Appendix B: the de-noising source separation algorithm[1]

This appendix is a not-for-publication appendix, and it discusses the de-noising source separation algorithm.

Consider the following linear, noisy independent component model:

$$\mathbf{X} = \mathbf{A}\mathbf{F} + \mathbf{v}, \qquad \text{(A1)}$$

where $\mathbf{X} = [\mathbf{x}_1, ..., \mathbf{x}_N]'$, $\mathbf{F} = [\Delta\mathbf{f}_1, ..., \Delta\mathbf{f}_K]'$, $\mathbf{A} = [\mathbf{a}_1, ..., \mathbf{a}_K]$ and $\mathbf{v} = [\mathbf{v}_1, ..., \mathbf{v}_N]'$ in which $\mathbf{x}_j = [x_{j,1}, ..., x_{j,T}]$, $\Delta\mathbf{f}_i = [\Delta f_{i,1}, ..., \Delta f_{i,T}]$, $\mathbf{a}_i = [a_{1,i}, ..., a_{N,i}]'$ and $\mathbf{v}_j = [v_{j,1}, ..., v_{j,T}]$ for j=1,...,N, i=1,…,K and $N \geq K$. The data matrix **X** has been centered and pre-whitened as described in appendix A. The matrix **F** consists of K different sources; **A** is the mixing matrix that consists of K mixing vectors, $\mathbf{a}_i$; and **v** consists of i.i.d. Gaussian noises. Furthermore, we assume that **F** has zero mean, unit variance and finite moments up to the fourth order and that sources are statistically independent and are independent of those of **v**. The sources in **F** are latent variables. The mixing coefficients in matrix **A** are also assumed to be unknown. The problem is to estimate the mixing matrix (**A**) and the independent sources (**F**) using only the data matrix (**X**).

The de-noising source separation (DSS) algorithm can be justified as an expectation-maximization (EM) algorithm for sources separation that is typically used to obtain maximum likelihood estimates when part of the data are missing. One way to execute the EM algorithm in a linear model is to assume that the missing data consist of sources and hence the mixing matrix needs to be re-estimated. The DSS algorithm proceeds by alternating between two steps. In the E-step, the posterior distribution of the sources is constructed based on the given data as well as the current estimates of the mixing matrix using Bayes' theorem. In the M-step, the mixing matrix is fitted to the new sources estimates. In short, sources are denoised in the E-step and a new mixing vector is re-estimated in the M-step:

E-step: compute $q(\mathbf{F}) = p(\mathbf{F} \mid \mathbf{A}, \mathbf{X}) = p(\mathbf{X} \mid \mathbf{A}, \mathbf{F})\, p(\mathbf{F}) / p(\mathbf{X} \mid \mathbf{A})$. (A2)

M-step: find $\mathbf{A}_{new} = \arg\max_{\mathbf{A}} \mathrm{E}_{q(\mathbf{F})}\left[\log p(\mathbf{F} \mid \mathbf{X}, \mathbf{A})\right] = \mathbf{C}_{\mathbf{XF}} \mathbf{C}_{\mathbf{FF}}^{-1}$, (A3)

[1] We thank Professor Valpola's discussion and comments.

where,

$$\mathbf{C}_{\mathbf{XF}} = \frac{1}{T}\sum_{t=1}^{T} E\left[\mathbf{x}_t \Delta\mathbf{f}_t' \mid \mathbf{X}, \mathbf{A}\right] = \frac{1}{T}\sum_{t=1}^{T} \mathbf{x}_t E\left[\Delta\mathbf{f}_t' \mid \mathbf{X}, \mathbf{A}\right] = \mathbf{X}\mathbf{F}' / T \ ,$$

$$\mathbf{C}_{\mathbf{FF}} = \frac{1}{T}\sum_{t=1}^{T} E\left[\Delta\mathbf{f}_t \Delta\mathbf{f}_t' \mid \mathbf{X}, \mathbf{A}\right] = \widehat{\mathbf{F}\mathbf{F}'} / T \ ,$$

$\Delta\mathbf{f}_t = \left[\Delta f_{1,t}, \Delta f_{2,t}, ..., \Delta f_{K,t}\right]'$ and $\mathbf{x}_t = \left[x_{1,t}, x_{2,t}, ..., x_{N,t}\right]'$. Since data have been pre-whitened, the transformed data are orthogonal and their variance is equal to unity, i.e. $\mathbf{C}_{\mathbf{XX}} = \mathbf{XX}' / T = \mathbf{I}$. Furthermore, it is common to fix the variance of the sources to unity: $\mathbf{FF}' / T = \mathbf{I}$. With this restriction, the mixing matrix $\mathbf{A}$ is orthogonal after pre-whitening the data. This is because: $E\left(\mathbf{XX}'\right) = \mathbf{AFF}'\mathbf{A}' / T + \mathbf{\Omega} = \mathbf{I}$, where $\mathbf{\Omega} = \text{diag}(\sigma_1^2, \ \sigma_2^2, \ ..., \ \sigma_N^2)$ is the covariance of noises in which σ_j^2 is the variance of $v_{j,t}$. If $\mathbf{\Omega}$ is proportional to the covariance of data that is due to sources, i.e., $\mathbf{\Omega} \propto \mathbf{AA}'$, then $\mathbf{AA}' \propto \mathbf{I}$. This indicates that the mixing matrix $\mathbf{A}$ is orthogonal. The DSS-algorithm alternates between computing the posterior distribution of sources given the current point estimates (E-step) and then using the posterior distribution of the sources to compute new maximum likelihood estimates (M-step).

Because sources are independent and noises are Gaussian, the likelihood of **F**: $L(\mathbf{F}) = p(\mathbf{X}|\mathbf{A}, \mathbf{F})$ can be factorized as:

$$L(\mathbf{F}) = D\prod_i \left(\exp\left[-\frac{1}{2}(\Delta\mathbf{f}_i - \mathbf{b}_i\mathbf{X})\Sigma_i^{-1}(\Delta\mathbf{f}_i - \mathbf{b}_i\mathbf{X})' \right] \right),$$

where D is a constant, $\mathbf{b}_i$ is the ith row vector of $\mathbf{A}^{-1}$ and Σ_i is a $T \times T$ matrix with $\sigma_i^2 \mathbf{b}_i \mathbf{b}_i'$ on the diagonal elements. The posterior mean $p(\Delta\mathbf{f}_i \mid \mathbf{A}, \mathbf{x}_i, \sigma_i^2\mathbf{I})$ is obtained as:

$$\widehat{\Delta\mathbf{f}_i} = E(\Delta\mathbf{f}_i \mid \mathbf{A}, \mathbf{x}_i, \sigma_i^2\mathbf{I}) \approx \Delta\mathbf{f}_i^0 + \sigma_i^2 h(\Delta\mathbf{f}_i^0)(\mathbf{A}'\mathbf{A})^{-1} = \Delta\mathbf{f}_i^0 + \sigma_i^2 h(\Delta\mathbf{f}_i^0) \ . \qquad \text{(A4)}$$

where $\Delta\mathbf{f}_i^0 = b_i\mathbf{X}$ is the noisy estimates of the ith source, which is the mode of the likelihood and $h(\Delta\mathbf{f}_i) = \dfrac{\partial \log p(\Delta\mathbf{f}_i)}{\partial(\Delta\mathbf{f}_i)}$. The matrix $\mathbf{A}$ is orthogonal after pre-whitening implying that $\mathbf{A}'\mathbf{A} = \mathbf{I}$. Equation (A4) indicates that the expectation $E(\Delta\mathbf{f}_i \mid \mathbf{A}, \mathbf{x}_i, \sigma_i^2\mathbf{I})$ is a function of $b_i\mathbf{X}$, i.e. the posterior depends on data only. The expectation can be approximated by

$\mathrm{E}_{q(\mathbf{F})}\left[\mathbf{F} \mid \mathbf{X}, \mathbf{A}\right] = \mathbf{F}^0 + \mathbf{\Psi}\left(\frac{\partial \log p(\mathbf{F})}{\partial \mathbf{F}}\right)_{\mathbf{F}^0}$, $\mathbf{\Psi} = \mathrm{diag}(\sigma_1^2, \sigma_2^2, ..., \sigma_K^2)$. By substituting the above approximations into (A3), we obtain:

$$\hat{\mathbf{A}} = \mathbf{X}\hat{\mathbf{F}}'/T \approx \mathbf{X}\mathbf{F}^{0\prime}/T + \mathbf{X}\mathbf{H}(\mathbf{F}^{0\prime})\mathbf{\Psi}'/T,$$
$$= \mathbf{A} + \mathbf{X}\mathbf{H}(\mathbf{F}^{0\prime})\mathbf{\Psi}'/T.$$

Särelä and Valpola (2005) and Valpola and Pajunen (2004) point out that it is not necessary to approximate: $\widehat{\Delta \mathbf{f}}_i \approx \Delta \mathbf{f}_i^0 + \sigma_i^2 h(\Delta \mathbf{f}_i^0)$ in the E-step. It can be any method that uses $\Delta \mathbf{f}_i^0$ to compute $\widehat{\Delta \mathbf{f}_i}$: $\widehat{\Delta \mathbf{f}_i} = \boldsymbol{\rho}(\Delta \mathbf{f}_i^0)$. Särelä and Valpola (2005) called $\boldsymbol{\rho}(\Delta \mathbf{f}_i^0)$ the de-noising function. The de-noising here is to remove the noise from other sources and the Gaussian noise in the data. The de-noising should reflect the prior knowledge of the source characteristics, and the choice of the de-noising function is implicitly connected to the distribution of sources.

