Federal Reserve Bank of Cleveland Working Paper Series

Online-Only Appendix for An Empirical Evaluation of Some Long-Horizon Macroeconomic Forecasts

Kurt G. Lunsford and Kenneth D. West

Working Paper No. 24-20

September 2024

Suggested citation: Lunsford, Kurt G. and Kenneth D. West. 2024. "Online-Only Appendix for An Empirical Evaluation of Some Long-Horizon Macroeconomic Forecasts." Working Paper No. 24-20. Federal Reserve Bank of Cleveland. https://doi.org/10.26509/frbc-wp-202420.

Federal Reserve Bank of Cleveland Working Paper Series
ISSN: 2573-7953

Online-Only Appendix for

An Empirical Evaluation of Some Long-Horizon Macroeconomic Forecasts

Kurt G. Lunsford*   Kenneth D. West†

Contents

*Federal Reserve Bank of Cleveland, P.O. Box 6387, Cleveland, OH 44101; kurt.lunsford@clev.frb.org

†Department of Economics, University of Wisconsin, Madison, WI 53706; kdwest@wisc.edu

A Data Countries and Samples

The following bullets list the countries for each variable and their respective samples:

- **Per capita real GDP growth, CPI inflation, and population growth:** We use 17 countries: AUS, BEL, CAN, CHE, DEU, DNK, ESP, FIN, FRA, GBR, ITA, JPN, NLD, NOR, PRT, SWE, and USA. All samples are 1871-2020.

- **Labor productivity growth:** We use 18 countries. BEL, CAN, CHE, CHL, DEU, DNK, ESP, FIN, FRA, GBR, GRC, IRL, ITA, JPN, NLD, SWE, and USA have samples of 1871-2022. NOR has a sample of 1876-2022.

- **Broad money growth:** We use 12 countries. The countries and respective samples are AUS (1871-2020), CAN (1872-2020), CHE (1881-2020), DNK (1871-2020), FIN (1871-2020), GBR (1871-2020), ITA (1871-2020), JPN (1871-2020), NOR (1871-2020), PRT (1871-2020), SWE (1872-2020), USA (1871-2020).

- **Total equity returns:** We use 11 countries. The countries and respective samples are AUS (1870-2020), BEL (1870-2020), DEU (1870-2020), DNK (1873-2020), FRA (1870-2020), GBR (1871-2020), ITA (1870-2020), NOR (1881-2020), PRT (1871-2020), SWE (1871-2020), USA (1872-2020).

- **Short-term nominal interest rate:** We use 9 countries. The countries and respective samples are CHE (1870-2020), DNK (1875-2020), ESP (1870-2020), FIN (1870-2020), GBR (1870-2020), NLD (1870-2020), PRT (1880-2020), SWE (1870-2020), USA (1870-2020).

- **Long-term nominal interest rate:** We use 12 countries. AUS, CAN, DNK, FRA, GBR, ITA, JPN, NOR, PRT, SWE, and USA have samples 1870-2020. CHE has a sample of 1880-2020.

- **Real exchange rate:** We use 16 countries. AUS, BEL, CAN, CHE, DEU, DNK, ESP, FIN, FRA, GBR, ITA, NLD, NOR, PRT, and SWE have samples of 1870-2020. JPN has a sample of 1873-2020. We compute the real exchange rate as US CPI times nominal exchange rate (expressed in local currency over US currency) divided by home country CPI.

- **Investment to GDP ratio:** We use 7 countries. ESP, FIN, BGR, ITA, SWE, and USA have samples of 1870-2020. CAN has a sample of 1871-2020.

B Formulas for the iid, AR(1), and Random Walk Models

For the iid, random walk, and AR(1) models, we treat the future average realization, $\bar{x}_{\tau,h}$, as normally distributed with a mean, $f_{\tau,h}$, and a variance, $V_{\tau,h}$. The three different models have different forms for $f_{\tau,h}$ and $V_{\tau,h}$, and we provide derivations in Appendix E. We estimate $f_{\tau,h}$ and $V_{\tau,h}$ with both a rolling estimation sample, $\{x_{\tau-R+1}, \ldots x_\tau\}_{\tau=R}^{T-h}$, and a recursive estimation sample, $\{x_1, \ldots, x_\tau\}_{\tau=R}^{T-h}$. For the formulas reported in this appendix and in Appendix C, we only use recursive sample notation with the understanding that rolling sample notation takes a parallel form.

The iid Model. The iid model assumes that the data generating process (DGP) for x_t is $x_t = \mu + u_t$, in which u_t is iid with a mean of zero and variance of σ^2. We estimate μ and σ^2 with $\hat{\mu}_\tau = \tau^{-1} \sum_{t=1}^{\tau} x_t$ and $\hat{\sigma}_\tau^2 = (\tau-1)^{-1} \sum_{t=1}^{\tau} (x_t - \hat{\mu})^2$. We then use

$$\hat{f}_{\tau,h}^{iid} = \hat{\mu}_\tau, \tag{B.1}$$

$$\hat{V}_{\tau,h}^{iid} = [(1/h) + (1/\tau)]\hat{\sigma}_\tau^2. \tag{B.2}$$

The Random Walk Model. The random walk model assumes that the DGP for x_t is a random walk with no drift, $x_t = x_{t-1} + u_t$, in which u_t is iid with a mean of zero and variance of σ^2. We estimate σ^2 with $\hat{\sigma}_\tau^2 = (\tau-1)^{-1} \sum_{t=2}^{\tau} (x_t - x_{t-1})^2$. We then use

$$\hat{f}_{\tau,h}^{rw} = x_\tau, \tag{B.3}$$

$$\hat{V}_{\tau,h}^{rw} = (h+1)(2h+1)\hat{\sigma}_\tau^2/(6h). \tag{B.4}$$

The AR(1) Model. The AR(1) model assumes that the DGP for x_t is $x_t = \rho_0 + \rho_1 x_{t-1} + u_t$, in which u_t is iid with a mean of zero and variance of σ^2. We first estimate ρ_0 and ρ_1 with ordinary least squares and then bias-adjust following Yamamoto and Kunitomo (1984), denoting the bias-adjusted estimates with $\tilde{\rho}_0$ and $\tilde{\rho}_1$ (we suppress dependence on τ for notational simplicity). If $\tilde{\rho}_1 \geq 1$, then we forecast with the random walk model. If $\tilde{\rho}_1 < 1$, then we proceed with the AR(1) model. We compute $\tilde{u}_t = x_t - \tilde{\rho}_0 - \tilde{\rho}_1 x_{t-1}$ and $\tilde{\sigma}_\tau^2 = (\tau-3)^{-1} \sum_{t=2}^{\tau} \tilde{u}_t^2$. We allow for uncertainty in $\tilde{\rho}_0$ when deriving our forecast distribution but assume that $\tilde{\rho}_1 = \rho_1$. We use

$$\hat{f}^{ar1}_{\tau,h} = \frac{\tilde{\rho}_0}{1-\tilde{\rho}_1} + \frac{1}{h}(\tilde{\rho}_1 + \tilde{\rho}_1^2 + \cdots + \tilde{\rho}_1^h)\left(x_\tau - \frac{\tilde{\rho}_0}{1-\tilde{\rho}_1}\right), \tag{B.5}$$

$$\hat{V}^{ar1}_{\tau,h} = [1 + (1+\tilde{\rho}_1)^2 + \cdots + (1 + \tilde{\rho}_1 + \cdots + \tilde{\rho}_1^{h-1})^2]\tilde{\sigma}^2_\tau/h^2 \tag{B.6}$$
$$+[1 + (1+\tilde{\rho}_1) + \cdots + (1 + \tilde{\rho}_1 + \cdots \tilde{\rho}_1^{h-1})]^2\tilde{\sigma}^2_\tau/(h^2(\tau-1)).$$

Notice that if $\tilde{\rho}_1 = 0$, then Equation (B.6) reduces to $\hat{V}^{ar1}_{\tau,h} = \tilde{\sigma}^2_\tau/h + \tilde{\sigma}^2_\tau/(\tau-1)$. This variance estimate parallels $\hat{V}^{iid}_{\tau,h}$ in Equation (B.2), except with $\tau - 1$ in place of τ to account for one less estimate of u_t with the AR(1) model than with the iid model.

