to say, the prediction error utilizing all frequencies $(0, \pi)$ yields the identical prediction error that would be computed from the commonly used method.

The prediction error for $t = R+h+1$ is then defined as the difference between the observed $y_t$ and the predicted value thereof, $\widehat{y}_t$. Hence, the prediction error is $e_t = y_t - \widehat{y}_t$. We repeat the procedure above to compute the prediction errors from $t = R + h + 1$ to $t = T$.

## 2.5 The Evaluation of Out-of-Sample Forecasting

Since Meese and Rogoff (1983), one way to measure an exchange rate model's performance is its forecasting ability, especially in comparison to that of the random walk model. For the purpose of assessing the performance of exchange rate models relative to the random walk model in the frequency domain, we employ the ratio of mean squared errors:

$$MSE \text{ ratio} = \frac{\frac{1}{T-R-h}\sum_{t=R+h+1}^{T} e_t^2}{\frac{1}{T-R-h}\sum_{t=R+h+1}^{T} re_t^2},$$

where $re_t^2$ is a square of the prediction error assuming the exchange rate follows the random walk process. More precisely, in such a case, $re_t = \widehat{y}_t - \overline{y}_t$, where $\overline{y}_t$ is the average of $y_t$ over the estimation sample. An MSE ratio greater than one implies that the random walk model better predicts the h-period ahead exchange rate, and vice versa. Further, it is possible to implement a test of whether the out-of-sample forecasting of the exchange rate based on the band spectral regression dominates the null of a random walk process. Given the fact that our alternative hypothesis nests the null hypothesis of the random walk[7], we employ the MSE-F test (e.g., Clark and McCracken, 2013) to test the equal forecast accuracy in the population. The asymptotic distribution for this test statistic is not standard, and therefore, we compute the p-values for each test statistic by a fixed regressor bootstrap (Clark and McCracken, 2012).

# 3 Dealing with Unknown Frequency Bands

The previous section finds that some bands are more important than other bands in discovering the relationship between dependent and independent variables, as well as in forecasting the dependent variable that is out-of-sample. However, one problem that arises in practical studies is how to choose the frequency bands that are utilized in the band spectral regression. Here, we consider two different procedures that select frequencies in our band spectral regressions. They are Bayesian Model Averaging (BMA) and LASSO, as explained and assessed in the following subsections.

[7] Because nested models are tested, a popular accuracy test such as the one proposed by Diebold and Mariano (1995) cannot be used. To see why our models are nested, note that our model under the alternative is

$$y = \pi_1 + \Psi (X - \Pi_2) \beta_A + \Psi^c (X - \Pi_2) \beta_{A^c} + \varepsilon,$$

while the model under the null is

$$y = \pi_1 + \varepsilon.$$

### 3.1 Bayesian Model Averaging (BMA) and the Band Spectral Regression

It is clear from equation (12) that our band spectral regression is a linear regression. Hence, BMA (Fernandez et al., 2001; Magnus et al. 2010, among others) is readily applicable and it can improve our out-of-sample forecasting. Let $\mathcal{M}_i$ be a band spectral regression model utilizing frequency band $i$ and let $N$ be the number of frequency bands under consideration. Since our band spectral regression equation (14), which uses a specific band $i$ in Table 1 can be regarded as model $i$ ($\mathcal{M}_i$), we can rewrite it as:

$$Model\ \mathcal{M}_i \ : \ \ y = \pi_i + X_{2i}\beta_i + \varepsilon_i.$$

For example, model 1 ($\mathcal{M}_1$) uses the high frequency band, $(\pi/2, \pi)$, and model 3 ($\mathcal{M}_3$) uses the low frequency band, $(0, \pi/2)$. In this case, the regressor $X_{2,1}$ is $[W^+AWX, W^+A^cWX]$, where the diagonal elements of the selecting matrix $A$ are 1's only for those corresponding to frequencies between $\pi/2$ and $\pi$, and 0's otherwise. Then, BMA forecast is $\sum_{i=1}^{N} p\left(\mathcal{M}_i|y\right)\widehat{y}_{t,i}$, where $p\left(\mathcal{M}_i|y\right)$ is the posterior model probability and $\widehat{y}_{t,i}$ is the forecast based on $\mathcal{M}_1$. Here, we employ Zellner's (1986) g-priors:

$$\begin{aligned} p\left(\sigma^2|\mathcal{M}_i\right) &\propto \sigma^{-2},\ p\left(\pi|\sigma^2, \mathcal{M}_i\right) \propto 1 \\ p\left(\beta_i|\pi, \sigma^2, \mathcal{M}_i\right) &\propto \left|\sigma^2 V_{0i}\right|^{-1/2} \exp\left\{-\frac{\beta_i' V_{0i}^{-1}\beta_i}{2\sigma^2}\right\} \end{aligned}$$

where $\beta_i = \begin{bmatrix} \beta_A' & \beta_{A^c}' \end{bmatrix}'$ and $V_{0i}^{-1} = g_i X_{2i} M_1 X_{2i}'$ with $g_i = \max\left\{T, k^2\right\}$ and $M_1 = I_T - \iota\iota'/T$, with $\iota = [[1, \ldots, 1]'$.

Then, the posterior model probability for model $\mathcal{M}_i$ is given by

$$p\left(\mathcal{M}_i|y\right) = c\left(\frac{g_i}{1+g_i}\right)^{k_{2i}/2} \left(y' M_1 B_i M_1 y\right)^{-(T-k_1)/2}, \tag{15}$$

where

$$B_i = \frac{g_i}{1+g_i} M_1 + \frac{1}{1+g_i}\left\{M_1 - M_1 X_{2i}\left(X_{2i}' M_1 X_{2i}\right)^{-1} X_{2i}' M_1\right\}$$

and $c$ is a normalizing constant.[8] Once the posterior probabilities $p\left(\mathcal{M}_i|y\right)$ are computed, we proceed to find the out-of-sample prediction errors.

[8] See Magnus et al. (2010) for details.

We also consider the BMA forecast with equal probabilities across all the frequency bands. As is well known (Morley and Panovska, 2019; Geweke and Amisano, 2011; among others), BMA forecasting with equal probabilities often outperforms BMA with posterior model probabilities that are computed from equation (15). We investigate whether the out-of-sample forecasting of exchange rates exhibits a similar tendency.

### 3.2 LASSO for Band Spectral Regression

#### 3.2.1 The Set up for LASSO

In the previous section, BMA requires prespecified frequency bands, and then a researcher can assess the relative importance of each band. Here, we propose utilizing the LASSO regression (Tibshirani, 1996) to search for the frequencies over which a regression is applied, without prespecifying frequency bands. Originally, LASSO was designed to deal with big data and a fairly large number of unknown parameters (coefficients) that are expected to be zero. Because we need to shut down irrelevant frequencies in the spirit of the band spectral regression, as the selection matrix $A$ has many zeros, LASSO is suitable for this problem. The question here is, more precisely, "How good would an out-of-sample forecast be if one were allowed to use 'some' frequencies?" Once the means of $y$ and $X$ are subtracted from those variables[9] and renamed as $y_c$ and $X_c$, respectively, our model (provided that $\beta_{A^c} = 0$) can be estimated from the following regression:

$$\begin{aligned} y_c &= \Psi X_c \beta_A + u_c \\ &= P\zeta + u_c, \end{aligned} \tag{16}$$

where the matrix $P$ consists of the data and known parameters and the vector $\zeta$ consists of $\beta_A$ and the diagonal elements of $A$. In equation (16), we specify $P$ as a $T \times T^2 k$ matrix and $\zeta$ as $T^2 k \times 1$. Here, our goal is to estimate the unknown parameter vector $\beta_A$ and the diagonal elements of $A$. While they are not identified separately, we are able to find the vector $\zeta$ that minimizes the mean squared error of the out-of-sample forecast.

[9] Note, as pointed out by Corbae (2002), that time domain detrending does not lead to a biased estimator when the only deterministic term in the model is an intercept.

This specification, (16), and the fact that $\zeta$ is a vector with most of its elements (possibly) being zeros, leads us to the following LASSO problem to estimate the unknown parameter vector $\zeta$:

$$\widehat{\zeta}_L = \arg\min \left(y_c - P\zeta\right)' \left(y_c - P\zeta\right) - \lambda \left\|\zeta\right\|, \tag{17}$$

where $\|\zeta\|$ is the L-1 norm for a $T^2k \times 1$ vector $\zeta$, i.e., $\|\zeta\| = \sum_{i=1}^{T^2k} |\zeta_i|$.

### 3.2.2 Searching for $\lambda$ That Minimizes the MSE Ratio

To implement LASSO, we employ the MATLAB command "`lasso`" to compute $\widehat{\zeta}_L$. The only unknown parameter in the minimization problem (17) is $\lambda$, which is often called a penalty parameter that is needed for "`lasso.`" Often, cross-validation is utilized for selecting $\lambda$. We, too, select $\lambda$ in a way that minimizes the mean squared prediction errors (i.e., the best out-of-sample forecast is made at the selected value of $\lambda$) relative to that of the random walk model. To this end, we compute the MSE ratio for a given $\lambda$. For each value of $\lambda$, the rolling window is applied and then the MSE ratio is computed. We repeat for different values of $\lambda$, such as $\lambda_1, \lambda_2, \ldots, \lambda_M$; and then we choose $\lambda$ that is the lowest value among $(\lambda_1, \lambda_2, \ldots, \lambda_M)$ that minimizes the MSE ratio and denote $\lambda^*$. Note that the frequencies chosen to compute the MSE ratio vary across rolling windows even if the hyper parameter $\lambda$ is fixed across the rolling window.[10]

It is important to keep in mind that $\widehat{\zeta}_L$ becomes a vector of zeros if a large number is chosen as the penalty parameter, $\lambda$. In such a case, the out-of-sample forecast from the LASSO becomes the mean value of $y$ in the estimation subsample, resulting in the MSE ratio becoming 1 because the out-of-sample forecast of a random walk model is simply the average of the depreciation rate.