If we assume that sources have lower frequencies than noise and hence set $\boldsymbol{\rho}(\Delta \mathbf{f}_i)$ to be low-pass filtering, we will end up with sources that have the most low-frequency components. On the other hand, if the de-noising function is chosen to be a shrinkage function, it then leaves large components of sources relatively untouched but removes small components. In such a case, the extracted sources have heavy tails and thus a super-Gaussian distribution (Valpola and Särelä, 2004). In other words, we can regard the E-step as filtering away the noise by exploiting the prior distribution of sources, and the de-noising function can but need not be based on the E-step derived from a generative model.

In practice, de-noising functions can easily be designed without explicitly starting from objective functions, and good de-noising results in a fast and accurate separation algorithm (Valpola and Särelä, 2004). Särelä and Valpola (2005) has shown that the de-noising function is linear if the source signals are Gaussian. Furthermore, the kurtosis-based FastICA algorithm is a special case of DSS with a nonlinear de-noising: $\boldsymbol{\rho}(\Delta \mathbf{f}_i) = \Delta \mathbf{f}_i^3$. Brunnermeier and Pedersen (2009) point out that significant skewness appears in exchange rate changes, which is also confirmed in Table 1. We therefore adopt skewness to measure the distance to normality. The objective function is $g(\Delta f_i)$=$[\sum_{t=1}^{T} \Delta f_{it}^3/T]/[\sum_{t=1}^{T} \Delta f_{it}^2/T]^{3/2}$=$[\sum_{t=1}^{T} \Delta f_{it}^3/T]$, since the source variance has been fixed to unity. Using the method of gradient ascend, we can derive the function $\boldsymbol{\rho}(\Delta\mathbf{f}_i)$ from $g(\Delta\mathbf{f}_i)$. This yields $\nabla_{\mathbf{f}} g(\Delta\mathbf{f}_i) = (3/T)\Delta\mathbf{f}_i^2$, where $\Delta\mathbf{f}_i^2 = [\Delta f_{i1}^2, \Delta f_{i2}^2, ..., \Delta f_{iT}^2]$. Hence, $\boldsymbol{\rho}(\Delta\mathbf{f}_i) = \Delta\mathbf{f}_i^2$.

In the EM algorithm, all sources are estimated simultaneously. However, pre-whitening the data allows us to extract sources one by one (Hyvärinen et al. 2001). The one-unit version of the DSS algorithm is described as follows:

(1) choose an initial weighting vector $\mathbf{w}_i^0$,

(2) $\Delta\mathbf{f}_i^0 = \mathbf{w}_i^{0\prime}\mathbf{X}$,

(3) $\Delta\mathbf{f}_i^1 = \boldsymbol{\rho}(\Delta\mathbf{f}_i^0)$,

(4) $\mathbf{w}_i^1 = \mathbf{X}\Delta\mathbf{f}_i^{1\prime}$,

(5) Let $\mathbf{w}_i^* = \mathbf{w}_i^1 / \left\|\mathbf{w}_i^1\right\|$, where $\left\|\mathbf{w}_i^1\right\|$ is the Euclidean norm of $\mathbf{w}_i^1$.

(6) If not converged, go back to (2) with $\mathbf{w}_i^0$ replaced by $\mathbf{w}_i^*$.

The noisy estimate of the ith source based on the mode of the likelihood is obtained from the second step. The third step corresponds to the expectation of $\Delta\mathbf{f}_i$ over $q(\mathbf{F})$. The de-noising function, $\boldsymbol{\rho}(\Delta\mathbf{f}_i)$, is a row-factor-valued function of a row-vector argument. The prior information of sources is presented in a form of a de-noising function. The fourth step is the re-estimation step which constructs the new ML estimates of the mixing vector. The M-step of the EM algorithm is completed by normalizing the mixing vector, which is the fifth step. The normalization step changes the contributions of the sources by an equal fraction.

The above one-unit DSS algorithm estimates one independent component. To estimate several independent components, we need to run the above algorithm several times with vectors $\mathbf{w}_1, \ldots, \mathbf{w}_K$. To prevent different vectors from converging to the same maxima, it is necessary to orthogonalize the vectors $\mathbf{w}_1, \ldots, \mathbf{w}_K$ after every iteration. Under the assumption that the data are pre-whitened and that the sources are statistically independent and non-Gaussian, the orthonormal basis spanned by the mixing vector corresponds to the fixed point of the DSS algorithm (Särelä and Valpola, 2005).

It is worth noting that the de-noising operation may usually lose information regarding the desired sources. However, it is only required that the operation cancels even more noise than the sources. Even if the de-noising discards parts of the sources or creates nonexistent sources, the re-estimation steps restore them. This is because the estimation steps (4) and (5) constrain the source $\Delta\mathbf{f}_i$ to the subspace spanned by the data. Unless the distorted source is correlated with some of the sources existing in the data set, it has no contribution to $\mathbf{w}_i^1$.

Appendix C: Table Appendix

This appendix is a not-for-publication appendix, and it provides empirical results in the robustness section.

Case 1: Late (1999Q1-2011Q2) and P90 (1990Q1-2011Q2) samples.
Table A1. Out-of-sample forecasts under the Late and P90 samples

Case 2: $\hat{E}^{P}_{i,t}$ and $\hat{E}^{I}_{i,t}$ are constructed based on the source number being determined by the *CPV* and BIC_3 criterions, respectively.
Table A2.1. Out-of-sample forecasts—*CPV*
Table A2.2. Out-of-sample forecasts—BIC_3

Case 3: Forecast accuracy is evaluated by the Max *t*-statistic (adj.) of Hubrich and West (2010).
Table A3. Tests of equal accuracy across IC-nested models

Case 4. The initial estimation period is sixteen and twelve years.
Table A4.1. Out-of-sample forecasts with the initial estimation period being sixteen years
Table A4.2. Out-of-sample forecasts with the initial estimation period being twelve years

Case 5: Removing Korea.
Table A5. Out-of-sample forecasts with Korea being removed

Case 6: Both PCs and ICs are constructed based on the two-step method.
Table A6. Out-of-sample forecasts with both PCs and ICs being constructed from the two-step method

Case 7: Measure non-Gaussianity with kurtosis.
Table A7. Out-of-sample forecasts when non-Gaussianity is measured by kurtosis

Case 8: The benchmark model being the random walk with drift.
Table A8. Out-of-sample forecasts with the benchmark being the random walk with drift

Case 9: The predictability of the UK and JAP sources.
Table A9.1. Out-of-sample forecasts with the UK-based exchange rates
Table A9.2. Out-of-sample forecasts with the JAP-based exchange rates

Case 10. The Rolling forecasting scheme.
Table A10.1: Out-of-sample forecasts with the rolling window being ten years
Table A10.2: Out-of-sample forecasts with the rolling window being fourteen years

Case 1: Late (1999Q1-2010Q2) and P90 (1990Q1-2010Q2) samples.

Table A1. Out-of-sample forecasts under the Late and P90 samples

PC: $e_{i,t+h} - e_{i,t} = c^P_{i,h} + \beta^P_h(\hat{E}^P_{i,t} - e^n_{i,t}) + u^P_{i,t+h}$, i=1,…, N; h=1,…,H.

IC: $e_{i,t+h} - e_{i,t} = c^I_{i,h} + \beta^I_h(\hat{E}^I_{i,t} - e^n_{i,t}) + u^I_{i,t+h}$,

PC+k : $e_{i,t+h} - e_{i,t} = c^P_{i,k,h} + \beta^P_{k,h}(\hat{E}^P_{i,t} - e^n_{i,t}) + \gamma^P_{k,h}(z_{i,k,t} - e_{i,t}) + u^P_{i,k,t+h}$, k =TR, M, PPP.

IC+k: $e_{i,t+h} - e_{i,t} = c^I_{i,k,h} + \beta^I_{k,h}(\hat{E}^I_{i,t} - e^n_{i,t}) + \gamma^I_{k,h}(z_{i,k,t} - e_{i,t}) + u^I_{i,k,t+h}$.