C Formulas for the Müller and Watson (2016) Models

The Müller and Watson (2016) (MW) forecasting models are based on extracting low-frequency patterns from the sample $\{x_1, \ldots, x_\tau\}$ by using a small number, q, of slowly cycling cosine waves. The tth observation of the jth cosine wave is given by $\psi_{j,t} = \sqrt{2}\cos(\pi j(t-1/2)/\tau)$. The jth cosine wave completes one cycle in $2\tau/j$ periods. So, for example, if the estimation sample size is 48 years and the number of cosine waves is $q = 8$, then the cosine waves have periods from 12 to 96 years.

We use averages to extract the low-frequency patterns in the data. First, we compute the simple average $\hat{\beta}_{\tau,0} = \tau^{-1}\sum_{t=1}^{\tau} x_t$. Then, we compute the cosine-weighted averages or "cosine transforms," $\hat{\beta}_{\tau,j} = \tau^{-1}\sum_{t=1}^{\tau}\psi_{j,t}x_t$ for $j = 1, \ldots, q$. The MW forecast densities are then constructed from $\hat{\beta}_{\tau,0}, \ldots, \hat{\beta}_{\tau,q}$. By focusing on the low-frequency information in $\{x_1, \ldots, x_\tau\}$, MW effectively reduce the dimension of the data from τ down to $q+1$.

The MW models are based on the asymptotic properties of the sample averages. Let $\hat{\beta}_{\tau,1:q} = [\hat{\beta}_{1,\tau}, \ldots, \hat{\beta}_{q,\tau}]'$ be the vector of cosine transforms and $y_{\tau,h} = \bar{x}_{\tau,h} - \hat{\beta}_{\tau,0}$ be the future average of x_t centered on the in-sample average. Then, we use

$$\tau^{1-\kappa}\begin{bmatrix}\hat{\beta}_{\tau,1:q} \\ y_{\tau,h}\end{bmatrix} \sim N(0,\Sigma), \quad \Sigma = \begin{bmatrix}\Sigma_{\beta\beta} & \Sigma_{\beta y} \\ \Sigma_{y\beta} & \Sigma_{yy}\end{bmatrix}, \tag{C.1}$$

in which κ is a scaling factor that depends on the relevant model for x_t: $\kappa = 1/2$ for the MW0 model, $\kappa = 3/2$ for the MW1 model, and $\kappa = 1/2 + d$ for the MWd model. The covariance matrix,

Σ, also depends on the model for x_t, and we provide details for computing Σ in Appendix F.

The MW0 Model. The DGP for x_t is $x_t = \mu + u_t$, in which u_t is a mean zero and integrated of order zero or $I(0)$ process. For this model, MW compute the covariance matrix, Σ, in (C.1) analytically. Let σ^2_{lrv} be the long-run variance of u_t. Then, $\Sigma_{\beta\beta} = \sigma^2_{lrv} I_q$, in which I_q is the $(q \times q)$ identity matrix, $\Sigma_{y\beta} = \Sigma'_{\beta y}$ is a $(1 \times q)$ matrix of zeros, and $\Sigma_{yy} = [(1/h)+(1/\tau)](\tau\sigma^2_{lrv})$. MW then show that $\bar{x}_{\tau,h}$ has a generalized Student-t distribution with q degrees of freedom, a location parameter of $\hat{\beta}_{\tau,0}$, which will be the point forecast $\hat{f}^{MW0}_{\tau,h}$, and a scale parameter of $\sqrt{[(1/h) + (1/\tau)](\tau\hat{\beta}'_{\tau,1:q}\hat{\beta}_{\tau,1:q}/q)}$.

We make two observations, comparing the MW0 model to the iid model. First, the point forecasts of the MW0 and iid models are the same: the in-sample average. That is, $\hat{f}^{iid}_{\tau,h} = \hat{\mu}_\tau = \hat{\beta}_{\tau,0} = \hat{f}^{MW0}_{\tau,h}$. Second, the densities and the forecast intervals of the MW0 and iid models only differ in two ways. First, the MW0 forecast density is Student-t with q degrees of freedom, while the iid forecast density is Normal. This difference reflects MW's data reduction from a large number, τ, of observations to a small number, q, of cosine transforms. Second, the scale parameters are $\sqrt{[(1/h) + (1/\tau)]}$ times the square root of the estimated long-run variances for the respective model: $\sqrt{\hat{\sigma}^2_\tau}$ for the iid model and $\sqrt{\tau\hat{\beta}'_{\tau,1:q}\hat{\beta}_{\tau,1:q}/q}$ for the MW0 model.[1]

The MW1 Model. The DGP for x_t is $x_t = \mu + u_t$, in which u_t is a mean zero and integrated of order one or $I(1)$ process. Unlike for the MW0 model, MW do not provide an analytical form for every element of Σ when u_t is $I(1)$. Because of this, we use an approximation of Σ, providing formulas in Appendix F. Given the appropriate form of Σ, MW show that $\bar{x}_{\tau,h}$ has a generalized Student-t density with q degrees of freedom, a location parameter of $\hat{\beta}_{\tau,0} + \Sigma_{y\beta}\Sigma^{-1}_{\beta\beta}\hat{\beta}_{\tau,1:q}$, which will be the point forecast $\hat{f}^{MW1}_{\tau,h}$, and a scale parameter of $\sqrt{(\Sigma_{yy} - \Sigma_{y\beta}\Sigma^{-1}_{\beta\beta}\Sigma_{\beta y})(\hat{\beta}'_{\tau,1:q}\Sigma^{-1}_{\beta\beta}\hat{\beta}_{\tau,1:q}/q)}$.

The MWd Model. The DGP for x_t is $x_t = \mu + u_t$, in which u_t is a mean zero and fractionally integrated or $I(d)$ process with fractional parameter $d \in (-0.5, 1)$. We treat d as unknown and use a Bayesian approach to construct the forecast density. Following MW, we set a grid of potential values of d, $\{-0.4, -0.2, 0, 0.2, 0.4, 0.6, 0.8, 1\}$, and use a prior of uniform mass on each grid point. The resulting Bayes predictive density is a weighted average of generalized Student-t densities with q degrees of freedom. We provide further details in Appendix F.

[1] See Section 3.1 of Müller and Watson (2015) for a discussion of using cosine transforms to estimate the long-run variance of an $I(0)$ process.

D Additional Tables

In this appendix, we show tables with additional results from our pseudo out-of-sample analysis. They are as follows.