Because $\lambda$ is not allowed to vary across rolling windows, one practical use of this setup is to take the $\lambda^*$ in this study and apply it toward forecasting future exchange rates that have not been observed as of

[10] Similar to the band spectral regression, an out-of-sample forecast is computed by the rolling method: Once $\widehat{\zeta}_L$ is found in the estimation sample for a given $\lambda$, the estimation sample is added to a new observation, and the oldest observation is subtracted to find the fitted value vector $\widehat{y}_c$, whose last entry becomes the out-of-sample forecast after being adjusted for the intercept. More precisely, the vector of the out-of-sample forecast for the demeaned process $y_c$ is

$$\widehat{y}_c = P\widehat{\zeta}_L.$$

Then,

$$\widehat{y} = \widehat{k} + \widehat{y}_c$$

where

$$\widehat{k} = \overline{y} - \overline{\widehat{y}_c},$$

and $\overline{y}$ and $\overline{\widehat{y}_c}$ are the averages of $y$ and $\widehat{y}_c$, respectively. Those averages are computed over the estimation subsample.

yet.

### 3.2.3 The Estimation of Relevant Frequencies

Once we find a $\lambda$ that minimizes the MSE ratio, we are able to estimate the selecting matrix $A$. To this end, we utilize the well-known fact (e.g., Engle, 1964) that for $\beta_A$ to be a vector of real numbers, both $\pi + h$ and $\pi - h$ should be included in the regression. Hence, it is possible that we specify our LASSO regression differently so that the size of $P$ is $T \times (T+1)\, k/2$ when $T$ is odd and $T \times (T+2)\, k/2$ when $T$ is even. However, this specification does not allow us to clearly find the $\lambda$ that minimizes the MSE ratio because such a specification does not fit the LASSO regression, which has a greater number of regressors than the sample size. For details, see Online Appendix 2.

Then, it is determined that the product of the $i$-th coefficient, $\beta_A$, and $j$-th frequency, $A_{jj}$, is zero if the corresponding element of $\widehat{\zeta}_L$ is less than 1e-5. Note that our focus here is not to determine whether $A_{jj}$ is one but to find relevant frequencies for the out-of-sample forecast. Because $\widehat{\zeta}_L$ is estimated for each rolling window, we shall plot the estimated $A$ for each frequency and time (rolling window) in order to show how relevant frequencies change over time. The averages of the estimated $A$ over time (rolling window) will also be displayed. For more details, see Online Appendix 2.

## 4 Empirical Results

### 4.1 Data

Our data are an extended version of those used by Molodtsova and Papell (2009), taken from the International Financial Statistics by the International Monetary Fund. They are monthly data from March, 1973 through February, 2017, and we utilize the data for the US, Canada, Japan, Switzerland, the UK, and Australia. As in Molodtsova and Papell (2009), each exchange rate is the local currency price of the US dollar. The data on industrial production indexes are used for output gaps ($y_t^G$). To obtain $y_t^G$, we follow Molodtsova and Papell (2009) and use the following three detrending methods: (1) a linear time trend, (2) a quadratic trend, and (3) the Hodrick and Prescott (1997, hereafter, HP) filter with a smoothing parameter of 14,400 that is applied to the data up to $t-1$. For $m_t$ and $p_t$, we use the M1 and consumer price index data, respectively. See Appendix 1 for more details.

### 4.2 The Band Spectral Regression

#### 4.2.1 The Models

We apply the band spectral regression for equation (3) with three different sets of variables for the regressor $w_t$: PPP (equation 4), Monetary (equation 5 with $\phi = 0$ and 1 and we call them Money 0 and Money 1, respectively), the Taylor rule with a fixed set of coefficients (denoted hereafter "fixed"; equation 6) and the heterogeneous Taylor rule (denoted hereafter "hetero"; equation 7). For the Taylor rule model, our output gaps are derived from three different detrending methods, namely, linear detrending, quadratic detrending, and HP detrending. Therefore, both the fixed coefficients Taylor rule model and the heterogeneous coefficients Taylor rule model have three variations. We also apply the band spectral regression for Mark's empirical model (1) with the PPP and monetary fundamentals for $h = 6$, 12, and 24. Table 2 summarizes the models and their equation numbers. In all cases, we find the MSE ratios for the model that imposes $\beta_{A^c} = 0$ (restricted, and are denoted "R" in Tables 3 through 11), and the model without such a restriction (unrestricted , and are denoted "U" in Tables 3 through 11).

Table 2: The Summary of Models

| Equation | Forecasting Horizon: $h = 1$ | | |
|---|---|---|---|
| (4) | PPP | | |
| (5) | Money 0 ($\phi = 0$) | Money 1 ($\phi = 1$) | |
| (6) | Taylor, fixed, linear detrend | Taylor, fixed, quadratic detrend | Taylor, fixed, HP detrend |
| (7) | Taylor, hetero, linear detrend | Taylor, hetero, quadratic detrend | Taylor, hetero, HP detrend |
| | Forecasting Horizon: $h = 6$, 12, and 24 | | |
| (4) | PPP | | |
| (5) | Money 0 ($\phi = 0$) | Money 1 ($\phi = 1$) | |

#### 4.2.2 The Comparison of the Models for h=1

Tables 3, 4, and 5 demonstrate the MSE ratios associated with the band spectral regression of (3) for each of the exchange rate models. The MSE ratios are tabulated for each frequency band, and the last row shows the MSE ratio for each model utilizing all frequencies. Hence, those numbers appearing in the last row are the MSE ratios for regular linear (time domain) regressions. Numbers in bold font indicate that Clark and McCracken's (2012) test rejects the null hypothesis of predictability equal to the random walk model at the 5% significance level.

From the results presented in Table 3, we find that PPP works only for the Swiss franc and the Japanese yen. For the latter, the superiority of the PPP model, relative to the random walk model, is confirmed for all frequencies, meaning that the time domain regression can yield such a forecast. Hence, we do not have a strong reason to use the band spectral regression. Table 4 suggests that M0 outperforms the random walk model for Canada and Australia, in the business cycle frequencies, whereas M1 beats the random walk model for Canada, Japan, the UK, and Australia in the business cycle frequencies.

When the Taylor rule fundamentals are employed, as shown in Table 5, the out-of-sample forecast becomes a little better. With a set of fixed coefficients, regressions within low frequencies are found to work well for Canada and Australia, and regressions within business cycle frequencies perform well for Switzerland and Australia. However, for those two countries, we have to conclude that the time domain regression can forecast future exchange rates as well (except for the model detrended by the HP filter). According to Table 5, allowing the coefficients of the Taylor rule to be free parameters to be estimated, does not result in more rejections. This means that making the Taylor rule more flexible in terms of its coefficients does not improve the accuracy of the out-of-sample forecast.

Given the results provided here, we proceed to investigate the following models' forecasting power for longer forecasting horizons. They are: PPP fundamentals and monetary fundamentals with $\phi = 1$.

Table 3: MSE Ratios for the PPP Model

| Frequency Band | Canada | | Japan | | Switzerland | | UK | | Australia | |
|---|---|---|---|---|---|---|---|---|---|---|
| | R | U | R | U | R | U | R | U | R | U |
| High 1 | 2.658 | 2.660 | 4.681 | 4.598 | 5.554 | 5.561 | 4.829 | 4.895 | 4.003 | 4.010 |
| High 2 | 2.529 | 2.518 | 4.139 | 4.039 | 4.572 | 4.518 | 4.506 | 4.466 | 3.604 | 3.560 |
| Low 1 | 1.001 | 2.653 | 1.007 | 4.630 | **0.993** | 5.513 | 1.006 | 4.962 | 1.004 | 3.985 |
| Low 2 | 1.002 | 2.533 | 1.008 | 4.089 | 0.998 | 4.714 | 1.005 | 4.528 | 1.005 | 3.646 |
| Middle | 1.294 | 1.301 | 1.556 | 1.540 | 1.740 | 1.771 | 1.766 | 1.836 | 1.445 | 1.456 |
| Business Cycle | 1.002 | 1.003 | 1.003 | **0.984** | 1.005 | **0.985** | 1.025 | 1.042 | 1.009 | 1.007 |
| All | 1.001 | 1.001 | 1.005 | 1.005 | **0.992** | **0.992** | 1.011 | 1.011 | 1.002 | 1.002 |

Notes: 1) The names in the "Frequency Band"columns correspond to those in Table 1. For example, "High 1"is $(\pi/2, \pi)$. 2) The values in the "R"columns are MSE ratios from the restricted model, imposing $\beta_{A^c} = 0$; and The values in the "U"columns are MSE ratios from the unrestricted model, without imposing $\beta_{A^c} = 0$. 3) The numbers in bold indicate that the MSE-F test rejects the null hypothesis of the equal predictability at the 5 % significance level. See Online Appendix 3 for more details.