Model	Median *TU* #*TU*<1 or (*CW*<-1.282)				$\overline{TU}$	$\overline{CW1}$	$\overline{CW2}$
	h=1	4	8	12			
			Late/*N*=9				
PC	1.011	1.044	1.083	1.158	0.19	0.14	0.22
	2(0)	1(1)	2(2)	2(2)			
IC	**0.996**	**0.982**	**0.952**	**0.92**	**0.97**	**0.53**	**0.78**
	8(3)	9(2)	9(8)	9(6)			
PC+TR	1.011	1.044	1.090	1.156	0.19	0.14	0.22
	2(0)	1(1)	2(2)	2(2)			
IC+TR	**0.994**	**0.980**	**0.952**	**0.921**	**0.97**	**0.61**	**0.78**
	8(4)	9(4)	9(8)	9(6)			
PC+M	1.018	1.100	1.242	1.413	0.17	0.03	0.06
	1(0)	2(0)	2(0)	1(1)			
IC+M	**0.993**	**0.981**	1.060	1.010	0.47	0.19	0.11
	5(2)	5(3)	3(1)	4(1)			
PC+PPP	1.000	**0.983**	**0.951**	**0.909**	**0.64**	0.39	**0.50**
	4(2)	6(3)	7(5)	6(4)			
IC+PPP	**0.991**	**0.955**	**0.864**	**0.794**	**0.94**	**0.53**	**0.67**
	8(3)	8(4)	9(6)	9(6)			
			P90/*N*=9				
PC	1.009	1.03	1.049	1.052	0.33	0.08	0.11
	4(0)	2(1)	3(1)	3(1)			
IC	**0.998**	**0.987**	**0.958**	**0.922**	**0.94**	**0.53**	**0.56**
	8(3)	8(6)	9(6)	9(4)			
PC+TR	1.008	1.028	1.051	1.052	0.33	0.08	0.11
	4(0)	2(1)	3(1)	3(1)			
IC+TR	**0.994**	**0.983**	**0.956**	**0.920**	**0.94**	**0.50**	**0.67**
	8(3)	8(3)	9(6)	9(6)			
PC+M	1.011	1.017	1.118	1.156	0.36	0.14	0.11
	3(1)	3(2)	4(2)	3(0)			
IC+M	**0.996**	**0.982**	1.014	**0.967**	**0.58**	0.39	0.39
	6(3)	6(4)	4(4)	5(3)			
PC+PPP	**0.998**	**0.984**	**0.955**	**0.92**	**0.69**	**0.69**	**0.89**
	5(3)	6(6)	7(8)	7(8)			
IC+PPP	**0.992**	**0.966**	**0.929**	**0.879**	**0.83**	**0.72**	**0.89**
	6(4)	8(6)	8(8)	8(8)			

Notes: $e_{i,t}$ is the log nominal exchange rate, $\hat{E}^P_{i,t}$ and $\hat{E}^I_{i,t}$ are PC- and IC-based fundamental exchange rates constructed with the source number being determined by the IC_{p2} rule. $z_{i,TR,t}$, $z_{i,M,t}$ and $z_{i,PPP,t}$ are fundamental exchange rates measured by the Taylor-rule (TR), monetary (M) and purchasing power parity (PPP) fundamentals, respectively. PC and IC indicate a PC-based and an IC-based model, respectively. PC(IC)+k indicates that the forecasts are constructed based on the kth fundamental augmented PC-based (IC-based) model. The benchmark is the driftless random walk: $e_{i,t+1} - e_{i,t} = u_{i,t+1}$. The numbers in the first line of each row under columns 2-5 are the medians of the *TU*-statistics, and they appear in bold if they are less than 1.0. *TU* is the ratio of the root mean square errors from a model relative to that from the driftless random walk. The model is superior to the driftless random walk if the *TU*-statistic is less than one. The *CW*-statistic examines if the two models have the same mean square prediction errors, and the hypothesis of equal accuracy is rejected if the *CW*-statistic is less than -1.282. There are two numbers in the second line of each row for each column under columns 2-5. The first number is the number of counties having the *TU*-statistic (across N) less than 1.0, and the second number, in parentheses, denotes the number of currencies for which the *CW*-statistic rejects the hypothesis of equal forecast accuracy at the 10% level. $\overline{TU}$ is the percentage of the *TU*-statistics less than 1.0 over all horizons. $\overline{CW1}$ and $\overline{CW2}$ denote the rejection percentages of the *CW*-statistic over all horizons and over the medium and long horizons, respectively. These percentages are boldfaced if they are greater than or equal to 0.5. Late and P90 refer to the periods 1999-2011 and 1990-2011, respectively.

Case 2: $\hat{E}^{P}_{i,t}$ **and** $\hat{E}^{I}_{i,t}$ **are constructed based on the source number being determined by the *CPV* and *BIC₃* criterions, respectively**.

Table A2.1. Out-of-sample forecasts—***CPV***

$$\text{PC:}\quad e_{i,t+h}-e_{i,t}=c^{P}_{i,h}+\beta^{P}_{h}(\hat{E}^{P}_{i,t}-e^{n}_{i,t})+u^{P}_{i,t+h}\,,\ i=1,\ldots,N;\ h=1,\ldots,H.$$

$$\text{IC:}\quad e_{i,t+h}-e_{i,t}=c^{I}_{i,h}+\beta^{I}_{h}(\hat{E}^{I}_{i,t}-e^{n}_{i,t})+u^{I}_{i,t+h}\,,$$

$$\text{PC+k:}\quad e_{i,t+h}-e_{i,t}=c^{P}_{i,k,h}+\beta^{P}_{k,h}(\hat{E}^{P}_{i,t}-e^{n}_{i,t})+\gamma^{P}_{k,h}(z_{i,k,t}-e_{i,t})+u^{P}_{i,k,t+h}\,,\ k=TR,\ M,\ PPP.$$

$$\text{IC+k:}\quad e_{i,t+h}-e_{i,t}=c^{I}_{i,k,h}+\beta^{I}_{k,h}(\hat{E}^{I}_{i,t}-e^{n}_{i,t})+\gamma^{I}_{k,h}(z_{i,k,t}-e_{i,t})+u^{I}_{i,k,t+h}$$

Model	Median *TU* #*TU*<1 or (*CW*<-1.282)				$\overline{TU}$	$\overline{CW1}$	$\overline{CW2}$
	h=1	4	8	12			
			Early/*N*=17				
PC	1.0150	1.017	1.036	1.158	0.28	0.10	0.12
	4(2)	6(1)	5(2)	4(2)			
IC	**0.999**	**0.991**	**0.950**	**0.965**	**0.69**	0.29	0.44
	11(0)	13(5)	13(8)	10(7)			
PC+TR	1.008	1.017	1.041	1.155	0.28	0.09	0.09
	5(3)	7(0)	4(2)	3(1)			
IC+TR	**0.997**	**0.988**	**0.942**	**0.965**	**0.75**	0.37	**0.50**
	14(2)	13(6)	13(9)	11(8)			
PC+M	1.012	1.017	1.034	1.242	0.32	0.16	0.18
	5(3)	6(2)	6(4)	5(2)			
IC+M	**0.995**	**0.970**	**0.927**	**0.938**	**0.82**	0.46	**0.62**
	14(2)	15(8)	15(11)	12(10)			
PC+PPP	1.010	1.021	1.083	1.310	0.34	0.26	0.32
	5(3)	7(4)	5(6)	6(5)			
IC+PPP	**0.993**	**0.948**	**0.888**	**0.961**	**0.65**	**0.50**	**0.50**
	11(6)	13(11)	11(9)	9(8)			
			Long/*N*=9				
PC	1.022	1.094	1.192	1.309	0.17	0.00	0.00
	2(0)	1(0)	2(0)	1(0)			
IC	1.000	**0.986**	**0.956**	**0.946**	**0.75**	0.36	**0.56**
	4(0)	7(3)	7(5)	9(5)			
PC+TR	1.018	1.084	1.198	1.323	0.17	0.00	0.00
	2(0)	1(0)	2(0)	1(0)			
IC+TR	**0.999**	**0.983**	**0.958**	**0.935**	**0.83**	0.33	**0.56**
	5(0)	7(2)	9(5)	9(5)			
PC+M	1.010	1.066	1.121	1.220	0.17	0.06	0.11
	2(0)	1(0)	2(1)	1(1)			
IC+M	1.002	**0.992**	**0.993**	**0.986**	**0.56**	0.31	0.39
	4(1)	5(3)	5(3)	6(4)			
PC+PPP	1.021	1.083	1.146	1.228	0.22	0.25	0.33
	2(1)	2(2)	2(3)	2(3)			
IC+PPP	**0.991**	**0.982**	**0.939**	**0.922**	**0.72**	**0.67**	**0.83**
	6(3)	7(6)	7(8)	6(7)			

Notes: $\hat{E}^{P}_{i,t}$ and $\hat{E}^{I}_{i,t}$ are constructed based on the source number determined by the criterion of cumulative percentage of total variance (*CPV*). Early and Long refer to the periods 1973-1998 and 1973-2011, respectively. Others are the same as those in Table A1.

Table A2.2. Out-of-sample forecasts—***BIC₃***

$$\text{PC:}\quad e_{i,t+h}-e_{i,t}=c^{P}_{i,h}+\beta^{P}_{h}(\hat{E}^{P}_{i,t}-e^{n}_{i,t})+u^{P}_{i,t+h}\text{ , } i=1,\ldots,N;\ h=1,\ldots,H.$$

$$\text{IC:}\quad e_{i,t+h}-e_{i,t}=c^{I}_{i,h}+\beta^{I}_{h}(\hat{E}^{I}_{i,t}-e^{n}_{i,t})+u^{I}_{i,t+h}\text{ ,}$$

$$\text{PC+k:}\quad e_{i,t+h}-e_{i,t}=c^{P}_{i,k,h}+\beta^{P}_{k,h}(\hat{E}^{P}_{i,t}-e^{n}_{i,t})+\gamma^{P}_{k,h}(z_{i,k,t}-e_{i,t})+u^{P}_{i,k,t+h}\text{ , } k=TR,\ M,\ PPP.$$

$$\text{IC+k:}\quad e_{i,t+h}-e_{i,t}=c^{I}_{i,k,h}+\beta^{I}_{k,h}(\hat{E}^{I}_{i,t}-e^{n}_{i,t})+\gamma^{I}_{k,h}(z_{i,k,t}-e_{i,t})+u^{I}_{i,k,t+h}$$