- Table D.1 shows coverage rates for a different categorization of variables that are plausibly stationary and plausibly non-stationary. To categorize the variables, we use the maximum likelihood estimate of d from the MWd model estimated on the longest estimation sample (the $h = 10$ recursive sample). For example, for per capita GDP growth for the USA, this sample is 1871-2010. If the estimate of d is less than 0.5, then we say the variable is plausibly stationary. Otherwise, we say the variable is plausibly non-stationary.

- Table D.2 shows the number of countries per variable that are plausibly stationary using the method in the previous bullet.

- Table D.3 shows actual coverage rates for each of our 10 variables by forecasting model and forecast horizon. As in our main results, we show medians and IQRs. Take per capita real GDP growth as an example. We have 17 countries times 2 sampling schemes to give 34 coverage rates. We then show the median and IQR across these 34 coverage rates.

- Table D.4 shows probability integral transform (PIT) rates by forecasting model and forecast horizon. For this table, we collect the PITs separately for the stationary and non-stationary variables. By construction, the value of a PIT is between 0 and 1. We then report the fraction of the PITs that fall in the intervals $[0, 0.2)$, $[0.2, 0.4)$, $[0.4, 0.6)$, $[0.6, 0.8)$, and $[0.8, 1.0]$. Because the PITs for a well-calibrated model are uniformly distributed on $[0, 1]$, the ideal value is 0.2 for each interval.

- Table D.5 shows what we call PIT distances by forecasting model and forecast horizon. For the 5 intervals in the previous table, let r_k be the realized fraction of the PITs that fall into the kth interval. Then, we compute the PIT distance as $(1/5)\sum_{k=1}^{5}|r_k - 0.2|$, which is a distance from the realized PIT rates to the ideal PIT rates.

  We make six comments about the distances reported in Table D.5. First, for plausibly stationary variables for $h = 10$, the MW0, AR(1), and MWd models have the smallest distances, consistent with our coverage rate results. Second, for plausibly stationary variables for $h = 25$, the AR(1) and MWd have the smallest PIT distances with little increase in distance compared to $h = 10$. Third, for plausibly stationary variables, all models show big increases in PIT distances from $h = 25$ to $h = 50$. Fourth, for plausibly non-stationary variables for $h = 10$, the random walk and MW1 models have reasonably small PIT distances (comparable to the AR(1) model for stationary variables), consistent with our coverage rate results. Fifth, for plausibly non-stationary variables, the random walk model shows little increase in PIT distance from $h = 10$ to $h = 25$ and has the same PIT distance at $h = 25$ as the AR(1) and MWd models for the stationary variables. Sixth, for plausibly non-stationary variables, all models except the iid model show big increases in PIT distances from $h = 25$ to $h = 50$.

- Table D.6 summarizes the continuous ranked probability score (CRPS) results. Let $F_{\tau,h}(\cdot)$ be the cumulative distribution function for a forecast distribution made with sample $\{x_t\}_{t=1}^{\tau}$ for horizon h. Then, we compute

$$CRPS_{\tau,h} = \int_{-\infty}^{\infty} [F_{\tau,h}(y) - \mathbf{1}(y \geq \bar{x}_{\tau,h})]^2 dy$$

for $\tau = R, \ldots, T - h$ and take the average. For the iid, AR(1), and random walk models, which have normal distributions, we use the CRPS formula on page 367 of Gneiting and Raftery (2007). For the MW0 and MW1 models, which have Student-t distributions, we use the formulas on page 25 of Jordan, Krüger, and Lerch (2019). For the MWd model, we use numerical integration. As with the Winkler score table in the body of the paper, we report CRPS results in values relative to the iid model. We also report the fraction of samples for a given forecast horizon in which each model has the lowest CRPS.

- Table D.7 summarizes the root mean squared prediction error (RMSPE) results. We compute RMSPE with $\sqrt{P^{-1} \sum_{\tau=R}^{T-h} (\bar{x}_{\tau,h} - \hat{f}_{\tau,h})^2}$, in which the point forecast, $\hat{f}_{\tau,h}$, is the mean of the forecast distribution. Point forecasts for the iid and MW0 models are the same, and we report results for these models jointly. As with the Winkler score table in the body of the paper, we report RMSPE results in values relative to the iid model. We also report the fraction of samples for a given forecast horizon in which each model has the lowest RMSPE.

- Table D.8 summarizes the mean absolute prediction error (MAPE) results. We compute the MAPE with $P^{-1}\sum_{\tau=R}^{T-h}|\bar{x}_{\tau,h} - \hat{f}_{\tau,h}|$ in which the point forecast, $\hat{f}_{\tau,h}$, is the median of the forecast distribution. We use the median because the MAPE is a consistent scoring (or loss) function for the median of the forecast distribution (while mean squared errors are a consistent scoring function for the mean) (Gneiting, 2011). Use of the median rather than the mean of the forecast distributions only affects forecasts of the MWd model; mean and median are the same in all other models. Point forecasts for the iid and MW0 models are the same, and we report results for these models jointly. As with the Winkler score table in the body of the paper, we report MAPE results in values relative to the iid model. We also report the fraction of samples for a given forecast horizon in which each model has the lowest MAPE.

- Table D.9 summarizes the absolute forecast bias results. We compute absolute forecast bias as $|P^{-1}\sum_{\tau=R}^{T-h}(\bar{x}_{\tau,h} - \hat{f}_{\tau,h})|$ in which the point forecast, $\hat{f}_{\tau,h}$, is the mean of the forecast distribution. Point forecasts for the iid and MW0 models are the same, and we report results for these models jointly. As with the Winkler score table in the body of the paper, all absolute bias results are reported in values relative to the iid model. We also report the fraction of samples for a given forecast horizon in which each model has the lowest absolute bias.

Table D.1: Coverage rates of nominal 68 percent forecast intervals: medians and IQRs

	(1)	(2)	(3a)	(3b)	(4a)	(4b)
(1)			Stationary variables		Nonstationary variables	
(2)			180 samples		92 samples	
(3)	horizon	model	median coverage	IQR	median coverage	IQR
(4)	10	iid	0.50	(0.38, 0.66)	0.14	(0.11, 0.26)
(5)	10	MW0	0.63	(0.53, 0.73)	0.36	(0.27, 0.51)
(6)	10	AR(1)	0.69	(0.61, 0.78)	0.56	(0.46, 0.69)
(7)	10	MWd	0.72	(0.65, 0.80)	0.58	(0.45, 0.70)
(8)	10	RW	0.95	(0.88, 0.97)	0.72	(0.59, 0.84)
(9)	10	MW1	0.76	(0.71, 0.83)	0.63	(0.51, 0.75)
(10)	25	iid	0.42	(0.26, 0.63)	0.10	(0.06, 0.16)
(11)	25	MW0	0.57	(0.41, 0.69)	0.24	(0.16, 0.37)
(12)	25	AR(1)	0.65	(0.53, 0.77)	0.41	(0.28, 0.61)
(13)	25	MWd	0.65	(0.54, 0.76)	0.45	(0.27, 0.63)
(14)	25	RW	0.98	(0.95, 0.99)	0.75	(0.42, 0.89)
(15)	25	MW1	0.86	(0.79, 0.91)	0.65	(0.33, 0.82)
(16)	50	iid	0.28	(0.13, 0.59)	0.04	(0.00, 0.13)
(17)	50	MW0	0.49	(0.22, 0.69)	0.13	(0.00, 0.23)
(18)	50	AR(1)	0.54	(0.32, 0.79)	0.30	(0.21, 0.49)
(19)	50	MWd	0.57	(0.32, 0.79)	0.35	(0.26, 0.56)
(20)	50	RW	1.00	(0.98, 1.00)	0.79	(0.46, 0.94)
(21)	50	MW1	0.92	(0.87, 0.94)	0.69	(0.37, 0.88)

Notes:

1. Stationary variables are defined as those having a maximum likelihood value of d in the MWd model less than 0.5 over the longest available sample.
2. Non-stationary variables are defined as those having a maximum likelihood value of d in the MWd model larger than 0.5 over the longest available sample.
3. We show the number of countries per variable that have plausibly stationary series in Table D.2.
4. The number of samples for each group in row (2) is the number of variables in that group times the two sampling schemes (rolling and recursive).
5. In the list of models in column (2), we use the shorthand "RW" for the random walk model.
6. Medians and IQRs (interquartile ranges) were constructed as described in the notes to Table 5.1 in the body of the paper.