Table 4: MSE Ratios for the Monetary Models

| Frequency Band | Canada | | Japan | | Switzerland | | UK | | Australia | |
|---|---|---|---|---|---|---|---|---|---|---|
| | R | U | R | U | R | U | R | U | R | U |
| | | | | | M0 | | | | | |
| High 1 | 1.158 | 1.503 | 1.688 | 1.815 | 2.793 | 2.867 | 2.729 | 2.707 | 2.409 | 2.455 |
| High 2 | 1.842 | 1.772 | 1.985 | 2.094 | 4.133 | 4.147 | 3.393 | 3.304 | 3.018 | 3.054 |
| Low 1 | 1.006 | 1.491 | 1.002 | 1.815 | 1.000 | 2.835 | 1.012 | 2.707 | 1.002 | 2.445 |
| Low 2 | 1.004 | 1.803 | 1.000 | 2.128 | 1.000 | 4.147 | 1.014 | 3.304 | 1.001 | 3.033 |
| Middle | 1.206 | 1.155 | 1.320 | 1.468 | 1.960 | 2.031 | 1.721 | 1.738 | 1.406 | 1.431 |
| Business Cycle | **0.987** | 1.010 | 1.007 | 1.006 | 1.003 | 1.020 | 0.999 | 1.003 | **0.992** | **0.989** |
| All | 1.013 | 1.013 | 1.018 | 1.018 | 1.013 | 1.013 | 1.013 | 1.013 | 1.006 | 1.006 |
| | | | | | M1 | | | | | |
| High 1 | 1.292 | 1.249 | 1.511 | 1.626 | 2.822 | 2.905 | 2.135 | 2.187 | 2.219 | 2.275 |
| High 2 | 1.479 | 1.440 | 1.771 | 1.887 | 4.199 | 4.224 | 2.863 | 2.855 | 2.793 | 2.838 |
| Low 1 | 1.006 | 1.217 | 1.001 | 1.625 | 0.998 | 2.866 | 1.029 | 2.187 | 1.004 | 2.259 |
| Low 2 | 1.003 | 1.457 | 1.001 | 1.902 | 0.999 | 4.224 | 1.028 | 2.855 | 1.002 | 2.788 |
| Middle | 1.136 | 1.107 | 1.201 | 1.337 | 1.959 | 2.038 | 1.548 | 1.616 | 1.341 | 1.373 |
| Business Cycle | **0.988** | 1.007 | 1.008 | **0.996** | 1.000 | 1.017 | **0.993** | 1.019 | **0.989** | **0.987** |
| All | 1.012 | 1.012 | 1.015 | 1.015 | 1.010 | 1.010 | 1.036 | 1.036 | 1.007 | 1.007 |

Notes: 1) The names in the "Frequency Band"columns correspond to those in Table 1. For example, "High 1"is $(\pi/2, \pi)$. 2) The values in the "R"columns are MSE ratios from the restricted model, imposing $\beta_{A^c} = 0$; and The values in the "U"columns are MSE ratios from the unrestricted model, without imposing $\beta_{A^c} = 0$. 3) The numbers in bold indicate that the MSE-F test rejects the null hypothesis of the equal predictability at the 5 % significance level. See Online Appendix 3 for more details.

Table 5: MSE Ratios for Taylor Rule Models with Fixed Coefficients

| Frequency Band | Canada | | Japan | | Switzerland | | UK | | Australia | |
|---|---|---|---|---|---|---|---|---|---|---|
| | R | U | R | U | R | U | R | U | R | U |
| | | | | | Linear Detrending | | | | | |
| High 1 | 1.067 | 1.075 | 1.258 | 1.263 | 1.960 | 1.941 | 1.431 | 1.473 | 1.682 | 1.673 |
| High 2 | 1.282 | 1.281 | 1.149 | 1.136 | 2.466 | 2.379 | 1.640 | 1.640 | 1.304 | 1.279 |
| Low 1 | **0.988** | 1.075 | 1.009 | 1.260 | 0.992 | 1.923 | 1.005 | 1.435 | **0.987** | 1.668 |
| Low 2 | **0.988** | 1.299 | 1.010 | 1.218 | 1.002 | 2.409 | 1.004 | 1.733 | **0.991** | 1.272 |
| Middle | 1.107 | 1.124 | 1.011 | 1.024 | 1.216 | 1.230 | 1.337 | 1.397 | 1.101 | 1.100 |
| Business Cycle | 1.001 | 0.991 | 1.004 | 0.998 | 1.031 | **0.971** | 1.043 | 1.056 | 1.012 | **0.986** |
| All | 0.995 | 0.995 | 1.006 | 1.006 | **0.975** | **0.975** | 1.015 | 1.015 | **0.982** | **0.982** |
| | | | | | Quadratic Detrending | | | | | |
| High 1 | 1.066 | 1.072 | 1.269 | 1.277 | 1.958 | 1.944 | 1.456 | 1.504 | 1.682 | 1.671 |
| High 2 | 1.285 | 1.282 | 1.170 | 1.160 | 2.458 | 2.376 | 1.655 | 1.656 | 1.302 | 1.275 |
| Low 1 | **0.988** | 1.072 | 1.009 | 1.272 | **0.990** | 1.925 | 1.008 | 1.466 | **0.986** | 1.665 |
| Low 2 | **0.988** | 1.300 | 1.009 | 1.238 | 1.000 | 2.410 | 1.007 | 1.759 | **0.990** | 1.268 |
| Middle | 1.106 | 1.122 | 1.015 | 1.029 | 1.216 | 1.231 | 1.337 | 1.399 | 1.100 | 1.098 |
| Business Cycle | 1.002 | 0.992 | 1.003 | 0.997 | 1.031 | **0.972** | 1.046 | 1.055 | 1.012 | **0.986** |
| All | 0.995 | 0.995 | 1.007 | 1.007 | **0.973** | **0.973** | 1.019 | 1.019 | **0.981** | **0.981** |
| | | | | | HP Detrending | | | | | |
| High 1 | 1.045 | 1.052 | 1.280 | 1.280 | 2.023 | 2.032 | 1.417 | 1.454 | 1.701 | 1.701 |
| High 2 | 1.221 | 1.222 | 1.178 | 1.159 | 2.529 | 2.464 | 1.608 | 1.597 | 1.314 | 1.298 |
| Low 1 | **0.989** | 1.052 | 1.008 | 1.274 | 1.001 | 2.012 | 1.005 | 1.418 | **0.991** | 1.694 |
| Low 2 | **0.991** | 1.238 | 1.007 | 1.240 | 1.007 | 2.505 | 1.004 | 1.691 | 0.994 | 1.292 |
| Middle | 1.083 | 1.097 | 1.021 | 1.030 | 1.242 | 1.274 | 1.319 | 1.371 | 1.106 | 1.111 |
| Business Cycle | 1.007 | 0.997 | 1.007 | 1.039 | 1.027 | **0.990** | 1.042 | 1.049 | 1.013 | 0.992 |
| All | 0.995 | 0.995 | 1.005 | 1.005 | 0.990 | 0.990 | 1.015 | 1.015 | **0.987** | **0.987** |

Notes: 1) The names in the "Frequency Band"columns correspond to those in Table 1. For example, "High 1"is $(\pi/2, \pi)$. 2) The values in the "R"columns are MSE ratios from the restricted model, imposing $\beta_{A^c} = 0$; and The values in the "U"columns are MSE ratios from the unrestricted model, without imposing $\beta_{A^c} = 0$. 3) The numbers in bold indicate that the MSE-F test rejects the null hypothesis of the equal predictability at the 5 % significance level. See Online Appendix 3 for more details.

Table 6: The MSE Ratios for the Taylor Rule Models with Heterogenous Coefficients

| Frequency Band | Canada | | Japan | | Switzerland | | UK | | Australia | |
|---|---|---|---|---|---|---|---|---|---|---|
| | R | U | R | U | R | U | R | U | R | U |
| | | | | | Linear Detrending | | | | | |
| High 1 | 1.720 | 1.765 | 2.143 | 2.161 | 1.990 | 1.947 | 1.909 | 2.024 | 1.503 | 1.487 |
| High 2 | 1.557 | 1.563 | 1.811 | 1.859 | 2.288 | 2.174 | 1.753 | 1.784 | 1.439 | 1.358 |
| Low 1 | 1.037 | 1.779 | 1.039 | 2.170 | 1.037 | 1.942 | 1.123 | 2.084 | 1.026 | 1.481 |
| Low 2 | 1.035 | 1.555 | 1.038 | 1.825 | 1.041 | 2.166 | 1.127 | 1.843 | 1.040 | 1.378 |
| Middle | 1.156 | 1.187 | 1.172 | 1.209 | 1.236 | 1.227 | 1.230 | 1.357 | 1.086 | **1.049** |
| Business Cycle | 1.026 | 1.075 | 1.001 | 1.252 | 1.045 | 1.447 | 1.121 | 1.438 | 1.059 | 1.416 |
| All | 1.036 | 1.036 | 1.054 | 1.054 | 1.040 | 1.040 | 1.134 | 1.134 | 1.029 | 1.029 |
| | | | | | Quadratic Detrending | | | | | |
| High 1 | 1.611 | 1.624 | 1.949 | 1.958 | 2.106 | 2.076 | 1.625 | 1.698 | 1.471 | 1.463 |
| High 2 | 1.495 | 1.479 | 1.781 | 1.807 | 2.345 | 2.285 | 1.504 | 1.538 | 1.335 | 1.297 |
| Low 1 | 1.036 | 1.640 | 1.036 | 1.964 | 1.031 | 2.072 | 1.109 | 1.747 | 1.015 | 1.459 |
| Low 2 | 1.034 | 1.472 | 1.034 | 1.780 | 1.037 | 2.209 | 1.120 | 1.579 | 1.027 | 1.342 |
| Middle | 1.162 | 1.173 | 1.155 | 1.181 | 1.249 | 1.241 | 1.188 | 1.310 | 1.106 | **1.077** |
| Business Cycle | 1.036 | 1.053 | 0.998 | 1.283 | 1.037 | 1.249 | 1.115 | 1.420 | 1.061 | 1.329 |
| All | 1.026 | 1.026 | 1.054 | 1.054 | 1.015 | 1.015 | 1.111 | 1.111 | 1.019 | 1.019 |
| | | | | | HP Detrending | | | | | |
| High 1 | 1.942 | 1.920 | 3.100 | 3.119 | 2.907 | 2.994 | 2.544 | 2.762 | 1.659 | 1.525 |
| High 2 | 1.858 | 1.842 | 2.969 | 2.957 | 2.984 | 3.045 | 2.414 | 2.489 | 1.625 | 1.504 |
| Low 1 | **0.981** | 1.931 | 1.060 | 3.134 | 1.019 | 2.989 | 1.020 | 2.828 | **0.981** | 1.518 |
| Low 2 | **0.987** | 1.832 | 1.051 | 2.916 | 1.025 | 3.049 | 1.016 | 2.549 | 1.007 | 1.545 |
| Middle | 1.215 | 1.241 | 1.371 | 1.438 | 1.372 | 1.461 | 1.416 | 1.607 | 1.141 | **1.064** |
| Business Cycle | 1.001 | 1.036 | 1.013 | 1.134 | 1.012 | 1.165 | 1.080 | 1.314 | 1.067 | 1.169 |
| All | 1.015 | 1.015 | 1.084 | 1.084 | 1.031 | 1.031 | 1.091 | 1.091 | **0.980** | **0.980** |

Notes: 1) The names in the "Frequency Band" columns correspond to those in Table 1. For example, "High 1" is $(\pi/2, \pi)$. 2) The values in the "R" columns are MSE ratios from the restricted model, imposing $\beta_{A^c} = 0$; and The values in the "U" columns are MSE ratios from the unrestricted model, without imposing $\beta_{A^c} = 0$. 3) The numbers in bold indicate that the MSE-F test rejects the null hypothesis of the equal predictability at the 5 % significance level. See Online Appendix 3 for more details.