Model	Median *TU* #*TU*<1 or (*CW*<-1.282)				$\overline{TU}$	$\overline{CW1}$	$\overline{CW2}$
	h=1	4	8	12			
			Early/*N*=17				
PC	1.007	1.047	1.078	1.146	0.25	0.18	0.24
	4(1)	3(3)	4(5)	6(3)			
IC	**0.999**	**0.994**	**0.956**	**0.964**	**0.69**	0.28	0.44
	10(1)	13(3)	14(8)	10(7)			
PC+TR	1.005	1.049	1.077	1.148	0.25	0.19	0.24
	4(2)	3(3)	4(5)	6(3)			
IC+TR	**0.998**	**0.989**	**0.952**	**0.961**	**0.74**	0.32	0.47
	13(4)	12(2)	14(8)	11(8)			
PC+M	1.005	1.028	1.033	1.106	0.26	0.16	0.21
	5(2)	3(2)	4(4)	6(3)			
IC+M	**0.995**	**0.976**	**0.935**	**0.929**	**0.79**	0.46	**0.56**
	12(5)	15(7)	15(9)	12(10)			
PC+PPP	**0.999**	**0.982**	**0.948**	**0.926**	**0.57**	0.43	**0.65**
	9(2)	10(5)	10(10)	10(12)			
IC+PPP	**0.994**	**0.949**	**0.883**	**0.911**	**0.63**	**0.51**	**0.53**
	11(6)	12(11)	11(9)	9(9)			
			Long/*N*=9				
PC	1.013	1.054	1.089	1.126	0.31	0.14	0.17
	4(1)	3(1)	3(1)	1(2)			
IC	1.000	**0.991**	**0.955**	**0.941**	**0.78**	0.47	**0.61**
	4(2)	8(4)	7(6)	9(5)			
PC+TR	1.014	1.064	1.082	1.116	0.31	0.14	0.17
	4(1)	3(1)	3(1)	1(2)			
IC+TR	1.000	**0.988**	**0.958**	**0.936**	**0.81**	0.44	**0.67**
	4(2)	8(2)	8(6)	9(6)			
PC+M	1.007	1.009	1.064	1.090	0.36	0.19	0.28
	4(1)	3(1)	3(3)	3(2)			
IC+M	1.002	**0.997**	1.013	1.001	**0.50**	0.31	0.28
	4(3)	6(3)	4(2)	4(3)			
PC+PPP	1.003	**0.992**	**0.945**	**0.955**	**0.58**	**0.56**	**0.67**
	4(3)	5(5)	6(6)	6(6)			
IC+PPP	**0.991**	**0.972**	**0.927**	**0.910**	**0.72**	**0.72**	**0.83**
	6(5)	7(6)	7(8)	6(7)			

Notes: $\hat{E}^{P}_{i,t}$ and $\hat{E}^{I}_{i,t}$ are constructed based on the source number determined by the criterion of BIC_3. Early and Long refer to the periods 1973-1998 and 1973-2011, respectively. Others are the same as those in Table A1.

Case 3: Forecast accuracy is evaluated by the Max *t*-statistic (adj.) of Hubrich and West (2010).

Table A3. Tests of equal accuracy across IC-nested models

	Early/*N*=17, m=4, R0=56				Long/*N*=9, m=4,R0=56			
	h=1	4	8	12	*h*=1	4	8	12
	P= 48	45	41	37	P= 90	87	83	79
AUT	0.422	**0.002**	**0.001**	**0.001**	0.468	**0.077**	**0.053**	**0.042**
CAN	0.953	0.783	0.355	0.150	0.315	0.206	0.225	0.120
DEN	0.228	**0.095**	**0.002**	0.210	0.258	0.125	**0.077**	0.191
JAP	0.534	0.845	0.701	0.203	0.225	0.394	0.204	0.213
KOR	**0.066**	0.122	**0.074**	**0.011**	**0.057**	**0.097**	**0.044**	**0.019**
NOR	**0.051**	**0.026**	**0.004**	**0.012**	0.111	**0.053**	**0.081**	**0.078**
SWD	0.342	**0.091**	**0.005**	**0.028**	**0.092**	**0.005**	**0.002**	**0.005**
SWZ	**0.013**	**0.011**	**0.001**	**0.000**	**0.075**	**0.031**	**0.017**	**0.029**
UK	0.464	0.276	**0.020**	**0.047**	0.436	0.237	**0.012**	0.188
AUS	**0.097**	**0.047**	**0.005**	0.145	--	--	--	--
BEL	**0.086**	**0.002**	**0.001**	**0.031**	--	--	--	--
FIN	**0.075**	**0.058**	**0.016**	**0.024**	--	--	--	--
FRN	**0.056**	0.101	**0.011**	**0.047**	--	--	--	--
GER	**0.018**	**0.000**	**0.008**	**0.059**	--	--	--	--
ITA	0.345	0.178	0.144	0.169	--	--	--	--
NET	**0.012**	**0.000**	**0.004**	**0.074**	--	--	--	--
SPN	0.411	0.298	0.265	0.286	--	--	--	--

Notes: *h*, P, R0, *N* and m are forecast horizons, the number of out-of-sample forecasts, the number of in-sample observations initially, the number of countries in the panel and the number of other models nesting the random walk, respectively. AUT, CAN, DEN, JAP, KOR, NOR, SWD, SWZ, UK, AUS, BEL, FIN, FRN, GER, ITA, NET and SPN indicate Australia, Canada, Denmark, Japan, Korea, Norway, Sweden, Switzerland, the United Kingdom, Austria, Belgium, Finland, France, Germany, Italy, Netherland and Spain. The numbers in the Table are p-values of the Max *t*-statistic (adj.). Critical values of the Max *t*-statistic (adj.) are constructed based on 1,000 bootstraps for constructing the sample correlation matrix between the adjusted difference in MSPEs across models and the random walk, and on 10,000 Monte-Carlo simulations for obtaining critical values of the test statistic. Values in bold denote significance at the 10% level. '- -' indicates that no statistic is constructed. Due to data availability, the sample period varies with fundamentals. We construct the Max *t*-statistic (adj.) with a balanced panel; hence, the Early period starts from 1973Q1 and ends in 1998Q4, but the Long period starts from 1973Q1 and ends in 2009Q2.

Case 4. The initial estimation period is sixteen and twelve years, respectively.

Table A4.1. Out-of-sample forecasts with the initial estimation period being sixteen years

PC: $e_{i,t+h} - e_{i,t} = c^P_{i,h} + \beta^P_h(\hat{E}^P_{i,t} - e^n_{i,t}) + u^P_{i,t+h}$, i=1,…, N; h=1,…,H.

IC: $e_{i,t+h} - e_{i,t} = c^I_{i,h} + \beta^I_h(\hat{E}^I_{i,t} - e^n_{i,t}) + u^I_{i,t+h}$,

PC+k : $e_{i,t+h} - e_{i,t} = c^P_{i,k,h} + \beta^P_{k,h}(\hat{E}^P_{i,t} - e^n_{i,t}) + \gamma^P_{k,h}(z_{i,k,t} - e_{i,t}) + u^P_{i,k,t+h}$, k =TR, M, PPP.

IC+k: $e_{i,t+h} - e_{i,t} = c^I_{i,k,h} + \beta^I_{k,h}(\hat{E}^I_{i,t} - e^n_{i,t}) + \gamma^I_{k,h}(z_{i,k,t} - e_{i,t}) + u^I_{i,k,t+h}$

Model	Median *TU* #*TU*<1 or (*CW*<-1.282)				$\overline{TU}$	$\overline{CW1}$	$\overline{CW2}$
	h=1	4	8	12			
			Early/*N*=17				
PC	1.002	1.01	1.006	0.952	0.46	0.29	0.38
	8(3)	6(4)	8(7)	9(6)			
IC	**0.998**	**0.993**	**0.954**	**0.913**	**0.85**	0.43	**0.68**
	15(2)	13(4)	15(8)	15(15)			
PC+TR	**0.999**	1.008	1.001	**0.953**	0.49	0.32	0.44
	9(3)	7(4)	8(8)	9(7)			
IC+TR	**0.994**	**0.985**	**0.944**	**0.905**	**0.87**	0.46	0.71
	15(4)	14(3)	15(9)	15(15)			
PC+M	**0.998**	**0.996**	1.019	**0.979**	**0.53**	0.32	0.44
	9(3)	9(4)	8(8)	10(7)			
IC+M	**0.994**	**0.973**	**0.935**	**0.882**	**0.90**	**0.51**	**0.68**
	15(5)	16(7)	15(11)	15(12)			
PC+PPP	**0.995**	**0.987**	**0.930**	**0.890**	**0.69**	**0.51**	**0.71**
	12(4)	10(7)	12(11)	13(13)			
IC+PPP	**0.986**	**0.939**	**0.870**	**0.808**	**0.79**	**0.57**	**0.68**
	16(6)	14(10)	13(11)	11(12)			
			Long/*N*=9				
PC	1.005	1.026	1.049	1.063	0.33	0.08	0.11
	3(0)	3(1)	3(1)	3(1)			
IC	**0.998**	**0.987**	**0.961**	**0.931**	**0.89**	**0.50**	**0.56**
	7(3)	8(5)	8(6)	9(4)			
PC+TR	1.005	1.031	1.052	1.058	0.33	0.08	0.11
	3(0)	3(1)	3(1)	3(1)			
IC+TR	**0.994**	**0.984**	**0.959**	**0.929**	**0.92**	**0.56**	**0.61**
	7(4)	8(5)	9(6)	9(5)			
PC+M	1.012	1.018	1.118	1.156	0.33	0.19	0.22
	3(1)	3(2)	4(3)	2(1)			
IC+M	**0.997**	**0.992**	1.043	**0.983**	**0.56**	0.36	0.33
	6(3)	5(4)	4(3)	5(3)			
PC+PPP	**0.999**	**0.978**	**0.965**	**0.955**	**0.67**	**0.67**	**0.89**
	5(3)	6(5)	6(8)	7(8)			
IC+PPP	**0.990**	**0.980**	**0.929**	**0.895**	**0.78**	**0.72**	**0.89**
	6(4)	7(6)	7(8)	8(8)			

Notes: The initial estimation window is sixteen years. Early and Long refer to the periods 1973-1998 and 1973-2011, respectively. Others are the same as those in Table A1.