Table D.2: Frequency of stationarity by variable

(1)	(1) variable	(2) no. of countries	(3) no. of countries with $\hat{d} < 0.5$
(2)	GDP growth	17	17
(3)	productivity growth	18	18
(4)	CPI inflation	17	15
(5)	money growth	12	10
(6)	population growth	17	9
(7)	equity returns	11	11
(8)	short-term interest	9	0
(9)	long-term interest	12	0
(10)	real exchange rate	16	10
(11)	I/Y ratio	7	0

Notes:

1. This table reports the number of countries per variable that have plausibly stationary series.

2. $\hat{d}$ is the maximum likelihood estimate of d using the longest available sample of data. For example, for per capita GDP growth, we estimate $\hat{d}$ on 1871-2010 (the recursive sample for $h = 10$).

Table D.3: Coverage rates of nominal 68 percent forecast intervals: medians and IQRs

	(1)	(2)	(3a)	(3b)	(4a)	(4b)	(5a)	(5b)	(6a)	(6b)	(7a)	(7b)
(1)			GDP growth		Productivity growth		CPI inflation		Money growth		Population growth	
(2)			17 countries		18 countries		17 countries		12 countries		17 countries	
(3)			34 samples		36 samples		34 samples		24 samples		34 samples	
(4)	horizon	model	median coverage	IQR	median coverage	IQR	median coverage	IQR	median coverage	IQR	median coverage	IQR
(5)	10	iid	0.65	(0.55, 0.71)	0.56	(0.47, 0.72)	0.48	(0.36, 0.56)	0.41	(0.37, 0.45)	0.28	(0.20, 0.40)
(6)	10	MW0	0.69	(0.58, 0.80)	0.61	(0.52, 0.69)	0.68	(0.62, 0.79)	0.62	(0.56, 0.68)	0.58	(0.42, 0.64)
(7)	10	AR(1)	0.70	(0.61, 0.80)	0.63	(0.57, 0.71)	0.79	(0.73, 0.84)	0.66	(0.62, 0.74)	0.70	(0.56, 0.76)
(8)	10	MWd	0.71	(0.63, 0.78)	0.71	(0.61, 0.77)	0.81	(0.76, 0.85)	0.75	(0.68, 0.80)	0.71	(0.62, 0.76)
(9)	10	RW	0.96	(0.95, 0.98)	0.97	(0.95, 0.99)	0.95	(0.93, 0.96)	0.94	(0.90, 0.97)	0.88	(0.80, 0.92)
(10)	10	MW1	0.76	(0.74, 0.80)	0.77	(0.74, 0.83)	0.83	(0.78, 0.88)	0.80	(0.76, 0.84)	0.77	(0.70, 0.83)
(11)	25	iid	0.70	(0.49, 0.76)	0.58	(0.41, 0.69)	0.32	(0.26, 0.51)	0.30	(0.22, 0.39)	0.24	(0.17, 0.32)
(12)	25	MW0	0.69	(0.55, 0.84)	0.57	(0.43, 0.71)	0.60	(0.44, 0.72)	0.56	(0.43, 0.67)	0.50	(0.33, 0.59)
(13)	25	AR(1)	0.72	(0.58, 0.83)	0.59	(0.49, 0.71)	0.67	(0.58, 0.73)	0.65	(0.52, 0.72)	0.67	(0.48, 0.79)
(14)	25	MWd	0.65	(0.55, 0.78)	0.65	(0.55, 0.77)	0.71	(0.63, 0.83)	0.66	(0.64, 0.73)	0.69	(0.60, 0.76)
(15)	25	RW	0.99	(0.97, 1.00)	0.99	(0.98, 1.00)	0.97	(0.96, 0.99)	0.99	(0.97, 1.00)	0.94	(0.88, 0.98)
(16)	25	MW1	0.88	(0.85, 0.92)	0.86	(0.84, 0.91)	0.88	(0.80, 0.91)	0.86	(0.79, 0.92)	0.85	(0.81, 0.91)
(17)	50	iid	0.43	(0.22, 0.65)	0.57	(0.26, 0.79)	0.20	(0.08, 0.47)	0.13	(0.08, 0.27)	0.15	(0.07, 0.32)
(18)	50	MW0	0.58	(0.26, 0.75)	0.50	(0.30, 0.77)	0.49	(0.19, 0.84)	0.48	(0.28, 0.69)	0.41	(0.17, 0.58)
(19)	50	AR(1)	0.58	(0.33, 0.76)	0.58	(0.33, 0.79)	0.53	(0.37, 0.87)	0.56	(0.22, 0.87)	0.55	(0.25, 0.77)
(20)	50	MWd	0.45	(0.28, 0.67)	0.50	(0.33, 0.74)	0.70	(0.56, 0.91)	0.65	(0.51, 0.85)	0.76	(0.38, 0.87)
(21)	50	RW	1.00	(1.00, 1.00)	1.00	(0.98, 1.00)	0.98	(0.98, 1.00)	0.99	(0.98, 1.00)	0.96	(0.94, 0.98)
(22)	50	MW1	0.92	(0.89, 0.95)	0.91	(0.87, 0.97)	0.94	(0.88, 0.94)	0.94	(0.92, 0.96)	0.92	(0.89, 0.94)
(23)			Equity returns		Short-term rates		Long-term rates		Real exchange rates		I/Y ratio	
(24)			11 countries		9 countries		12 countries		16 countries		7 countries	
(25)			22 samples		18 samples		24 samples		32 samples		14 samples	
(26)	horizon	model	median coverage	IQR	median coverage	IQR	median coverage	IQR	median coverage	IQR	median coverage	IQR
(27)	10	iid	0.62	(0.57, 0.73)	0.11	(0.10, 0.13)	0.14	(0.11, 0.17)	0.20	(0.12, 0.26)	0.05	(0.04, 0.12)
(28)	10	MW0	0.68	(0.61, 0.73)	0.26	(0.23, 0.28)	0.36	(0.32, 0.40)	0.44	(0.31, 0.59)	0.30	(0.25, 0.36)
(29)	10	AR(1)	0.72	(0.63, 0.80)	0.46	(0.40, 0.50)	0.43	(0.35, 0.54)	0.65	(0.59, 0.71)	0.62	(0.55, 0.71)
(30)	10	MWd	0.68	(0.65, 0.75)	0.42	(0.39, 0.48)	0.50	(0.39, 0.57)	0.66	(0.64, 0.70)	0.59	(0.56, 0.77)
(31)	10	RW	0.90	(0.88, 0.93)	0.66	(0.62, 0.73)	0.46	(0.35, 0.66)	0.75	(0.72, 0.78)	0.74	(0.62, 0.85)
(32)	10	MW1	0.63	(0.59, 0.69)	0.50	(0.44, 0.55)	0.51	(0.41, 0.57)	0.70	(0.67, 0.74)	0.71	(0.63, 0.81)
(33)	25	iid	0.55	(0.40, 0.63)	0.09	(0.08, 0.11)	0.10	(0.07, 0.14)	0.08	(0.07, 0.11)	0.04	(0.00, 0.08)
(34)	25	MW0	0.59	(0.46, 0.66)	0.21	(0.19, 0.24)	0.25	(0.21, 0.33)	0.21	(0.15, 0.33)	0.12	(0.05, 0.23)
(35)	25	AR(1)	0.63	(0.52, 0.72)	0.30	(0.26, 0.34)	0.25	(0.16, 0.30)	0.56	(0.47, 0.65)	0.59	(0.53, 0.66)
(36)	25	MWd	0.55	(0.49, 0.64)	0.25	(0.22, 0.33)	0.27	(0.23, 0.36)	0.54	(0.46, 0.60)	0.54	(0.51, 0.65)
(37)	25	RW	0.96	(0.94, 0.99)	0.48	(0.36, 0.76)	0.24	(0.16, 0.44)	0.82	(0.77, 0.87)	0.83	(0.68, 0.89)
(38)	25	MW1	0.76	(0.74, 0.85)	0.34	(0.27, 0.42)	0.27	(0.21, 0.44)	0.77	(0.72, 0.82)	0.75	(0.65, 0.85)
(39)	50	iid	0.40	(0.09, 0.57)	0.06	(0.04, 0.12)	0.00	(0.00, 0.06)	0.06	(0.03, 0.19)	0.00	(0.00, 0.06)
(40)	50	MW0	0.44	(0.16, 0.60)	0.12	(0.08, 0.15)	0.07	(0.00, 0.17)	0.19	(0.10, 0.34)	0.00	(0.00, 0.10)
(41)	50	AR(1)	0.54	(0.25, 0.76)	0.26	(0.16, 0.30)	0.26	(0.23, 0.35)	0.44	(0.32, 0.56)	0.38	(0.27, 0.43)
(42)	50	MWd	0.36	(0.20, 0.56)	0.32	(0.23, 0.36)	0.31	(0.22, 0.37)	0.51	(0.30, 0.63)	0.28	(0.28, 0.46)
(43)	50	RW	1.00	(0.99, 1.00)	0.62	(0.43, 0.87)	0.29	(0.25, 0.63)	0.87	(0.85, 0.91)	0.55	(0.50, 0.85)
(44)	50	MW1	0.93	(0.86, 0.96)	0.44	(0.35, 0.52)	0.33	(0.28, 0.56)	0.83	(0.77, 0.85)	0.69	(0.43, 0.76)