### 4.2.3 Longer Horizon Forecasts

As is clear from Table 7, 8, and 9, the MSE ratio generally declines as the forecasting horizon ($h$) increases. Interestingly, however, this tendency does not necessarily mean that the accuracy of the forecast improves with $h$. As shown in Table 8, the monetary fundamentals model has fewer rejections in $h$ =24 than in $h$ =6 or 12. Conversely, the PPP fundamentals model (Table 7) performs reasonably well for higher $h$. In fact, the model beats the random walk model in the business cycle frequencies for all exchange rates when $h = 24$. In addition, the PPP fundamentals model works particularly well for Japan and the UK. For those two countries, even high frequencies yield significantly better-than-random walk forecasts, despite the fact that filtering associated with computing 24-month ahead forecast – shown in equation (9) – leaves only a small power spectrum for high frequencies. Interestingly, for those cases, the time domain regression also provides statistically significant forecasts. It indicates that almost all frequencies are so important for forecasting those exchange rates that even the band spectral regression assessing each band separately can confirm significant forecasts. On the contrary, the time domain regression cannot beat the random walk model but the business cycle can do so for Canada, Japan, Switzerland, and Australia with $h = 12$, and for Canada, Switzerland, and Australia with $h = 24$. Here, we confirm the usefulness of the band spectral regression for the purpose of out-of-sample forecasting.

When the monetary fundamentals model is considered in Table 8, the 6-month ahead Australian dollar can be forecast for any of the frequency bands except for the "all" frequencies, meaning that the time domain forecast is inferior to the band spectral regression. However, such forecasting power declines when $h$=12 or 24. It is worth noting that there is some evidence that the business cycle frequencies are important for the UK and Australia when $h = 12$ (and when $h = 24$ for Australia), and that the significance of business cycle frequencies is affirmed by the band spectral regression; the time domain regression cannot outperform the random walk model.

Table 7: MSE Ratios for the PPP Model with h=6, 12 and 24

| Frequency Band | Canada | | Japan | | Switzerland | | UK | | Australia | |
|---|---|---|---|---|---|---|---|---|---|---|
| | R | U | R | U | R | U | R | U | R | U |
| | | | | | h=6 | | | | | |
| High 1 | 0.941 | 0.934 | 0.880 | 0.862 | 0.920 | 0.922 | 0.937 | **0.783** | 0.901 | 0.879 |
| High 2 | 0.907 | 0.906 | 0.828 | 0.813 | 0.915 | 0.952 | 0.911 | **0.778** | 0.923 | 0.933 |
| Low 1 | 1.000 | 0.934 | 0.978 | 0.862 | 1.010 | 0.922 | **0.841** | **0.783** | 0.977 | 0.879 |
| Low 2 | 1.003 | 0.906 | 0.972 | 0.813 | 1.019 | 0.952 | **0.858** | **0.778** | 0.979 | 0.933 |
| Middle | 0.976 | 0.972 | 0.916 | 0.916 | 0.971 | 0.976 | 0.978 | **0.818** | 0.958 | 0.937 |
| Business Cycle | 0.999 | 1.010 | 0.942 | 1.001 | 0.951 | 0.976 | **0.899** | 0.939 | 1.013 | 1.069 |
| All | 0.998 | 0.998 | 1.012 | 1.012 | 1.018 | 1.018 | **0.845** | **0.845** | 1.001 | 1.001 |
| | | | | | h=12 | | | | | |
| High 1 | 0.978 | 0.926 | 0.847 | 0.784 | 0.978 | 0.952 | 1.029 | **0.696** | 0.912 | 0.761 |
| High 2 | 0.878 | 0.862 | 0.879 | 0.813 | 0.877 | 0.950 | 0.854 | **0.631** | 0.918 | 0.898 |
| Low 1 | 0.976 | 0.926 | 0.935 | 0.784 | 1.014 | 0.952 | **0.693** | **0.696** | 0.907 | 0.761 |
| Low 2 | 0.979 | 0.862 | **0.916** | 0.813 | 1.045 | 0.950 | **0.755** | **0.631** | **0.897** | 0.898 |
| Middle | 0.953 | 0.931 | 0.929 | 0.914 | 0.936 | 0.969 | 0.940 | **0.655** | 0.948 | 0.901 |
| Business Cycle | **0.782** | 0.814 | **0.675** | **0.749** | **0.774** | **0.777** | **0.639** | **0.599** | **0.704** | **0.816** |
| All | 0.980 | 0.980 | 1.016 | 1.016 | 1.018 | 1.018 | **0.672** | **0.672** | 0.979 | 0.979 |
| | | | | | h=24 | | | | | |
| High 1 | 0.940 | 0.851 | 0.819 | **0.656** | 0.943 | 0.859 | 0.931 | **0.584** | 0.907 | 0.729 |
| High 2 | 0.885 | 0.822 | 0.777 | **0.629** | 0.837 | 0.816 | **0.724** | **0.539** | 0.884 | 0.778 |
| Low 1 | 0.931 | 0.853 | **0.796** | **0.658** | 0.938 | 0.861 | **0.636** | **0.581** | 0.819 | 0.731 |
| Low 2 | 0.923 | 0.822 | **0.814** | **0.629** | 1.005 | 0.816 | **0.794** | **0.539** | 0.801 | 0.778 |
| Middle | 0.956 | 0.917 | 0.898 | **0.748** | 0.922 | 0.862 | **0.881** | **0.507** | 0.948 | 0.873 |
| Business Cycle | **0.642** | 0.755 | **0.499** | **0.571** | **0.509** | **0.444** | **0.628** | **0.465** | **0.517** | **0.666** |
| All | 0.973 | 0.973 | **0.840** | **0.840** | 0.900 | 0.900 | **0.502** | **0.502** | 0.942 | 0.942 |

Notes: 1) The names in the "Frequency Band"columns correspond to those in Table 1. For example, "High 1"is $(\pi/2, \pi)$. 2) The values in the "R"columns are MSE ratios from the restricted model, imposing $\beta_{A^c} = 0$; and The values in the "U"columns are MSE ratios from the unrestricted model, without imposing $\beta_{A^c} = 0$. 3) The numbers in bold indicate that the MSE-F test rejects the null hypothesis of the equal predictability at the 5 % significance level. See Online Appendix 3 for more details.

Table 8: MSE Ratios for the M0 Model with h=6, 12 and 24

| Frequency Band | Canada | | Japan | | Switzerland | | UK | | Australia | |
|---|---|---|---|---|---|---|---|---|---|---|
| | R | U | R | U | R | U | R | U | R | U |
| | | | | | h=6 | | | | | |
| High 1 | 0.665 | 0.673 | 0.884 | 0.910 | 0.837 | 0.851 | 0.922 | 0.943 | **0.673** | **0.649** |
| High 2 | 0.552 | 0.547 | 0.824 | 0.842 | 0.822 | 0.844 | 0.893 | 0.934 | **0.582** | **0.576** |
| Low 1 | 0.973 | 0.673 | 1.012 | 0.910 | 1.023 | 0.851 | 1.013 | 0.928 | 0.963 | **0.649** |
| Low 2 | 0.972 | 0.547 | 1.009 | 0.842 | 1.019 | 0.844 | 1.013 | 0.934 | 0.968 | **0.576** |
| Middle | 0.805 | 0.805 | 0.883 | 0.912 | 0.907 | 0.904 | 0.949 | 0.980 | **0.802** | **0.775** |
| Business Cycle | 0.978 | 1.043 | 1.067 | 1.073 | 0.970 | 1.253 | 1.038 | 1.072 | **0.936** | **0.909** |
| All | 1.008 | 1.008 | 1.032 | 1.032 | 1.053 | 1.053 | 1.021 | 1.021 | 0.965 | 0.965 |
| | | | | | h=12 | | | | | |
| High 1 | 0.671 | 0.617 | 0.837 | 0.900 | 0.763 | 0.801 | 0.960 | 0.923 | 0.699 | 0.638 |
| High 2 | 0.569 | 0.544 | 0.854 | 0.885 | 0.727 | 0.828 | 1.007 | 1.052 | 0.639 | 0.623 |
| Low 1 | 0.920 | 0.617 | 1.029 | 0.900 | 1.081 | 0.801 | 0.986 | 0.923 | 0.916 | 0.638 |
| Low 2 | 0.912 | 0.544 | 1.016 | 0.885 | 1.060 | 0.828 | 0.977 | 1.052 | 0.923 | 0.623 |
| Middle | 0.807 | 0.791 | 0.896 | 0.949 | 0.836 | 0.960 | 0.991 | 1.020 | 0.804 | 0.758 |
| Business Cycle | 0.971 | 1.143 | 1.049 | 1.033 | 0.879 | 1.383 | **0.797** | **0.811** | **0.820** | **0.842** |
| All | 0.999 | 0.999 | 1.071 | 1.071 | 1.186 | 1.186 | 1.030 | 1.030 | 0.931 | 0.931 |
| | | | | | h=24 | | | | | |
| High 1 | 0.672 | 0.592 | 0.784 | 0.811 | 0.754 | 0.966 | 0.958 | 0.754 | 0.675 | 0.550 |
| High 2 | 0.562 | 0.495 | 0.733 | 0.733 | 0.642 | 0.847 | 0.918 | 0.789 | 0.564 | 0.478 |
| Low 1 | 0.818 | 0.589 | 0.988 | 0.810 | 1.265 | 0.966 | 0.794 | 0.754 | 0.823 | 0.550 |
| Low 2 | **0.787** | 0.495 | 0.964 | 0.733 | 1.225 | 0.847 | 0.796 | 0.789 | 0.833 | 0.478 |
| Middle | 0.801 | 0.788 | 0.855 | 0.888 | 0.824 | 1.187 | 0.962 | 0.832 | 0.785 | 0.683 |
| Business Cycle | 0.887 | 1.244 | 1.075 | 1.134 | 0.921 | 1.944 | 1.005 | 1.495 | **0.629** | **0.702** |
| All | 1.011 | 1.011 | 1.060 | 1.060 | 1.495 | 1.495 | 0.850 | 0.850 | 0.861 | 0.861 |

Notes: 1) The names in the “Frequency Band”columns correspond to those in Table 1. For example, “High 1”is $(\pi/2, \pi)$. 2) The values in the “R”columns are MSE ratios from the restricted model, imposing $\beta_{A^c} = 0$; and The values in the “U”columns are MSE ratios from the unrestricted model, without imposing $\beta_{A^c} = 0$. 3) The numbers in bold indicate that the MSE-F test rejects the null hypothesis of the equal predictability at the 5 % significance level. See Online Appendix 3 for more details.