Table A4.2. Out-of-sample forecasts with the initial estimation period being twelve years

$$\text{PC:}\quad e_{i,t+h}-e_{i,t}=c^{P}_{i,h}+\beta^{P}_{h}(\hat{E}^{P}_{i,t}-e^{n}_{i,t})+u^{P}_{i,t+h}\text{ , } i=1,\ldots,N;\ h=1,\ldots,H.$$

$$\text{IC:}\quad e_{i,t+h}-e_{i,t}=c^{I}_{i,h}+\beta^{I}_{h}(\hat{E}^{I}_{i,t}-e^{n}_{i,t})+u^{I}_{i,t+h}\text{ ,}$$

$$\text{PC+k :}\quad e_{i,t+h}-e_{i,t}=c^{P}_{i,k,h}+\beta^{P}_{k,h}(\hat{E}^{P}_{i,t}-e^{n}_{i,t})+\gamma^{P}_{k,h}(z_{i,k,t}-e_{i,t})+u^{P}_{i,k,t+h}\text{ , } k=TR,\ M,\ PPP.$$

$$\text{IC+k:}\quad e_{i,t+h}-e_{i,t}=c^{I}_{i,k,h}+\beta^{I}_{k,h}(\hat{E}^{I}_{i,t}-e^{n}_{i,t})+\gamma^{I}_{k,h}(z_{i,k,t}-e_{i,t})+u^{I}_{i,k,t+h}\text{ .}$$

Model	Median *TU* #*TU*<1 or (*CW*<-1.282)				$\overline{TU}$	$\overline{CW1}$	$\overline{CW2}$
	h=1	4	8	12			
			Early/*N*=17				
PC	1.033	1.098	1.182	1.223	0.15	0.07	0.09
	4(2)	3(0)	1(1)	2(2)			
IC	1.002	**0.981**	**0.954**	**0.907**	**0.65**	0.28	0.44
	7(3)	11(1)	12(7)	14(8)			
PC+TR	1.029	1.095	1.179	1.225	0.16	0.07	0.09
	5(2)	3(0)	1(1)	2(2)			
IC+TR	1.000	**0.979**	**0.954**	**0.912**	**0.66**	0.34	0.47
	8(5)	11(2)	12(8)	14(8)			
PC+M	1.027	1.128	1.161	1.178	0.19	0.12	0.12
	5(2)	5(2)	1(2)	2(2)			
IC+M	**0.997**	**0.974**	**0.947**	**0.895**	**0.72**	0.46	**0.53**
	10(6)	12(7)	13(8)	14(10)			
PC+PPP	1.020	1.070	1.010	0.927	0.44	0.37	0.41
	6(5)	6(6)	8(6)	10(8)			
IC+PPP	**0.990**	**0.930**	**0.867**	**0.796**	**0.72**	**0.62**	**0.71**
	11(7)	12(11)	12(13)	14(11)			
			Long/*N*=9				
PC	1.02	1.07	1.117	1.171	0.11	0.08	0.11
	1(1)	1(0)	1(1)	1(1)			
IC	1.002	1.007	**0.984**	**0.940**	**0.53**	0.39	**0.67**
	2(1)	4(1)	6(5)	7(7)			
PC+TR	1.019	1.071	1.106	1.163	0.11	0.08	0.11
	1(1)	1(0)	1(1)	1(1)			
IC+TR	1.002	1.008	**0.983**	**0.935**	**0.56**	0.36	**0.61**
	3(1)	4(1)	6(4)	7(7)			
PC+M	1.010	1.055	1.127	1.274	0.25	0.08	0.11
	1(1)	2(0)	3(1)	3(1)			
IC+M	1.006	1.020	1.038	1.075	0.31	0.28	0.33
	3(2)	2(2)	2(2)	4(4)			
PC+PPP	1.004	1.038	**0.979**	**0.926**	**0.53**	0.47	**0.67**
	3(2)	4(3)	6(6)	6(6)			
IC+PPP	**0.992**	**0.965**	**0.948**	**0.933**	**0.72**	**0.78**	**0.94**
	6(5)	5(6)	7(8)	8(9)			

Notes: The initial estimation window is twelve years. Early and Long refer to the periods 1973-1998 and 1973-2011, respectively. Others are the same as those in Table A1.

Case 5: Removing Korea.

Table A5. Out-of-sample forecasts with Korea being removed

$$\text{PC:}\quad e_{i,t+h}-e_{i,t}=c^{P}_{i,h}+\beta^{P}_{h}(\hat{E}^{P}_{i,t}-e^{n}_{i,t})+u^{P}_{i,t+h}\ ,\ i=1,\ldots,N;\ h=1,\ldots,H.$$

$$\text{IC:}\quad e_{i,t+h}-e_{i,t}=c^{I}_{i,h}+\beta^{I}_{h}(\hat{E}^{I}_{i,t}-e^{n}_{i,t})+u^{I}_{i,t+h}\ ,$$

$$\text{PC+k:}\quad e_{i,t+h}-e_{i,t}=c^{P}_{i,k,h}+\beta^{P}_{k,h}(\hat{E}^{P}_{i,t}-e^{n}_{i,t})+\gamma^{P}_{k,h}(z_{i,k,t}-e_{i,t})+u^{P}_{i,k,t+h}\ ,\ k=TR,\ M,\ PPP.$$

$$\text{IC+k:}\quad e_{i,t+h}-e_{i,t}=c^{I}_{i,k,h}+\beta^{I}_{k,h}(\hat{E}^{I}_{i,t}-e^{n}_{i,t})+\gamma^{I}_{k,h}(z_{i,k,t}-e_{i,t})+u^{I}_{i,k,t+h}$$

Model	Median *TU* #*TU*<1 or (*CW*<-1.282)				$\overline{TU}$	$\overline{CW1}$	$\overline{CW2}$
	h=1	4	8	12			
			Early/*N*=16				
PC	1.007	1.04	1.087	1.113	0.23	0.08	0.09
	4(1)	3(1)	5(3)	3(0)			
IC	1.003	1.000	**0.978**	**0.989**	0.47	0.19	0.38
	0(0)	8(0)	13(6)	9(6)			
PC+TR	1.006	1.043	1.082	1.113	0.23	0.08	0.09
	4(1)	3(1)	5(3)	3(0)			
IC+TR	1.002	1.001	**0.971**	**0.980**	0.47	0.20	0.41
	3(0)	5(0)	13(6)	9(7)			
PC+M	1.009	1.011	1.043	1.064	0.31	0.13	0.16
	5(1)	5(2)	6(4)	4(1)			
IC+M	**0.999**	**0.985**	**0.949**	**0.956**	**0.73**	0.31	0.47
	10(0)	12(5)	13(8)	12(7)			
PC+PPP	1.005	1.012	**0.965**	**0.962**	**0.50**	0.38	**0.50**
	7(2)	6(6)	10(9)	9(7)			
IC+PPP	**0.999**	**0.960**	**0.929**	**0.953**	**0.56**	0.47	**0.50**
	8(5)	11(9)	9(8)	8(8)			
			Long/*N*=8				
PC	1.011	1.035	1.062	1.107	0.16	0.13	0.19
	0(0)	2(1)	2(2)	1(1)			
IC	1.001	1.004	1.016	**0.973**	0.47	0.19	0.31
	3(0)	3(1)	3(3)	6(2)			
PC+TR	1.012	1.042	1.056	1.099	0.13	0.13	0.19
	0(0)	1(1)	2(2)	1(1)			
IC+TR	1.002	1.006	1.009	**0.967**	0.47	0.19	0.31
	3(0)	3(1)	3(3)	6(2)			
PC+M	1.016	1.053	1.171	1.266	0.22	0.09	0.13
	1(0)	2(1)	3(1)	1(1)			
IC+M	1.001	1.021	1.031	1.017	0.44	0.25	0.31
	4(0)	3(3)	3(2)	4(3)			
PC+PPP	1.002	**0.998**	**0.981**	**0.977**	**0.53**	**0.50**	**0.81**
	3(0)	4(3)	5(7)	5(6)			
IC+PPP	**0.996**	**0.987**	**0.962**	**0.953**	**0.63**	**0.56**	**0.75**
	5(2)	5(4)	5(6)	5(6)			

Notes: Korea is removed from the panel. The panel size is sixteen for the Early sample and eight for the Long sample. Early and Long refer to the periods 1973-1998 and 1973-2011, respectively. Others are the same as those in Table A1.

Case 6: Both PCs and ICs are constructed based on the two-step method.

Table A6. Out-of-sample forecasts with both PCs and ICs being constructed from the two-step method

PC: $e_{i,t+h}-e_{i,t}=c^{P}_{i,h}+\beta^{P}_{h}(\hat{E}^{P}_{i,t}-e^{n}_{i,t})+u^{P}_{i,t+h}$, i=1,…, N; h=1,…,H.