Notes:
1. The number of samples is the number of countries times the two sampling schemes (rolling and recursive).
2. In the list of models in column (2), we use the shorthand "RW" for the random walk model.
3. Median refers to the median across the number of samples of a given variable. Consider the figure of 0.65 in row (5), column (3a). Per columns (1) and (2), the iid model is used to produce 34 sets of pseudo out-of-sample forecasts of GDP growth for $h = 10$. Actual coverage rates of forecast intervals with nominal 68 percent coverage are computed for each set of pseudo out-of-sample forecasts. 0.65 is the median value across these 34 coverage rates.
4. IQR is interquartile range. It shows the 25th and 75th percentiles of the coverage rates across the number of samples in rows (3) and (25). These percentiles are computed analogously to the median in the previous note.

Table D.4: Rates at which PITs fall into different intervals

(1)			Plausibly stationary variables					Plausibly non-stationary variables				
(2)	horizon	model	$[0.0, 0.2)$	$[0.2, 0.4)$	$[0.4, 0.6)$	$[0.6, 0.8)$	$[0.8, 1.0]$	$[0.0, 0.2)$	$[0.2, 0.4)$	$[0.4, 0.6)$	$[0.6, 0.8)$	$[0.8, 1.0]$
(3)	10	iid	0.29	0.16	0.15	0.14	0.26	0.42	0.05	0.04	0.04	0.45
(4)	10	MW0	0.22	0.20	0.20	0.17	0.21	0.32	0.13	0.08	0.10	0.38
(5)	10	AR(1)	0.18	0.22	0.23	0.17	0.20	0.26	0.20	0.16	0.13	0.25
(6)	10	MWd	0.15	0.24	0.23	0.18	0.19	0.26	0.22	0.14	0.12	0.26
(7)	10	RW	0.05	0.17	0.58	0.15	0.05	0.19	0.20	0.26	0.15	0.20
(8)	10	MW1	0.16	0.22	0.31	0.18	0.13	0.23	0.20	0.20	0.14	0.23
(9)	25	iid	0.24	0.14	0.12	0.14	0.37	0.38	0.03	0.02	0.03	0.54
(10)	25	MW0	0.19	0.16	0.16	0.18	0.30	0.33	0.07	0.05	0.07	0.48
(11)	25	AR1(1)	0.16	0.18	0.19	0.20	0.27	0.30	0.15	0.10	0.11	0.34
(12)	25	MWd	0.14	0.20	0.19	0.20	0.27	0.30	0.15	0.09	0.12	0.35
(13)	25	RW	0.03	0.15	0.66	0.14	0.03	0.20	0.17	0.23	0.16	0.24
(14)	25	MW1	0.10	0.24	0.36	0.19	0.10	0.23	0.17	0.18	0.15	0.28
(15)	50	iid	0.16	0.08	0.09	0.13	0.53	0.25	0.02	0.02	0.03	0.67
(16)	50	MW0	0.12	0.11	0.12	0.19	0.47	0.22	0.04	0.05	0.05	0.64
(17)	50	AR(1)	0.10	0.12	0.14	0.20	0.43	0.19	0.09	0.10	0.12	0.49
(18)	50	MWd	0.09	0.14	0.13	0.22	0.42	0.19	0.08	0.08	0.15	0.50
(19)	50	RW	0.02	0.11	0.73	0.13	0.01	0.08	0.17	0.24	0.19	0.32
(20)	50	MW1	0.05	0.23	0.41	0.23	0.07	0.10	0.16	0.21	0.17	0.36

Notes:

1. See Table 4.1 in the paper for categorization of variables as plausibly stationary or plausibly non-stationary.

2. In the list of models, we use the shorthand "RW" for the random walk model.

3. Values in the table show the fraction of PITs that fall into each interval. For each collection of variables, plausibly stationary and plausiby non-stationary variables, the values sum to 1 across the rows (aside from rounding).

4. For plausibly stationary variables, the number of PITs used to calculate the rates is 17,134 for $h = 10$, 14,374 for $h = 25$, and 9,774 for $h = 50$. For plausibly non-stationary variables, the number of PITs used to calculate the rates is 8,214 for $h = 10$, 6,894 for $h = 25$, and 4,694 for $h = 50$.