Table 9: MSE Ratios for the M1 Model with h=6, 12 and 24

| Frequency Band | Canada | | Japan | | Switzerland | | UK | | Australia | |
|---|---|---|---|---|---|---|---|---|---|---|
| | R | U | R | U | R | U | R | U | R | U |
| | | | | | h=6 | | | | | |
| High 1 | **0.658** | 0.659 | 0.864 | 0.899 | 0.821 | 0.844 | 0.968 | 1.006 | **0.669** | **0.652** |
| High 2 | 0.543 | 0.537 | 0.796 | 0.826 | 0.816 | 0.842 | 0.993 | 1.047 | **0.578** | **0.576** |
| Low 1 | 0.975 | 0.659 | 1.020 | 0.899 | 1.040 | 0.844 | 1.023 | 1.003 | 0.972 | **0.652** |
| Low 2 | 0.975 | 0.537 | 1.016 | 0.826 | 1.031 | 0.842 | 1.025 | 1.047 | 0.976 | **0.576** |
| Middle | 0.801 | 0.799 | 0.866 | 0.907 | 0.893 | 0.911 | 0.987 | 1.029 | **0.799** | **0.781** |
| Business Cycle | 0.979 | 1.033 | 1.067 | 1.064 | 0.985 | 1.334 | 1.011 | 1.037 | **0.940** | **0.919** |
| All | 1.003 | 1.003 | 1.043 | 1.043 | 1.078 | 1.078 | 1.027 | 1.027 | 0.974 | 0.974 |
| | | | | | h=12 | | | | | |
| High 1 | 0.660 | 0.605 | 0.817 | 0.916 | 0.740 | 0.818 | 1.001 | 0.999 | 0.691 | 0.649 |
| High 2 | 0.564 | 0.531 | 0.818 | 0.886 | 0.720 | 0.841 | 1.091 | 1.169 | 0.636 | 0.626 |
| Low 1 | 0.925 | 0.605 | 1.068 | 0.916 | 1.135 | 0.818 | 1.001 | 0.999 | 0.933 | 0.649 |
| Low 2 | 0.919 | 0.531 | 1.050 | 0.886 | 1.099 | 0.841 | 0.997 | 1.169 | 0.939 | 0.626 |
| Middle | 0.802 | 0.776 | 0.878 | 0.972 | 0.820 | 0.994 | 1.025 | 1.073 | 0.804 | 0.771 |
| Business Cycle | 0.969 | 1.135 | 1.062 | 1.020 | 0.894 | 1.532 | **0.810** | **0.800** | **0.843** | **0.859** |
| All | 0.989 | 0.989 | 1.121 | 1.121 | 1.267 | 1.267 | 1.033 | 1.033 | 0.946 | 0.946 |
| | | | | | h=24 | | | | | |
| High 1 | 0.676 | 0.601 | 0.750 | 0.857 | 0.721 | 1.029 | 0.967 | 0.811 | 0.668 | 0.556 |
| High 2 | 0.567 | 0.504 | 0.685 | 0.763 | 0.612 | 0.863 | 0.954 | 0.883 | 0.567 | 0.483 |
| Low 1 | 0.842 | 0.602 | 1.071 | 0.856 | 1.392 | 1.029 | 0.802 | 0.811 | 0.833 | 0.557 |
| Low 2 | 0.818 | 0.504 | 1.034 | 0.763 | 1.319 | 0.863 | 0.822 | 0.883 | 0.844 | 0.483 |
| Middle | 0.800 | 0.781 | 0.827 | 0.953 | 0.802 | 1.290 | 0.968 | 0.851 | 0.785 | 0.686 |
| Business Cycle | 0.880 | 1.257 | 1.071 | 1.134 | 0.882 | 2.096 | 0.928 | 1.017 | **0.641** | **0.692** |
| All | 0.998 | 0.998 | 1.170 | 1.170 | 1.697 | 1.697 | 0.817 | 0.817 | 0.860 | 0.860 |

Notes: 1) The names in the "Frequency Band" columns correspond to those in Table 1. For example, "High 1" is $(\pi/2, \pi)$. 2) The values in the "R" columns are MSE ratios from the restricted model, imposing $\beta_{A^c} = 0$; and The values in the "U" columns are MSE ratios from the unrestricted model, without imposing $\beta_{A^c} = 0$. 3) The numbers in bold indicate that the MSE-F test rejects the null hypothesis of the equal predictability at the 5 % significance level. See Online Appendix 3 for more details.

#### 4.2.4 Is Band Spectral Regression Data Snooping?

It is possible that we find a rejection of the null hypothesis of a better forecast than the random walk model at the 5% level by chance because we repeatedly test the null hypothesis over different frequency bands and different models. This is often called data snooping or data mining (see White, 2000). This is another reason why we use fixed regressor bootstrapping.

### 4.3 Bayesian Model Averaging (BMA)

#### 4.3.1 The Comparison of Three Models for h=1 with BMA

Given the results of band spectral regressions, we limit ourselves to considering the following models to see whether BMA improves the accuracy of out-of-sample forecasts and find out which frequency bands are relatively important in forecasting the future exchange rate. The models considered are: the PPP model, the monetary model with $\phi = 1$ (M1), and the Taylor rule model with fixed coefficients plus output gaps derived from quadratic detrending.

Since the posterior model probabilities are computed for each rolling window, we display the average of the posterior model probabilities over the forecasting windows in Figure 2. With posterior probabilities in hand, we are able to assess whether the out-of-sample forecast improves due to model averaging. We present the MSE ratios from BMA in Table 9. A bold number indicates that the MSE ratio is lower than any of the corresponding currency's MSE ratios found in the fixed band spectral regression in Subsection 4.2.1. Two things are clear from the comparison of these three models. First, for all cases, the MSE ratio of BMA with equal weights (probabilities) is always smaller than that with posterior probabilities. Second, such an observation may not be relevant because only one case (the Taylor rule model for Japan) outperforms the fixed band spectral regression.

Table 10: MSE Ratios of Bayesian Model Averaging using $p(\mathcal{M}_i|y)$ and Equal Weights for h=1

| Model | Canada | | Japan | | Switzerland | | UK | | Australia | |
|---|---|---|---|---|---|---|---|---|---|---|
| | R | U | R | U | R | U | R | U | R | U |
| | | | | $p(\mathcal{M}_i \| y)$ | | | | | | |
| PPP (h=1) | 1.344 | 2.599 | 1.557 | 4.412 | 1.944 | 4.916 | 1.872 | 4.510 | 1.693 | 3.640 |
| Monetary (h=1) | 1.027 | 1.173 | 1.049 | 1.418 | 1.808 | 3.668 | 1.409 | 2.550 | 1.613 | 2.542 |
| Taylor (h=1) | 1.040 | 1.209 | 1.006 | 1.175 | 1.318 | 2.325 | 1.177 | 1.440 | 1.103 | 1.420 |
| | | | | Equal Weights (1/7) | | | | | | |
| PPP (h=1) | 1.086 | 1.628 | 1.338 | 2.266 | 1.446 | 2.620 | 1.413 | 2.543 | 1.288 | 2.087 |
| Monetary (h=1) | 0.999 | 1.076 | 1.026 | 1.285 | 1.353 | 2.115 | 1.207 | 1.681 | 1.185 | 1.620 |
| Taylor (h=1) | 1.024 | 1.084 | **0.989** | 1.070 | 1.086 | 1.444 | 1.101 | 1.302 | 1.040 | 1.159 |

Notes: 1) The values in the "R"columns are MSE ratios from the restricted model, imposing $\beta_{A^c} = 0$; and The values in the "U"columns are MSE ratios from the unrestricted model, without imposing $\beta_{A^c} = 0$. 2) The bold number indicates that the MSE ratio is lower than any of the corresponding currency's MSE ratios found in the fixed band spectral regression in Subsection 4.2.1.

### 4.3.2 Longer Horizon Forecast with BMA

Posterior probabilities for higher $h$ (6, 12, and 24 month-ahead forecasts) reveal quite interesting features. As $h$ increases, the probabilities for the business cycle frequencies become larger, as well as those for the entire frequency bands. In the case of the PPP fundamentals displayed in Figure 3, the probability of the business cycle frequencies is approaching 1 for Switzerland, and greater than 0.9 for Canada and Japan when $h = 24$. However, such a phenomenon is more moderate when the monetary fundamentals are considered, yet the business cycle frequencies carry the highest probabilities for all countries when $h = 12$ and 24 (see Figure 4).