IC: $e_{i,t+h}-e_{i,t}=c^{I}_{i,h}+\beta^{I}_{h}(\hat{E}^{I}_{i,t}-e^{n}_{i,t})+u^{I}_{i,t+h}$,

PC+k : $e_{i,t+h}-e_{i,t}=c^{P}_{i,k,h}+\beta^{P}_{k,h}(\hat{E}^{P}_{i,t}-e^{n}_{i,t})+\gamma^{P}_{k,h}(z_{i,k,t}-e_{i,t})+u^{P}_{i,k,t+h}$, k =*TR, M, PPP*.

IC+k: $e_{i,t+h}-e_{i,t}=c^{I}_{i,k,h}+\beta^{I}_{k,h}(\hat{E}^{I}_{i,t}-e^{n}_{i,t})+\gamma^{I}_{k,h}(z_{i,k,t}-e_{i,t})+u^{I}_{i,k,t+h}$

Model	Median *TU* #*TU*<1 or (*CW*<-1.282)				$\overline{TU}$	$\overline{CW1}$	$\overline{CW2}$
	h=1	4	8	12			
			Early/*N*=17				
PC	1.006	1.014	1.015	1.075	0.22	0.06	0.12
	1(0)	1(0)	6(2)	7(2)			
IC	1.000	**0.999**	**0.976**	**0.976**	**0.59**	0.24	0.47
	7(0)	9(0)	14(8)	10(8)			
PC+TR	1.004	1.010	1.011	1.075	0.22	0.09	0.18
	1(0)	2(0)	5(3)	7(3)			
IC+TR	**0.999**	**0.998**	**0.971**	**0.975**	**0.66**	0.28	0.47
	12(3)	9(0)	14(8)	10(8)			
PC+M	1.002	**0.988**	**0.968**	**0.983**	**0.57**	0.26	0.44
	7(0)	11(3)	12(8)	9(7)			
IC+M	**0.996**	**0.985**	**0.934**	**0.945**	**0.76**	0.37	**0.50**
	11(2)	14(6)	15(9)	12(8)			
PC+PPP	**0.998**	**0.974**	**0.920**	**0.980**	**0.56**	0.40	0.44
	9(4)	11(8)	9(9)	9(6)			
IC+PPP	**0.995**	**0.953**	**0.882**	**0.898**	**0.63**	**0.51**	**0.53**
	11(6)	12(11)	11(9)	9(9)			
			Long/*N*=9				
PC	1.005	1.015	**0.997**	**0.969**	**0.53**	0.14	0.28
	4(0)	4(0)	5(2)	6(3)			
IC	1.001	1.002	**0.974**	**0.951**	**0.61**	0.33	0.44
	3(2)	3(2)	7(5)	9(3)			
PC+TR	1.004	1.015	1.001	0.972	0.47	0.11	0.22
	3(0)	4(0)	4(2)	6(2)			
IC+TR	1.001	1.001	**0.974**	**0.947**	**0.64**	0.31	**0.50**
	3(1)	4(1)	7(5)	9(4)			
PC+M	1.006	1.032	1.051	1.080	0.19	0.08	0.11
	2(0)	2(1)	1(1)	2(1)			
IC+M	1.004	1.014	1.047	1.025	0.31	0.31	0.28
	3(3)	3(3)	3(3)	2(2)			
PC+PPP	**0.993**	**0.987**	**0.940**	**0.922**	**0.67**	**0.61**	**0.83**
	5(3)	5(4)	7(8)	7(7)			
IC+PPP	**0.991**	**0.980**	**0.934**	**0.916**	**0.69**	**0.69**	**0.83**
	6(5)	6(5)	7(8)	6(7)			

Notes: The PCs and ICs are constructed based on the two-step method proposed by Bai and Ng (2002). Early and Long refer to the periods 1973-1998 and 1973-2011, respectively. Others are the same as those in Table A1.

Case 7: Measure non-Gaussianity with kurtosis.

Table A7. Out-of-sample forecasts when non-Gaussianity is measured by kurtosis

PC: $e_{i,t+h} - e_{i,t} = c^{P}_{i,h} + \beta^{P}_{h}(\hat{E}^{P}_{i,t} - e^{n}_{i,t}) + u^{P}_{i,t+h}$, *i*=1,…,*N; h*=1,…,*H.*

IC : $e_{i,t+h} - e_{i,t} = c^{I}_{i,h} + \beta^{I}_{h}(\hat{E}^{I}_{i,t} - e^{n}_{i,t}) + u^{I}_{i,t+h}$,

PC+k: $e_{i,t+h} - e_{i,t} = c^{P}_{i,k,h} + \beta^{P}_{k,h}(\hat{E}^{P}_{i,t} - e^{n}_{i,t}) + \gamma^{P}_{k,h}(z_{i,k,t} - e_{i,t}) + u^{P}_{i,k,t+h}$, *k* =*TR, M, PPP*;.

IC+k: $e_{i,t+h} - e_{i,t} = c^{I}_{i,k,h} + \beta^{I}_{k,h}(\hat{E}^{I}_{i,t} - e^{n}_{i,t}) + \gamma^{I}_{k,h}(z_{i,k,t} - e_{i,t}) + u^{I}_{i,k,t+h}$,

Model	Median *TU* #*TU*<1 or (*CW*<-1.282)				$\overline{TU}$	$\overline{CW1}$	$\overline{CW2}$
	h=1	4	8	12			
			Early/*N*=17				
PCM	1.009	1.042	1.070	1.141	0.26	0.09	0.12
	4(1)	5(1)	6(3)	3(1)			
ICM	1.003	1.004	1.007	**0982**	0.32	0.18	0.32
	1(0)	5(1)	7(4)	9(7)			
PC+TR	1.008	1.042	1.066	1.134	0.25	0.09	0.09
	3(2)	5(1)	6(3)	3(0)			
IC+TR	1.001	1.006	1.002	0.980	0.46	0.18	0.32
	6(1)	8(0)	8(4)	9(7)			
PC+M	1.007	1.032	1.055	1.058	0.32	0.18	0.21
	5(2)	5(3)	6(5)	6(2)			
IC+M	**0.999**	**0.992**	**0.977**	**0.958**	**0.57**	0.25	0.35
	9(1)	10(4)	9(6)	11(6)			
PC+PPP	1.008	1.005	**0.953**	**0.981**	**0.51**	0.43	**0.56**
	8(2)	8(8)	10(11)	9(8)			
IC+PPP	**0.995**	**0.957**	**0.924**	1.009	**0.59**	0.38	0.41
	11(4)	12(8)	9(9)	8(5)			
			Long/*N*=9				
PCM	1.015	1.051	1.059	1.103	0.22	0.08	0.11
	3(0)	2(1)	2(1)	1(1)			
ICM	1.000	**0.995**	**0.973**	**0.959**	**0.67**	0.17	0.17
	4(2)	6(1)	6(2)	8(1)			
PC+TR	1.014	1.059	1.061	1.095	0.22	0.08	0.11
	3(0)	2(1)	2(1)	1(1)			
IC+TR	1.001	**0.995**	**0.971**	**0.953**	**0.67**	0.19	0.22
	3(2)	7(1)	6(2)	8(2)			
PC+M	1.011	1.032	1.113	1.150	0.19	0.08	0.11
	1(0)	1(1)	3(1)	2(1)			
IC+M	1.005	1.036	1.047	1.065	0.25	0.22	0.22
	3(1)	3(3)	2(2)	1(2)			
PC+PPP	1.003	**0.991**	**0.962**	**0.969**	**0.56**	**0.53**	**0.78**
	4(1)	5(4)	5(8)	6(6)			
IC+PPP	**0.990**	**0.987**	**0.944**	**0.921**	**0.69**	**0.67**	**0.83**
	6(4)	6(5)	7(8)	6(7)			

Notes: Non-Gaussianity is measured by kurtosis. Early and Long refer to the periods 1973-1998 and 1973-2011, respectively. Others are the same as those in Table A1.

Case 8: The benchmark model is the random walk with drift.

Table A8: Out-of-sample forecasts with the benchmark being the random walk with drift

$$\text{PC: } e_{i,t+h} - e_{i,t} = c^P_{i,h} + \beta^P_h(\hat{E}^P_{i,t} - e^n_{i,t}) + u^P_{i,t+h},\ i=1,\ldots,N;\ h=1,\ldots,H.$$

$$\text{IC: } e_{i,t+h} - e_{i,t} = c^I_{i,h} + \beta^I_h(\hat{E}^I_{i,t} - e^n_{i,t}) + u^I_{i,t+h},$$

$$\text{PC+k: } e_{i,t+h} - e_{i,t} = c^P_{i,k,h} + \beta^P_{k,h}(\hat{E}^P_{i,t} - e^n_{i,t}) + \gamma^P_{k,h}(z_{i,k,t} - e_{i,t}) + u^P_{i,k,t+h},\ k = TR, M, PPP.$$

$$\text{IC+k: } e_{i,t+h} - e_{i,t} = c^I_{i,k,h} + \beta^I_{k,h}(\hat{E}^I_{i,t} - e^n_{i,t}) + \gamma^I_{k,h}(z_{i,k,t} - e_{i,t}) + u^I_{i,k,t+h}$$