Table D.5: Distance from actual PIT rates to ideal PIT rates

(1) (3)	(1) horizon	(2) model	(3) Plausibly stationary variables	(4) Plausibly non-stationary variables
(4)	10	iid	0.06	0.19
(5)	10	MW0	0.01	0.12
(6)	10	AR(1)	0.02	0.04
(7)	10	MWd	0.03	0.06
(8)	10	RW	0.15	0.02
(9)	10	MW1	0.05	0.02
(10)	25	iid	0.08	0.21
(11)	25	MW0	0.04	0.16
(12)	25	AR(1)	0.03	0.10
(13)	25	MWd	0.03	0.10
(14)	25	RW	0.18	0.03
(15)	25	MW1	0.08	0.04
(16)	50	iid	0.13	0.21
(17)	50	MW0	0.11	0.18
(18)	50	AR(1)	0.09	0.12
(19)	50	MWd	0.10	0.12
(20)	50	RW	0.21	0.06
(21)	50	MW1	0.11	0.07

Notes:

1. See Table 4.1 in the paper for categorization of variables as plausibly stationary or plausibly non-stationary.

2. In the list of models in column (2), we use the shorthand "RW" for the random walk model.

3. For the 5 intervals in Table D.4, let r_k be the value reported for the kth interval in Table D.4 for a given forecast horizon and model. This table shows the distance from those values to 0.2 for each model and forecast horizon, measured with $(1/5)\sum_{k=1}^{5}|r_k - 0.2|$.

Table D.6: CRPSs: medians and IQRs of relative values and fraction with minimum value

	(1)	(2)	(3a)	(3b)	(3c)	(4a)	(4b)	(4c)
(1)			Stationary variable			Non-stationary variables		
(2)			184 samples			88 samples		
(3)	horizon	model	median relative CRPS	IQR	fraction with min CRPS	median relative CRPS	IQR	fraction with min CRPS
(4)	10	iid	1.00	(1.00, 1.00)	0.28	1.00	(1.00, 1.00)	0.00
(5)	10	MW0	0.99	(0.94, 1.01)	0.09	0.87	(0.86, 0.90)	0.06
(6)	10	AR1	0.99	(0.91, 1.04)	0.33	0.57	(0.49, 0.70)	0.32
(7)	10	MWd	0.99	(0.90, 1.07)	0.27	0.62	(0.54, 0.74)	0.03
(8)	10	RW	2.45	(1.37, 3.55)	0.02	0.54	(0.48, 0.70)	0.50
(9)	10	MW1	1.54	(1.17, 2.01)	0.01	0.58	(0.50, 0.74)	0.09
(10)	25	iid	1.00	(1.00, 1.00)	0.25	1.00	(1.00, 1.00)	0.00
(11)	25	MW0	0.97	(0.92, 1.01)	0.29	0.90	(0.88, 0.92)	0.22
(12)	25	AR1	1.00	(0.94, 1.08)	0.27	0.84	(0.72, 0.94)	0.23
(13)	25	MWd	1.03	(0.92, 1.09)	0.17	0.83	(0.71, 0.94)	0.27
(14)	25	RW	3.85	(2.16, 6.89)	0.01	0.89	(0.73, 1.04)	0.22
(15)	25	MW1	2.07	(1.53, 3.01)	0.02	0.92	(0.77, 1.08)	0.07
(16)	50	iid	1.00	(1.00, 1.00)	0.27	1.00	(1.00, 1.00)	0.01
(17)	50	MW0	0.98	(0.91, 1.02)	0.28	0.92	(0.88, 0.94)	0.19
(18)	50	AR1	1.02	(0.93, 1.11)	0.23	0.87	(0.76, 0.99)	0.17
(19)	50	MWd	1.04	(0.93, 1.12)	0.20	0.84	(0.75, 0.93)	0.27
(20)	50	RW	5.60	(2.95, 11.28)	0.01	0.97	(0.77, 1.42)	0.20
(21)	50	MW1	2.85	(1.77, 4.29)	0.01	0.95	(0.76, 1.40)	0.15

Notes:

1. See Table 4.1 in the paper for categorization of variables as plausibly stationary or plausibly non-stationary.
2. In the list of models in column (2), we use the shorthand "RW" for the random walk model.
3. For each model in each sample, CRPS is expressed relative to the CRPS for the iid model in that sample. Medians and IQRs (interquartile ranges) of relative CRPSs were constructed as described in the notes to Table 5.1 in the body of the paper.
4. "fraction with min CRPS" reports the fraction of the samples for a given horizon in which the corresponding model has the lowest CRPS among the six models.

Table D.7: RMSPEs: medians and IQRs of relative values and fraction with minimum value

	(1)	(2)	(3a)	(3b)	(3c)	(4a)	(4b)	(4c)
(1)			Stationary variables			Non-stationary variables		
(2)			184 samples			88 samples		
(3)	horizon	model	median relative RMSPE	IQR	fraction with min RMSPE	median relative RMSPE	IQR	fraction with min RMSPE
(4)	10	iid/MW0	1.00	(1.00, 1.00)	0.40	1.00	(1.00, 1.00)	0.08
(5)	10	AR(1)	1.00	(0.97, 1.06)	0.27	0.71	(0.59, 0.84)	0.26
(6)	10	MWd	1.02	(0.96, 1.08)	0.31	0.74	(0.64, 0.86)	0.10
(7)	10	RW	2.01	(1.37, 2.89)	0.03	0.69	(0.56, 0.88)	0.48
(8)	10	MW1	1.70	(1.32, 2.18)	0.00	0.71	(0.59, 0.97)	0.08
(9)	25	iid/MW0	1.00	(1.00, 1.00)	0.56	1.00	(1.00, 1.00)	0.31
(10)	25	AR(1)	1.01	(0.99, 1.12)	0.24	0.98	(0.91, 1.10)	0.16
(11)	25	MWd	1.04	(1.00, 1.14)	0.18	0.98	(0.86, 1.10)	0.26
(12)	25	RW	2.68	(1.88, 3.93)	0.02	1.11	(0.94, 1.31)	0.23
(13)	25	MW1	2.25	(1.70, 2.94)	0.01	1.14	(0.96, 1.35)	0.05
(14)	50	iid/MW0	1.00	(1.00, 1.00)	0.61	1.00	(1.00, 1.00)	0.59
(15)	50	AR(1)	1.02	(1.00, 1.21)	0.22	1.05	(0.98, 1.26)	0.10
(16)	50	MWd	1.07	(1.01, 1.24)	0.16	1.05	(0.94, 1.16)	0.13
(17)	50	RW	3.36	(2.29, 5.35)	0.01	1.24	(1.02, 1.83)	0.14
(18)	50	MW1	2.94	(2.07, 3.79)	0.00	1.24	(1.02, 1.81)	0.05

Notes:

1. See Table 4.1 in the paper for categorization or variables as plausibly stationary or plausibly non-stationary.
2. In the list of models in column (2), we use the shorthand “RW” for the random walk model.
3. For each model in each sample, RMSPE is expressed relative to the RMSPE for the iid/MW0 model in that sample. Medians and IQRs (interquartile ranges) of the resulting relative RMSPEs were constructed as described in the notes to Table 5.1 in the body of the paper.
4. “fraction with min RMSPE” reports the fraction of the samples for a given horizon in which the corresponding model has the lowest RMSPE among the six models.