As Table 11 shows, the MSE ratio from BMA is promising for the PPP fundamentals model, while the monetary fundamentals model does not perform well, regardless of how the probabilities are computed (i.e., either the posterior model probability or equal weight). When PPP fundamentals are used, forecasts with the posterior probability ($p(\mathcal{M}_i|y)$) have lower MSE ratios compared to forecasts with the equal weight for $h = 12$ and 24. However, this tendency is less clear when $h = 6$.

In addition, all the (bold number) cases where BMA beats the fixed band regression (shown in Table 7) are BMA with posterior probabilities, not with equal probabilities – except for one case of the monetary fundamentals model, which overall performs poorly throughout this BMA exercise.

Table 11: The MSE Ratios of Bayesian Model Averaging using $p(\mathcal{M}_i|y)$ and Equal Weights

| Model | Canada | | Japan | | Switzerland | | UK | | Australia | |
|---|---|---|---|---|---|---|---|---|---|---|
| | R | U | R | U | R | U | R | U | R | U |
| | | | | | $p(\mathcal{M}_i\|y)$ | | | | | |
| PPP (h=6) | 0.999 | 0.967 | 0.984 | 0.970 | 0.973 | 0.986 | **0.840** | 0.857 | 0.980 | 1.011 |
| PPP (h=12) | 0.794 | **0.782** | 0.698 | **0.748** | **0.765** | 0.771 | 0.656 | 0.603 | 0.767 | 0.799 |
| PPP (h=24) | 0.656 | **0.742** | **0.510** | **0.560** | **0.509** | 0.444 | 0.562 | **0.447** | 0.587 | **0.624** |
| Monetary (h=6) | 0.939 | 0.918 | 1.032 | 1.024 | 0.956 | 1.164 | 1.023 | 1.034 | 0.817 | 0.711 |
| Monetary (h=12) | 0.949 | 1.089 | 1.051 | 1.004 | 1.019 | 1.224 | 0.828 | 0.803 | 0.804 | 0.795 |
| Monetary (h=24) | 0.896 | 1.209 | 1.070 | 1.108 | 1.118 | 1.614 | 0.972 | 1.018 | 0.609 | 0.669 |
| | | | | | Equal Weights (1/7) | | | | | |
| PPP (h=6) | 0.962 | 0.941 | 0.916 | 0.884 | 0.946 | 0.945 | 0.871 | 0.802 | 0.940 | 0.935 |
| PPP (h=12) | 0.901 | 0.879 | 0.836 | 0.797 | 0.863 | 0.876 | 0.719 | 0.613 | 0.837 | 0.834 |
| PPP (h=24) | 0.846 | 0.835 | 0.698 | 0.631 | 0.722 | 0.729 | 0.627 | 0.501 | 0.746 | 0.762 |
| Monetary (h=6) | 0.827 | 0.728 | 0.921 | 0.890 | 0.895 | 0.889 | 0.981 | 1.009 | 0.811 | 0.700 |
| Monetary (h=12) | 0.792 | 0.706 | 0.931 | 0.919 | 0.876 | 0.941 | 0.937 | 0.995 | 0.794 | 0.703 |
| Monetary (h=24) | 0.732 | 0.709 | 0.899 | 0.886 | 0.934 | 1.173 | 0.813 | **0.790** | 0.701 | 0.595 |

Notes: 1) The values in the "R" columns are MSE ratios from the restricted model, imposing $\beta_{A^c} = 0$; and The values in the "U" columns are MSE ratios from the unrestricted model, without imposing $\beta_{A^c} = 0$. 2) The bold number indicates that the MSE ratio is lower than any of the corresponding currency's MSE ratios found in the fixed band spectral regression in Subsection 4.2.1.

Given the fact that bold numbers indicate that BMA is useful for those cases (currencies), we should look at the figures of the average posterior probabilities for those cases again. As displayed in Figure 3 (and Figure 4 for the single case), clearly, the business cycle frequencies are the key to making better out-of-sample forecasts.

For BMA with higher $h$, we conclude that BMA improves the accuracy of the forecast using the PPP fundamentals, and the business cycle frequencies are particularly useful in forecasting the future exchange rate. However, the monetary fundamentals model performs poorly, and hence, we are not certain as to which frequency band is more helpful in increasing the accuracy of the forecast.

## 4.4 The LASSO Regression

### 4.4.1 Three Models for h=1 with LASSO

What does LASSO tell us about the exchange rate models? First, let us determine whether the MSE ratio computed by LASSO is smaller than any of the MSE ratios that are found in the band spectral regression in Tables 3, 4, and 6. As graphic examples, Figure 5 and Figure 6 (for $h = 6$) demonstrate that the MSE ratio can be less than 1 if $\lambda$ is selected appropriately. Not surprisingly, each country has a different $\lambda$ that minimizes the MSE ratio.

MSE ratios are presented in Table 12. With the help of LASSO, the PPP fundamental model improves the accuracy of out-of-sample forecasts for all exchange rates. Both monetary fundamentals and Taylor rule models somewhat succeed in gaining higher accuracy compared to the fixed band spectral regressions.

Table 12: Minimum MSE Ratios for h=1

| Model | Canada | Japan | Switzerland | UK | Australia |
|---|---|---|---|---|---|
| PPP | **0.991** | 1.000 | **0.978** | **0.993** | **0.992** |
| Monetary ($\phi = 1$) | 0.991 | **0.992** | 1.000 | **0.944** | 0.993 |
| Taylor (Fixed coefficients, Detrend 2) | 0.991 | **0.967** | 0.979 | **1.000** | **0.972** |

Note: The numbers in bold indicate that the MSE ratio is lower than any of the corresponding currency's MSE ratios found in the fixed band spectral regression in Subsection 4.2.1.

Note, once again, that the MSE ratio computed from LASSO has an upper bound of 1 because a sufficiently large $\lambda$ renders all elements of $A$ zero, making the model (16) equivalent to the random walk model. Here, we only present the estimated $A$ (both for the average over the rolling windows and for each rolling window) for those that outperform the fixed band spectral regression (i.e., cases where MSE ratios are presented in bold numbers).

Recall that the PPP fundamentals perform poorly in the band spectral regression for $h = 1$. Figures 7(a) and (b) in the Online Appendix show that LASSO can improve the forecast and give the answer to the question of why the band spectral regression does not work well. For Canadian dollars and Australian dollars, seemingly there is a structural break in the relationship between the fundamentals and the exchange rate. More precisely, LASSO gives a lower MSE ratio than that of the band spectral regression when $\lambda$ is selected in a way that eliminates all frequency components before some certain rolling window, implying that MSE ratios would be very high before some point due to the poor forecasting power of the model. In the early sample period, the relationship does not exist for all frequencies, but it suddenly appears in the later sample period. The estimated $A$ for the monetary fundamentals model in Figures 8(a) and (b) in the Online Appendix also indicate structural breaks for the UK pound. As for the Japanese yen, there seems to be relative importance in lower frequencies.

As for the Taylor rule fundamentals model in Figures 9(a) and (b) in the Online Appendix, the Japanese yen and the UK pound can be forecast by the Taylor rule only for the early period of their samples. These facts suggest that the reason why the Taylor rule fundamentals fail to forecast the future exchange rate, regardless of frequency, with the band spectral regression is that the link between the fundamentals and exchange rate was severed during the sample period. However, one can forecast the Australian dollar by the Taylor rule when low or all frequencies are utilized. The UK pound also has a structural break, but for most of the sample, the link between the fundamentals and the exchange rate is preserved for almost all frequencies. The forecast of the Swiss franc can be improved slightly, by using selective frequencies. Overall, however, the PPP fundamentals for the all sample period can forecast the Swiss franc quite well.

### 4.4.2 High h with LASSO

For longer forecasting horizons, Table 13 reveals that the PPP fundamentals model performs quite well with the exception of Japan. Interestingly, the monetary fundamentals model works well only for Japan with $h = 6$, 12, 24 and the UK with $h = 6$.

Table 13: Minimum MSE Ratios for h=6,12, and 24

| Model | Canada | Japan | Switzerland | UK | Australia |
|---|---|---|---|---|---|
| PPP (h=6) | **0.424** | 1.000 | **0.379** | **0.745** | **0.529** |
| PPP (h=12) | **0.305** | 1.000 | **0.286** | 0.816 | **0.372** |
| PPP (h=24) | **0.213** | 1.000 | **0.227** | 0.916 | **0.332** |
| Monetary (h=6) | 0.887 | **0.537** | 1.000 | **0.939** | 0.856 |
| Monetary (h=12) | 0.933 | **0.467** | 1.000 | 0.978 | 0.938 |
| Monetary (h=24) | 1.000 | **0.458** | 1.000 | 0.983 | 0.950 |

Note: The numbers in bold indicate that the MSE ratio is lower than any of the corresponding currency's MSE ratios found in the fixed band spectral regression in Subsection 4.2.1.

Why does LASSO sometimes fail to reach a lower MSE ratio than the one that the fixed band spectral regression reaches? Given the way we select a $\lambda$ that minimizes the MSE ratio, there are two possible explanations pertaining to the existence of abrupt changes, outliers, or structural breaks.

First, suppose there is a fixed band that consistently has the lowest squared error across rolling windows. Note that $\lambda$ is assumed to be the same for all rolling windows, and the relevant frequencies for forecasting are estimated as the solution to (17) within each rolling window, given $\lambda$. However, such estimated frequencies may vary with the rolling window if there are outliers in some rolling windows that push the estimated frequency away from the true ones in those windows. In such a case, the resulting MSE ratio from LASSO becomes larger than that from the fixed band spectral regression. Second, suppose there is a rolling window whose out-of-sample forecast yields a relatively large squared error due to abrupt changes in the variables, compared to other rolling windows. Then, LASSO would select a $\lambda$ that eliminates such a large squared error, which we call $\lambda^*$, but this $\lambda^*$ may not be optimal for the rest of the rolling windows.

Looking at Figures 10 through 12 in the Online Appendix, the PPP fundamentals work well because of the absence of a structural break in the link between the fundamentals and the exchange rate for all the horizons considered. Therefore, it is possible to make a forecast by properly choosing frequencies from low and middle ranges.[11] Additionally, consistent with the results from the band spectral regression and BMA, business cycle frequencies (to some extent, low frequencies as well) play an important role in forecasting out-of-sample exchange rates as $h$ increases.