Model	Median *TU* #*TU*<1 or (*CW*<-1.282)				$\overline{TU}$	$\overline{CW1}$	$\overline{CW2}$
	h=1	4	8	12			
			Early/*N*=17				
PC	**0.998**	**0.974**	**0.920**	**0.960**	**0.68**	0.40	0.44
	10(3)	11(9)	15(9)	10(6)			
IC	**0.988**	**0.939**	**0.843**	**0.856**	**0.97**	**0.84**	**0.88**
	16(11)	16(16)	17(15)	17(15)			
PC+TR	**0.996**	**0.977**	**0.918**	**0.961**	**0.69**	0.37	0.44
	11(3)	11(7)	15(10)	10(5)			
IC+TR	**0.9878**	**0.936**	**0.840**	**0.853**	**0.97**	**0.87**	**0.88**
	16(14)	16(15)	17(15)	17(15)			
PC+M	**0.996**	**0.959**	**0.901**	**0.902**	**0.76**	**0.53**	**0.68**
	12(3)	14(10)	14(12)	12(11)			
IC+M	**0.985**	**0.921**	**0.827**	**0.826**	**0.97**	**0.85**	**0.91**
	16(11)	16(16)	17(16)	17(15)			
PC+PPP	**0.996**	**0.948**	**0.835**	**0.803**	**0.79**	**0.59**	**0.65**
	14(7)	16(11)	13(11)	11(11)			
IC+PPP	**0.977**	**0.904**	**0.800**	**0.793**	**0.85**	**0.78**	**0.82**
	16(12)	15(13)	15(14)	12(14)			
			Long/*N*=9				
PC	1.001	**0.979**	**0.928**	**0.916**	**0.67**	**0.56**	**0.61**
	4(4)	6(5)	7(6)	7(5)			
IC	**0.988**	**0.947**	**0.862**	**0.815**	**1.00**	**0.97**	**1.00**
	9(8)	9(9)	9(9)	9(9)			
PC+TR	1.000	**0.985**	**0.929**	**0.920**	**0.67**	**0.58**	**0.67**
	4(4)	6(5)	7(6)	7(6)			
IC+TR	**0.984**	**0.940**	**0.858**	**0.814**	**1.00**	**0.97**	**1.00**
	9(8)	9(9)	9(9)	9(9)			
PC+M	**0.999**	**0.983**	**0.9922**	1.007	**0.53**	0.33	0.33
	5(2)	5(4)	5(3)	4(3)			
IC+M	**0.987**	**0.943**	**0.923**	**0.894**	**0.86**	**0.69**	**0.83**
	8(4)	8(6)	8(8)	7(7)			
PC+PPP	**0.991**	**0.938**	**0.853**	**0.835**	**0.75**	**0.81**	**0.89**
	6(5)	7(8)	7(8)	7(8)			
IC+PPP	**0.982**	**0.913**	**0.840**	**0.781**	**1.00**	**0.94**	**1.00**
	9(7)	9(9)	9(9)	9(9)			

Notes: The benchmark model is the random walk with drift. Early and Long refer to the periods 1973-1998 and 1973-2011, respectively. Others are the same as those in Table A1.

Case 9: The predictability of the UK and JAP sources.

Table A9.1. Out-of-sample forecasts with the UK-based exchange rates

$$\text{PC:}\quad e_{i,t+h} - e_{i,t} = c^P_{i,h} + \beta^P_h(\hat{E}^P_{i,t} - e^n_{i,t}) + u^P_{i,t+h}\text{ , } i\text{=1,…, } N\text{; } h\text{=1,…,}H.$$

$$\text{IC:}\quad e_{i,t+h} - e_{i,t} = c^I_{i,h} + \beta^I_h(\hat{E}^I_{i,t} - e^n_{i,t}) + u^I_{i,t+h}\text{ ,}$$

$$\text{PC+k :}\quad e_{i,t+h} - e_{i,t} = c^P_{i,k,h} + \beta^P_{k,h}(\hat{E}^P_{i,t} - e^n_{i,t}) + \gamma^P_{k,h}(z_{i,k,t} - e_{i,t}) + u^P_{i,k,t+h}\text{ , } k\text{ =}TR, M, PPP.$$

$$\text{IC+k:}\quad e_{i,t+h} - e_{i,t} = c^I_{i,k,h} + \beta^I_{k,h}(\hat{E}^I_{i,t} - e^n_{i,t}) + \gamma^I_{k,h}(z_{i,k,t} - e_{i,t}) + u^I_{i,k,t+h}$$

Model	Median *TU* #*TU*<1 or (*CW*<-1.282)				$\overline{TU}$	$\overline{CW1}$	$\overline{CW2}$
	h=1	4	8	12			
			Early/*N*=17				
PC	1.025	1.096	1.096	1.193	0.19	0.18	0.29
	4(1)	2(1)	2(2)	5(8)			
IC	1.000	1.007	**0.978**	**0.993**	**0.50**	0.32	0.44
	8(2)	8(5)	9(8)	9(7)			
PC+TR	1.028	1.084	1.100	1.199	0.18	0.18	0.29
	4(1)	2(1)	2(2)	4(8)			
IC+TR	**0.999**	**0.998**	**0.981**	1.005	**0.51**	0.34	0.41
	9(2)	9(7)	9(7)	8(7)			
PC+M	1.036	1.139	1.196	1.329	0.15	0.15	0.26
	4(1)	2(0)	2(2)	2(7)			
IC+M	1.009	1.018	**0.966**	**0.859**	**0.56**	0.35	**0.62**
	3(0)	7(3)	15(8)	13(13)			
PC+PPP	1.029	1.106	1.091	1.291	0.29	0.26	0.38
	6(2)	5(3)	5(5)	4(8)			
IC+PPP	**0.992**	**0.972**	**0.900**	**0.884**	**0.68**	**0.56**	**0.76**
	9(3)	13(9)	13(14)	11(12)			
			Long/*N*=9				
PC	1.014	1.071	1.078	1.076	0.25	0.22	0.28
	3(1)	3(2)	2(3)	1(2)			
IC	1.047	1.002	1.003	1.012	0.44	0.31	0.39
	4(1)	4(3)	4(4)	4(3)			
PC+TR	1.015	1.068	1.080	1.053	0.22	0.22	0.28
	2(1)	3(2)	2(3)	1(2)			
IC+TR	1.001	1.004	1.015	1.04	0.44	0.31	0.39
	4(1)	4(3)	4(4)	4(3)			
PC+M	1.039	1.266	1.747	2.262	0.08	0.33	0.44
	2(1)	1(3)	0(4)	0(4)			
IC+M	1.039	1.15	1.513	2.253	0.14	0.47	**0.61**
	2(1)	2(5)	1(6)	0(5)			
PC+PPP	1.027	1.108	1.147	1.206	0.33	**0.53**	**0.61**
	3(3)	3(5)	3(5)	3(6)			
IC+PPP	1.007	1.015	**0.99**	1.056	0.47	**0.61**	**0.72**
	4(2)	4(7)	5(7)	4(6)			

Notes: The base currency is the British pound. Early and Long refer to the periods 1973-1998 and 1973-2011, respectively. Others are the same as those in Table A1.

Table A9.2. Out-of-sample forecasts with the JAP-based exchange rates

$$\text{PC:}\quad e_{i,t+h}-e_{i,t}=c^{P}_{i,h}+\beta^{P}_{h}(\hat{E}^{P}_{i,t}-e^{n}_{i,t})+u^{P}_{i,t+h}\text{ , } i\text{=1,…, }N\text{; }h\text{=1,…,}H.$$

$$\text{IC:}\quad e_{i,t+h}-e_{i,t}=c^{I}_{i,h}+\beta^{I}_{h}(\hat{E}^{I}_{i,t}-e^{n}_{i,t})+u^{I}_{i,t+h}\text{ ,}$$

$$\text{PC+k :}\quad e_{i,t+h}-e_{i,t}=c^{P}_{i,k,h}+\beta^{P}_{k,h}(\hat{E}^{P}_{i,t}-e^{n}_{i,t})+\gamma^{P}_{k,h}(z_{i,k,t}-e_{i,t})+u^{P}_{i,k,t+h}\text{ , } k\text{ =TR, M, PPP.}$$

$$\text{IC+k:}\quad e_{i,t+h}-e_{i,t}=c^{I}_{i,k,h}+\beta^{I}_{k,h}(\hat{E}^{I}_{i,t}-e^{n}_{i,t})+\gamma^{I}_{k,h}(z_{i,k,t}-e_{i,t})+u^{I}_{i,k,t+h}$$

Model	Median *TU* #*TU*<1 or (*CW*<-1.282)				$\overline{TU}$	$\overline{CW1}$	$\overline{CW2}$
	h=1	4	8	12			
			Early/*N*=17				
PC	1.003	1.053	1.071	1.106	0.22	0.12	0.21
	6(1)	5(0)	2(2)	2(5)			
IC	0.987	1.015	1.015	1.005	0.47	0.32	0.35
	11(7)	6(3)	7(5)	8(7)			
PC+TR	1.01	1.062	1.068	1.104	0.22	0.12	0.21
	6(1)	5(0)	2(2)	2(5)			
IC+TR	0.995	1.016	1.016	**0.999**	0.49	0.28	0.35
	10(4)	7(3)	7(5)	9(7)			
PC+M	1.001	1.024	1.035	1.019	0.46	0.22	0.32
	8(3)	7(1)	8(4)	8(7)			
IC+M	**0.99**	**0.994**	**1.003**	**0.959**	**0.57**	0.46	**0.56**
	13(7)	9(5)	8(7)	9(12)			
PC+PPP	**0.987**	**0.946**	**0.888**	**0.765**	**0.91**	**0.68**	**0.88**
	15(6)	15(10)	16(15)	16(15)			
IC+PPP	1.003	1.045	1.089	1.19	0.34	0.32	0.41
	7(4)	5(4)	4(7)	7(7)			
			Long/*N*=9				
PC	1.012	1.041	1.085	1.143	0.08	0.08	0.17
	1(0)	1(0)	0(1)	1(2)			
IC	1.029	1.078	1.103	1.069	0.00	0.11	0.17
	0(0)	0(1)	0(1)	0(2)			
PC+TR	1.010	1.041	1.076	1.144	0.08	0.11	0.22
	1(0)	1(0)	0(2)	1(2)			
IC+TR	1.035	1.084	1.103	1.07	0.00	0.11	0.17
	0(0)	0(1)	0(1)	0(2)			
PC+M	1.000	1.014	**0.997**	**0.960**	**0.50**	0.25	0.44
	4(0)	4(1)	5(4)	5(4)			
IC+M	1.032	1.081	1.08	1.036	0.17	0.39	0.72
	0(0)	0(1)	3(5)	3(8)			
PC+PPP	**0.992**	**0.97**	**0.899**	**0.875**	**0.86**	**0.81**	**1.00**
	5(5)	8(6)	9(9)	9(9)			
IC+PPP	1.022	1.076	1.089	1.116	0.14	0.44	**0.72**
	1(0)	1(3)	2(6)	1(7)			

Notes: The base currency is the Japanese yen. Early and Long refer to the periods 1973-1998 and 1973-2011, respectively. Others are the same as those in Table A1.