Table D.8: MAPEs: medians and IQRs of relative values and fraction with minimum value

	(1)	(2)	(3a)	(3b)	(3c)	(4a)	(4b)	(4c)
(1)			Stationary variables			Non-stationary variables		
(2)			184 samples			88 samples		
(3)	horizon	model	median relative MAPE	IQR	fraction with min MAPE	median relative MAPE	IQR	fraction with min MAPE
(4)	10	iid/MW0	1.00	(1.00, 1.00)	0.26	1.00	(1.00, 1.00)	0.05
(5)	10	AR(1)	0.99	(0.95, 1.02)	0.38	0.64	(0.57, 0.80)	0.30
(6)	10	MWd	1.00	(0.93, 1.07)	0.29	0.73	(0.66, 0.85)	0.02
(7)	10	RW	1.43	(1.10, 2.14)	0.07	0.61	(0.54, 0.77)	0.60
(8)	10	MW1	1.44	(1.16, 1.89)	0.01	0.67	(0.60, 0.85)	0.03
(9)	25	iid/MW0	1.00	(1.00, 1.00)	0.45	1.00	(1.00, 1.00)	0.39
(10)	25	AR(1)	1.01	(0.99, 1.07)	0.27	0.97	(0.88, 1.11)	0.22
(11)	25	MWd	1.03	(0.99, 1.09)	0.24	1.02	(0.87, 1.13)	0.16
(12)	25	RW	2.07	(1.54, 2.94)	0.03	1.07	(0.84, 1.22)	0.22
(13)	25	MW1	1.90	(1.52, 2.45)	0.01	1.10	(0.93, 1.24)	0.02
(14)	50	iid/MW0	1.00	(1.00, 1.00)	0.49	1.00	(1.00, 1.00)	0.47
(15)	50	AR(1)	1.02	(0.99, 1.13)	0.28	1.03	(0.94, 1.24)	0.10
(16)	50	MWd	1.04	(1.00, 1.09)	0.20	1.02	(0.92, 1.09)	0.18
(17)	50	RW	2.47	(1.81, 3.70)	0.02	1.16	(0.96, 1.69)	0.14
(18)	50	MW1	2.42	(1.73, 3.21)	0.01	1.17	(0.98, 1.70)	0.11

Notes:

1. See Table 4.1 in the paper for categorization or variables as plausibly stationary of plausibly non-stationary.

2. In the list of models in column (2), we use the shorthand "RW" for the random walk model.

3. For each model in each sample, MAPE is expressed relative to the MAPE for the iid/MW0 model in that sample. Medians and IQRs (interquartile ranges) of the resulting relative MAPEs were constructed as described in the notes to Table 5.1 in the body of the paper.

4. "fraction with min MAPE" reports the fraction of the samples for a given horizon in which the corresponding model has the lowest MAPE among the six models.

Table D.9: Absolute biases: medians and IQRs of relative values and fraction with minimum value

	(1)	(2)	(3a)	(3b)	(3c)	(4a)	(4b)	(4c)
(1)			Stationary variables			Non-stationary variables		
(2)			184 samples			88 samples		
(3)	horizon	model	median relative bias	IQR	fraction with min bias	median relative bias	IQR	fraction with min bias
(4)	10	iid/MW0	1.00	(1.00, 1.00)	0.09	1.00	(1.00, 1.00)	0.10
(5)	10	AR(1)	0.94	(0.64, 1.12)	0.10	0.33	(0.17, 0.61)	0.15
(6)	10	MWd	0.79	(0.46, 1.27)	0.16	0.37	(0.19, 0.62)	0.18
(7)	10	RW	0.31	(0.11, 0.98)	0.35	0.21	(0.11, 0.47)	0.48
(8)	10	MW1	0.43	(0.17, 1.00)	0.30	0.23	(0.12, 0.53)	0.09
(9)	25	iid/MW0	1.00	(1.00, 1.00)	0.05	1.00	(1.00, 1.00)	0.09
(10)	25	AR(1)	0.96	(0.81, 1.03)	0.07	0.44	(0.14, 0.76)	0.18
(11)	25	MWd	0.79	(0.56, 1.06)	0.14	0.45	(0.18, 0.69)	0.16
(12)	25	RW	0.36	(0.14, 0.81)	0.32	0.35	(0.08, 0.57)	0.35
(13)	25	MW1	0.29	(0.10, 0.81)	0.43	0.35	(0.08, 0.57)	0.22
(14)	50	iid/MW0	1.00	(1.00, 1.00)	0.08	1.00	(1.00, 1.00)	0.27
(15)	50	AR(1)	0.98	(0.92, 1.02)	0.04	0.88	(0.72, 1.15)	0.17
(16)	50	MWd	0.92	(0.78, 1.01)	0.14	0.90	(0.75, 1.03)	0.11
(17)	50	RW	0.58	(0.26, 0.94)	0.30	0.88	(0.65, 1.56)	0.31
(18)	50	MW1	0.51	(0.26, 0.88)	0.43	0.88	(0.65, 1.52)	0.14

Notes:

1. See Table 4.1 in the paper for categorization of variables as plausibly stationary or plausibly non-stationary.

2. In the list of models in column (2), we use the shorthand "RW" for the random walk model.

3. For each model in each sample, absolute bias is expressed relative to the absolute bias for the iid/MW0 model in that sample. Medians and IQRs (interquartile ranges) of relative biases were constructed as described in the notes to Table 5.1 in the body of the paper.

4. "fraction with min bias" reports the fraction of the samples for a given horizon in which the corresponding model has the lowest absolute bias among the six models.

E Derivations of Forecast Distributions for the iid, AR(1), and Random Walk Models

In the paper, we use two sample schemes to estimate the parameters of the forecasting models: a recursive scheme and a rolling scheme. In this appendix, we only show parameter estimates with the recursive sample notation with the understanding that estimates with the rolling sample notation take a parallel form.

The iid Model. The model is $x_t = \mu + u_t$, in which u_t is iid with mean zero and variance σ^2. We use the estimates $\hat{\mu} = \tau^{-1} \sum_{t=1}^{\tau} x_t$ and $\hat{\sigma}^2 = (\tau - 1)^{-1} \sum_{t=1}^{\tau} (x_t - \hat{\mu})^2$. Then, we treat h and τ as sufficiently large so that $h^{1/2}[(x_{\tau+1} + \cdots + x_{\tau+h})/h - \mu]$ and $\tau^{1/2}(\hat{\mu} - \mu)$ are each normally distributed with $h^{1/2}[(x_{\tau+1} + \cdots + x_{\tau+h})/h - \mu] \sim N(0, \sigma^2)$ and $\tau^{1/2}(\hat{\mu} - \mu) \sim N(0, \sigma^2)$. We rearrange terms so that $(x_{\tau+1} + \cdots + x_{\tau+h})/h - \mu \sim N(0, \sigma^2/h)$ and $\hat{\mu} - \mu \sim N(0, \sigma^2/\tau)$.

With u_t being iid, $(x_{\tau+1} + \cdots + x_{\tau+h})/h - \mu$ and $\hat{\mu} - \mu$ are independent, yielding

$$[(x_{\tau+1} + \cdots + x_{\tau+h})/h - \mu] - [\hat{\mu} - \mu] \sim N(0, [(1/h) + (1/\tau)]\sigma^2).$$

Then,

$$(x_{\tau+1} + \cdots + x_{\tau+h})/h \sim N(\hat{\mu}, [(1/h) + (1/\tau)]\sigma^2),$$

and we plug $\hat{\sigma}^2$ in for σ^2 to compute the forecast distribution.