When monetary fundamentals are utilized, the Japanese yen is the only currency (exchange rate) for which LASSO finds a lower MSE ratio than the band spectral regression does. As Figures 13(a) and (b) in the Online Appendix show, we can draw similar conclusions as the PPP model: the lack of a structural break and the relative importance of low and business cycle frequencies allow for making a better forecast.

All in all, LASSO provides (i) satisfactory forecasting power relative to fixed band spectral regression, (ii) information as to which frequencies are important for an accurate forecast, and (iii) information as to why band spectral regression sometimes fails to forecast the future exchange rate. It can also provide information about the dynamic relationship between forecasting variables and exchange rates.

## 5 Conclusion

We propose that band spectral regression be utilized to improve the out-of-sample forecast of exchange rates. When a one-period-ahead forecast is considered, there is some evidence that the band spectral regression benefits us, especially when the Taylor rule fundamentals model is employed. However, when the forecasting horizon increases, the PPP fundamentals model is found to be powerful, and we can improve the out-of-sample forecast by selecting the appropriate frequency bands. BMA shows that placing a high weight on the business cycle frequency improves the accuracy of the out-of-sample forecasting of the PPP model (as well as the monetary fundamentals model) when a longer forecasting horizon is our focus. Without specifying the frequency bands prior to applying the regression, LASSO can generally provide

[11] Note that the MSE ratio of the band spectral regression reaches less than 1 but is not statistically significant. A drawback of the LASSO approach is that one cannot see whether the forecast is statistically significantly better than the random walk model.

better out-of-sample exchange rate forecasts for many cases but most patently for the PPP fundamentals model.

## References


[1] Baxter, M. and R. King (1999) "Measuring Business Cycles: Approximate Band-Pass Filters for Economic Time Series," *Review of Economics and Statistics* 81, 575-593.

[2] Burns, A.M. and W.C. Mitchell (1946) *Measuring Business Cycles*, National Bureau of Economic Research: New York.

[3] Clarida, R., J. Gali and M. Gertler (1998) "Monetary Rules in Practice: Some International Evidence," *European Economic Review* 42, 1033-1067.

[4] Clark, T. E. and M.W. McCracken (2012) "Reality Checks and Comparisons of Nested Predictive Models," *Journal of Business and Economic Statistics* 30 (1), 53-66.

[5] Clark, T. E. and M.W. McCracken (2013) "Advances in Forecast Evaluation," in Elliott, G. and A. Timmermann Eds. *Handbook of Economic Forecasting*,Volume 2B.

[6] Colombo, E. and M. Pelagatti (2020) "Statistical Learning and Exchange Rate Forecasting," *International Journal of Forecasting*, forthcoming.

[7] Corbae D., S. Ouliaris, and P.C.B. Phillips (2002) "Band Spectral Regression with Trending Data," *Econometrica* 70 (3), 1067-1109.

[8] Diebold, F.X. and R.S. Mariano (1995) "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics* 13, 253–263.

[9] Diebold, F.X. and M. Shin (2019) "Machine Learning for Regularized Survey Forecast Combination: Partially-Egalitarian LASSO and Its Derivatives," *International Journal of Forecasting*, forthcoming.

[10] Engel, C., Mark, N.C. and K.D. West (2007) "Exchange Rate Models Are Not As Bad As You Think," In Acemoglu, D., Rogoff, K. and M. Woodford (Eds.), *NBER Macroeconomics Annual*, University of Chicago Press.

[11] Engle, R. (1974) "Band Spectrum Regression," *International Economic Review*, 15, 1-11.

[12] Evans, M.D.D. (2011) *Exchange Rate Dynamics*, Princeton University Press.

[13] Fernández, C., E. Ley, and M.F.J. Steel (2001) "Benchmark Priors for Bayesian Model Averaging," *Journal of Econometrics* 100 (2001) 381-427.

[14] Geweke, J. and G. Amisano (2011) "Optimal Prediction Pools," *Journal of Econometrics* 164, 130–141.

[15] Hannan, E.J. (1965) "The Estimation of Relationships Involving Distributed Lags," *Econometrica* 33 (1), 206-224.

[16] Hodrick, R. and E. Prescott (1997) "Postwar US Business Cycles: An Empirical Investigation," *Journal of Money, Credit, and Banking* 29, 1-16.

[17] Magnus, J.R., O.Powell, and P.Prüfer (2010) "A Comparison of Two Model Averaging Techniques with an Application to Growth Empirics," *Journal of Econometrics* 154, 139-153.

[18] Mark, N.C. (1995) "Exchange Rate and Fundamentals: Evidence on Long-Horizon Predictability," *American Economic Review* 85, 201–218.

[19] Meese, R.A. and K. Rogoff (1983) "Empirical Exchange Rate Models of the Seventies: Do They Fit Out of Sample?" *Journal of International Economics* 14, 3–24.

[20] Molodtsova, T. and D.H. Papell (2009) "Out-of-Sample Exchange Rate Predictability with Taylor Rule Fundamentals," *Journal of International Economics* 77, 167-180.

[21] Morley, J.C. and I.B. Panovska (2019) "Is Business Cycle Asymmetry Intrinsic in Industrialized Countries?" *Macroeconomic Dynamics*, forthcoming.

[22] Pakko, M.R. (2002) "What Happens When the Technology Growth Trend Changes? Transition Dynamics, Capital Growth, and the 'New Economy,'" *Review of Economic Dynamics* 5, 376-407.

[23] Perron, P. and T. Wada (2009) "Let's Take a Break: Trends and Cycles in US Real GDP," *Journal of Monetary Economics* 56 (6), 749-765.

[24] Perron, P. and T. Wada (2016) "Measuring Business Cycles with Structural Breaks and Outliers: Applications to International Data," *Research in Economics*, 70 (2) 281-303.

[25] Perron, P. and Y. Yamamoto (2013) "Estimating and Testing Multiple Structural Changes in Linear Models Using Band Spectral Regressions," *Econometrics Journal* (16), 400-429.

[26] Sargent, T. (1987) *Macroeconomic Theory*, 2nd Ed. Boston : Academic Press.

[27] Sims, C. (1972) "The Role of Approximate Prior Restrictions in Distributed Lag Estimation," *Journal of the American Statistical Association*, 67 (337), 169-175.

[28] Tibshirani, R.J. (1996) "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society Series B*, (58) 267-288.

[29] White, H.(2000) "A Reality Check for Data Snooping," *Econometrica*, 68 (5), 1097-1126.

[30] Wada. T. (2012) "The Real Exchange Rate and Real Interest Differentials: The Role of the Trend-Cycle Decomposition," *Economic Inquiry* 50 (4) 968-987.

[31] Zellner, A. (1986) "On Assessing Prior Distributions and Bayesian Regression Analysis with g-prior Distributions," in: Goel, P.K., and Zellner, A. (Eds.), *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti*. North-Holland, Amsterdam, pp. 233 243.

**Appendix 1: The Data**

| Country | Data Start | Data End |
|---|---|---|
| Canada | March, 1973 | February, 2017 |
| Japan | March, 1973 | February, 2017 |
| Switzerland | March, 1973 | November, 2013 |
| UK | March, 1973 | April, 2006 |
| Sweden | January, 1991 | February, 2017 |
| Australia | March, 1973 | February, 2017 |

Notes: i) The money supply for the UK is not M1, which is not available for most sample periods. Instead, we use M0 for the UK. ii) Monthly industrial production data are not available for Switzerland and Australia. The monthly data of CPI are not available for Australia. For those cases, we follow Molodtsova and Papell (2009): use the Eviews command "quadratic-match average" to render quarterly data to monthly data.

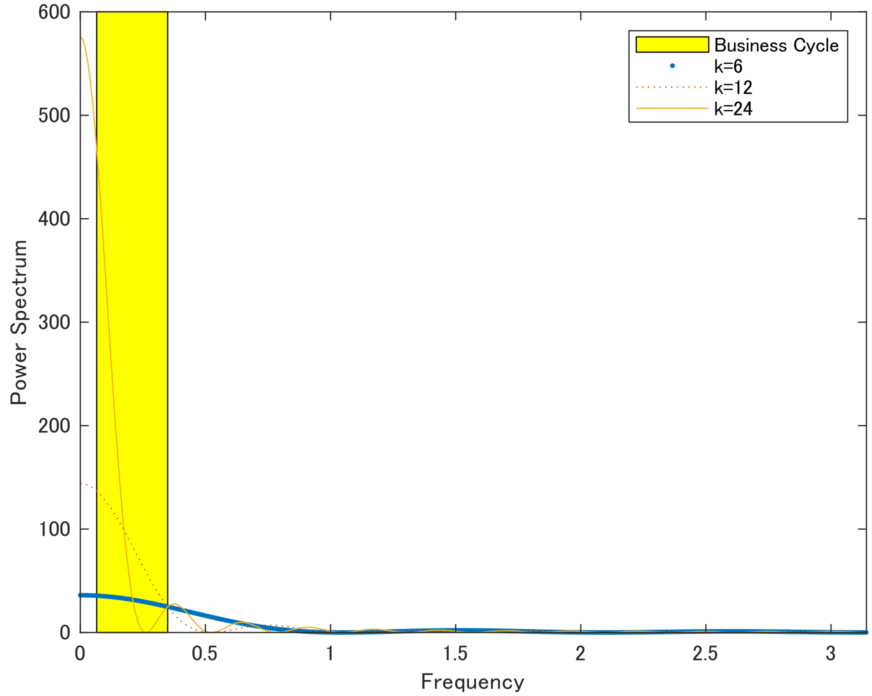


Figure 1: The Squared Gains for Filtering ($h = 6$, 12, and 24).