Case 10: The Rolling forecasting scheme.

Table A10.1: Out-of-sample forecasts with the rolling window being ten years

$$\text{PC:}\quad e_{i,t+h}-e_{i,t}=c^{P}_{i,h}+\beta^{P}_{h}(\hat{E}^{P}_{i,t}-e^{n}_{i,t})+u^{P}_{i,t+h}\text{, } i\text{=1,…, }N\text{; }h\text{=1,…,}H.$$

$$\text{IC:}\quad e_{i,t+h}-e_{i,t}=c^{I}_{i,h}+\beta^{I}_{h}(\hat{E}^{I}_{i,t}-e^{n}_{i,t})+u^{I}_{i,t+h}\text{,}$$

$$\text{PC+k:}\quad e_{i,t+h}-e_{i,t}=c^{P}_{i,k,h}+\beta^{P}_{k,h}(\hat{E}^{P}_{i,t}-e^{n}_{i,t})+\gamma^{P}_{k,h}(z_{i,k,t}-e_{i,t})+u^{P}_{i,k,t+h}\text{, } k\text{ =}TR, M, PPP.$$

$$\text{IC+k:}\quad e_{i,t+h}-e_{i,t}=c^{I}_{i,k,h}+\beta^{I}_{k,h}(\hat{E}^{I}_{i,t}-e^{n}_{i,t})+\gamma^{I}_{k,h}(z_{i,k,t}-e_{i,t})+u^{I}_{i,k,t+h}$$

Model	Median *TU* #*TU*<1 or (*CW*<-1.282)				$\overline{TU}$	$\overline{CW1}$	$\overline{CW2}$
	h=1	4	8	12			
			Early/*N*=17				
PC	1.037	1.136	1.265	1.444	0.10	0.06	0.09
	3(0)	1(1)	1(1)	2(2)			
IC	1.019	1.010	**0.984**	1.041	0.35	0.28	**0.56**
	0(0)	7(0)	10(10)	7(9)			
PC+TR	1.034	1.146	1.263	1.406	0.10	0.07	0.09
	3(1)	1(1)	1(1)	2(2)			
IC+TR	1.021	1.009	**0.970**	1.049	0.35	0.29	**0.53**
	1(0)	7(2)	10(10)	6(8)			
PC+M	1.022	1.133	1.194	1.289	0.19	0.24	0.24
	5(5)	4(3)	2(3)	2(5)			
IC+M	1.012	**0.998**	1.032	1.020	0.38	0.34	0.47
	2(0)	9(7)	8(7)	7(9)			
PC+PPP	1.023	1.117	1.132	1.156	0.24	0.26	0.44
	3(1)	2(2)	5(6)	6(9)			
IC+PPP	1.039	1.114	1.203	1.213	0.25	0.24	0.47
	2(0)	3(0)	5(6)	7(10)			
			Long/*N*=9				
PC	1.046	1.147	1.261	1.331	0.11	0.14	0.17
	0(1)	1(1)	2(2)	1(1)			
IC	1.013	1.005	1.018	1.054	0.31	0.33	**0.56**
	1(0)	3(2)	4(5)	3(5)			
PC+TR	1.050	1.163	1.278	1.324	0.11	0.08	0.06
	1(1)	1(1)	1(0)	1(1)			
IC+TR	1.012	1.007	**0.976**	1.040	0.39	0.36	**0.56**
	2(1)	3(2)	5(5)	4(5)			
PC+M	1.073	1.207	1.327	1.558	0.06	0.11	0.17
	0(0)	1(1)	1(2)	0(1)			
IC+M	1.070	1.172	1.256	1.306	0.17	0.22	0.28
	1(1)	2(2)	2(2)	1(3)			
PC+PPP	1.057	1.122	1.143	1.195	0.14	0.36	**0.61**
	0(0)	0(2)	2(6)	3(5)			
IC+PPP	1.049	1.159	1.276	1.371	0.03	0.42	**0.61**
	0(2)	0(4)	0(5)	1(6)			

Notes: A rolling scheme is adopted in out-of-sample contests with the rolling window being ten years. Early and Long refer to the periods 1973-1998 and 1973-2011, respectively. Others are the same as those in Table A1.

Table A10.2: Out-of-sample forecasts with the rolling window being fourteen years

$$\text{PC:}\quad e_{i,t+h}-e_{i,t}=c^{P}_{i,h}+\beta^{P}_{h}(\hat{E}^{P}_{i,t}-e^{n}_{i,t})+u^{P}_{i,t+h},\ i=1,\ldots,N;\ h=1,\ldots,H.$$

$$\text{IC:}\quad e_{i,t+h}-e_{i,t}=c^{I}_{i,h}+\beta^{I}_{h}(\hat{E}^{I}_{i,t}-e^{n}_{i,t})+u^{I}_{i,t+h},$$

$$\text{PC+k:}\quad e_{i,t+h}-e_{i,t}=c^{P}_{i,k,h}+\beta^{P}_{k,h}(\hat{E}^{P}_{i,t}-e^{n}_{i,t})+\gamma^{P}_{k,h}(z_{i,k,t}-e_{i,t})+u^{P}_{i,k,t+h},\ k=TR, M, PPP.$$

$$\text{IC+k:}\quad e_{i,t+h}-e_{i,t}=c^{I}_{i,k,h}+\beta^{I}_{k,h}(\hat{E}^{I}_{i,t}-e^{n}_{i,t})+\gamma^{I}_{k,h}(z_{i,k,t}-e_{i,t})+u^{I}_{i,k,t+h}$$

Model	Median *TU* #*TU*<1 or (*CW*<-1.282)				$\overline{TU}$	$\overline{CW1}$	$\overline{CW2}$
	h=1	4	8	12			
			Early/*N*=17				
PC	1.023	1.073	1.192	1.252	0.13	0.03	0.06
	2(0)	2(0)	2(2)	3(0)			
IC	1.004	1.005	1.054	1.077	0.28	0.15	0.26
	3(0)	4(1)	6(4)	6(5)			
PC+TR	1.015	1.069	1.178	1.246	0.12	0.06	0.09
	2(0)	2(1)	2(3)	2(0)			
IC+TR	1.001	**0.995**	1.059	1.065	0.43	0.22	0.29
	8(2)	9(3)	6(5)	6(5)			
PC+M	1.015	1.037	1.025	1.109	0.34	0.21	0.32
	4(1)	5(2)	8(6)	6(5)			
IC+M	**0.996**	**0.975**	**0.963**	1.018	**0.56**	0.31	0.41
	10(1)	11(6)	10(7)	7(7)			
PC+PPP	1.014	1.066	1.101	1.220	0.37	0.22	0.38
	6(0)	6(2)	7(6)	6(7)			
IC+PPP	1.005	1.002	1.101	1.272	0.35	0.25	0.35
	6(1)	8(4)	6(6)	4(6)			
			Long/*N*=9				
PC	1.023	1.073	1.112	1.189	0.06	0.03	0.00
	0(0)	1(1)	1(0)	0(0)			
IC	1.001	**0.996**	1.008	1.053	0.42	0.25	0.28
	3(1)	5(3)	3(3)	4(2)			
PC+TR	1.025	1.083	1.123	1.157	0.06	0.03	0.00
	0(0)	1(1)	1(0)	0(0)			
IC+TR	1.008	1.006	1.007	1.036	0.39	0.33	0.39
	4(2)	3(3)	3(3)	4(4)			
PC+M	1.035	1.131	1.274	1.499	0.08	0.03	0.00
	1(0)	1(1)	1(0)	0(0)			
IC+M	1.019	1.082	1.109	1.227	0.14	0.17	0.28
	1(0)	2(1)	1(3)	1(2)			
PC+PPP	1.025	1.092	1.101	1.175	0.17	0.42	**0.78**
	0(0)	1(1)	2(7)	3(7)			
IC+PPP	1.036	1.095	1.094	1.222	0.14	0.44	**0.72**
	1(1)	1(2)	2(7)	1(6)			

Notes: A rolling scheme is adopted in out-of-sample contests with the rolling window being fourteen years. Early and Long refer to the periods 1973-1998 and 1973-2011, respectively. Others are the same as those in Table A1.