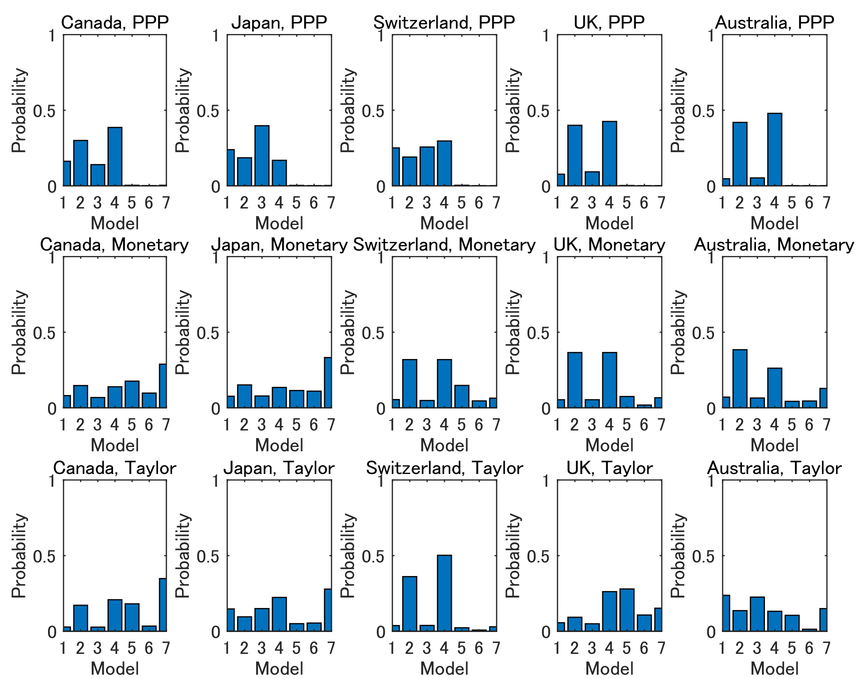


Figure 2: Posterior Model Probabilities for the PPP, Monetary, and Taylor Rule Fundamentals Models for $h = 1$. (Model numbers correspond to the frequency bands in Table 1)

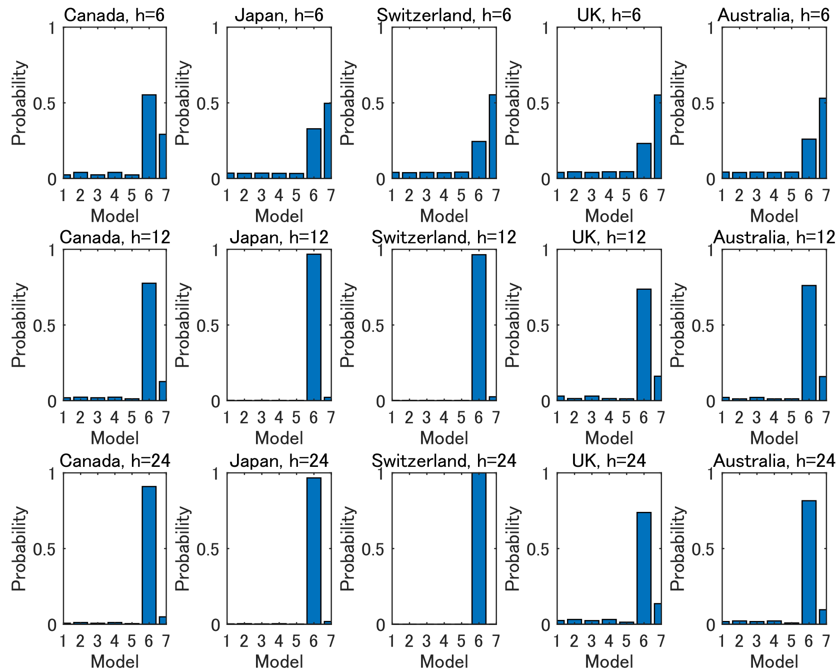


Figure 3: Posterior Model Probabilities for the PPP Fundamentals Model for $h = 6$, 12, and 24. (Model numbers correspond to the frequency bands in Table 1)

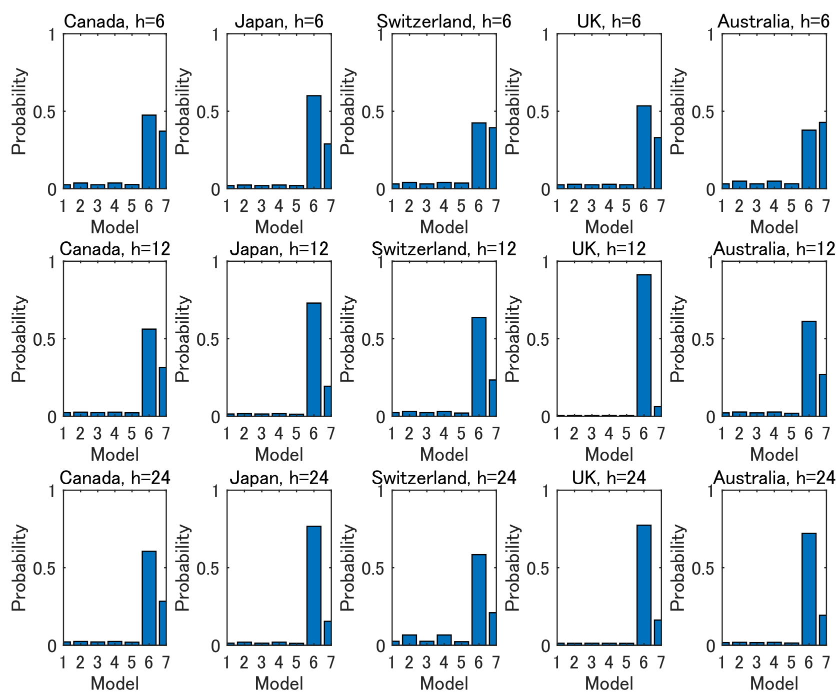


Figure 4: Posterior Model Probabilities for the Monetary Fundamentals Model for $h = 6$, 12, and 24. (Model numbers correspond to the frequency bands in Table 1)

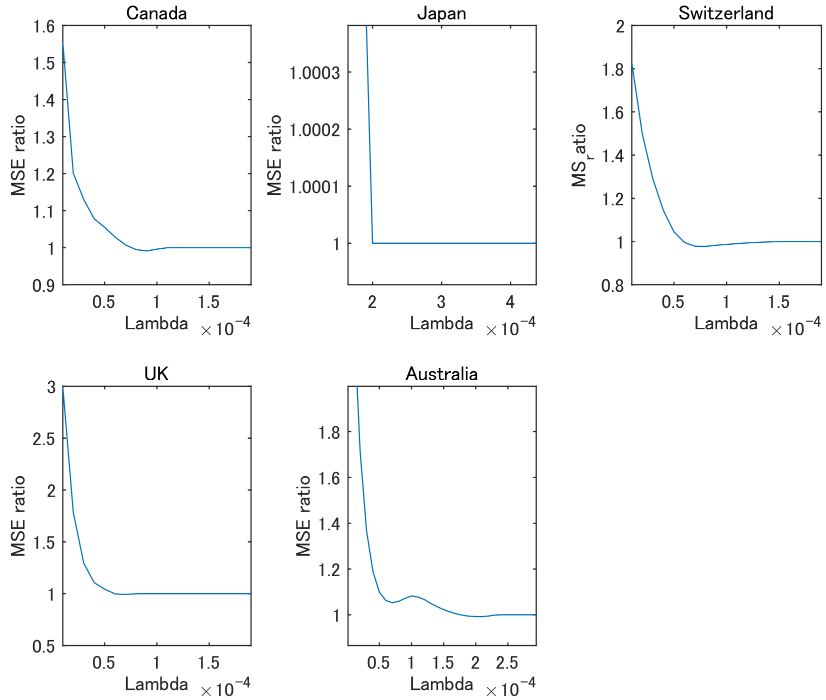


Figure 5: MSE ratio and $\lambda$ for the PPP fundamentals model with $h = 1$

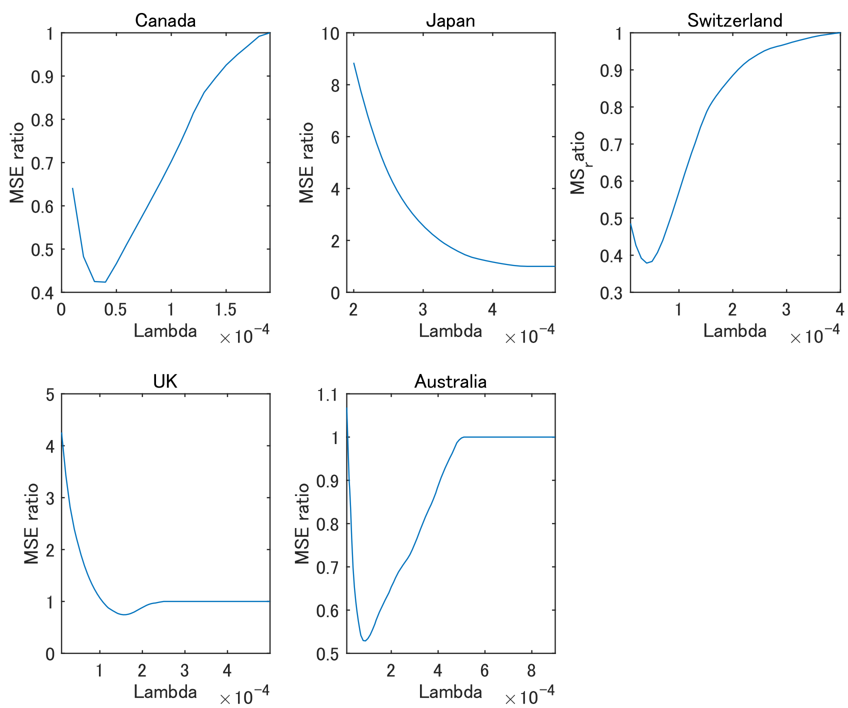


Figure 6: MSE ratio and $\lambda$ for the PPP fundamentals model with $h = 6$