The Random Walk Model. The model is $x_t = x_{t-1} + u_t$, in which u_t is iid with mean zero and variance σ^2. We estimate σ^2 with $\hat{\sigma}^2 = (\tau - 1)^{-1} \sum_{t=2}^{\tau} (x_t - x_{t-1})^2$. It is the case that

$$\begin{aligned}
(x_{\tau+1} + \cdots + x_{\tau+h})/h - x_\tau &= [(x_{\tau+1} - x_\tau) + \cdots + (x_{\tau+h} - x_\tau)]/h \\
&= [u_{\tau+1} + (u_{\tau+1} + u_{\tau+2}) + \cdots + (u_{\tau+1} + \cdots u_{\tau+h})]/h \\
&= hu_{\tau+1}/h + (h-1)u_{\tau+2}/h + \cdots u_{\tau+h}/h \\
&= v_{\tau+1} + v_{\tau+2} + \cdots v_{\tau+h}.
\end{aligned}$$

In the last line, we use $v_{\tau+j} = (h - j + 1)u_{\tau+j}/h$ so that $v_{\tau+j}$ and $v_{\tau+i}$ are independent for $j \neq i$ with $E(v_{\tau+j}) = 0$ and $E(v_{\tau+j}^2) = [(h - j + 1)/h]^2\sigma^2$. Then, we assume that $v_{\tau+j}$ for $j = 1, 2, \ldots$

satisfies Lindeberg's condition and that h is sufficiently large to yield

$$\frac{v_{\tau+1} + v_{\tau+2} + \cdots v_{\tau+h}}{\sqrt{\sum_{j=1}^{h} [(h-j+1)/h]^2 \sigma^2}} \sim N(0,1).$$

Using $\sum_{j=1}^{h} j^2 = h(h+1)(2h+1)/6$ from Equation 16.1.10 in Hamilton (1994), we compute $\sum_{j=1}^{h} [(h-j+1)/h]^2 \sigma^2 = (h+1)(2h+1)\sigma^2/(6h)$ so that $v_{\tau+1} + v_{\tau+2} + \cdots v_{\tau+h} \sim N(0, (h+1)(2h+1)\sigma^2/(6h))$. Hence, $(x_{\tau+1} + \cdots + x_{\tau+h})/h - x_\tau \sim N(0, (h+1)(2h+1)\sigma^2/(6h))$ and

$$(x_{\tau+1} + \cdots + x_{\tau+h})/h \sim N(x_\tau, (h+1)(2h+1)\sigma^2/(6h)),$$

and we plug $\hat{\sigma}^2$ in for σ^2 to compute the forecast distribution.

The AR(1) Model. The model is $x_t = \rho_0 + \rho_1 x_{t-1} + u_t$, in which u_t is iid with mean zero and variance σ^2. We compute $\hat{\rho}_0$ and $\hat{\rho}_1$ with ordinary least squares, suppressing notational dependence on τ for convenience. Ordinary least squares estimates imply

$$\hat{\rho}_0 = \frac{1}{\tau-1} \left(\sum_{t=2}^{\tau} x_t - \hat{\rho}_1 \sum_{t=2}^{\tau} x_{t-1} \right). \tag{E.1}$$

Then, using $x_t = \rho_0 + \rho_1 x_{t-1} + u_t$, we have

$$\hat{\rho}_0 = \rho_0 + (\rho_1 - \hat{\rho}_1) \frac{1}{\tau-1} \sum_{t=2}^{\tau} x_{t-1} + \frac{1}{\tau-1} \sum_{t=2}^{\tau} u_t. \tag{E.2}$$

Next, we bias-adjust the ordinary least squares estimates. Yamamoto and Kunitomo (1984) show that the asymptotic bias of $\hat{\rho}_0$ is $(1+3\rho_1)\rho_0/((\tau-1)(1-\rho_1))$ and that the asymptotic bias of $\hat{\rho}_1$ is $-(1+3\rho_1)/(\tau-1)$. Then, we compute

$$\begin{aligned} \tilde{\rho}_0 &= \hat{\rho}_0 - (1+3\hat{\rho}_1)\hat{\rho}_0/((\tau-1)(1-\hat{\rho}_1)) \\ \tilde{\rho}_1 &= \hat{\rho}_1 + (1+3\hat{\rho}_1)/(\tau-1). \end{aligned}$$

These bias adjustments imply

$$\tilde{\rho}_0 = \left(\frac{1-\tilde{\rho}_1}{1-\hat{\rho}_1} \right) \hat{\rho}_0. \tag{E.3}$$

Hence, the mean of x_t implied by the ordinary least squares estimates, given by $\hat{\rho}_0/(1-\hat{\rho}_1)$, is unchanged by the bias adjustment.

As noted in the paper, we only forecast with the AR(1) model if $\tilde{\rho}_1 < 1$. If $\tilde{\rho}_1 \geq 1$, we forecast with the random walk model. If $\tilde{\rho}_1 < 1$, we compute $\tilde{u}_t = x_t - \tilde{\rho}_0 - \tilde{\rho}_1 x_{t-1}$ and $\tilde{\sigma}^2 = (\tau - 3)^{-1}\sum_{t=2}^{\tau}\tilde{u}_t^2$. Then, we compute the period-by-period forecasts recursively, using $\tilde{x}_{\tau+1} = \tilde{\rho}_0 + \tilde{\rho}_1 x_\tau$ for the one-step-ahead forecast and $\tilde{x}_{\tau+s} = \tilde{\rho}_0 + \tilde{\rho}_1\tilde{x}_{\tau+s-1}$ for the multi-step-ahead forecasts. Hence, we can write the s-step-ahead forecast error as

$$x_{\tau+s} - \tilde{x}_{\tau+s} = \rho_0\sum_{j=0}^{s-1}\rho_1^j - \tilde{\rho}_0\sum_{j=0}^{s-1}\tilde{\rho}_1^j + (\rho_1^s - \tilde{\rho}_1^s)x_\tau + \sum_{j=0}^{s-1}\rho_1^j u_{\tau+s-j}. \tag{E.4}$$

To simplify the analysis, we then assume that $\tilde{\rho}_1 = \rho_1$, yielding

$$x_{\tau+s} - \tilde{x}_{\tau+s} = -\left(\sum_{j=0}^{s-1}\tilde{\rho}_1^j\right)(\tilde{\rho}_0 - \rho_0) + \sum_{j=0}^{s-1}\tilde{\rho}_1^j u_{\tau+s-j}. \tag{E.5}$$

The first term on the right-hand side can then be manipulated as follows

$$\begin{aligned}
-\left(\sum_{j=0}^{s-1}\tilde{\rho}_1^j\right)(\tilde{\rho}_0 - \rho_0) &= -\left(\sum_{j=0}^{s-1}\tilde{\rho}_1^j\right)[(\tilde{\rho}_0 - \hat{\rho}_0) - (\hat{\rho}_0 - \rho_0)] \\
&= \left(\sum_{j=0}^{s-1}\tilde{\rho}_1^j\right)\left[\frac{(\tilde{\rho}_1 - \hat{\rho}_1)\hat{\rho}_0}{1-\hat{\rho}_1} - (\rho_1 - \hat{\rho}_1)\frac{1}{\tau-1}\sum_{t=2}^{\tau}x_{t-1} - \frac{1}{\tau-1}\sum_{t=2}^{\tau}u_t\right] \\
&= \left(\sum_{j=0}^{s-1}\tilde{\rho}_1^j\right)(\tilde{\rho}_1 - \hat{\rho}_1)\left[\frac{\hat{\rho}_0}{1-\hat{\rho}_1} - \frac{1}{\tau-1}\sum_{t=2}^{\tau}x_{t-1}\right] - \left(\sum_{j=0}^{s-1}\tilde{\rho}_1^j\right)\frac{1}{\tau-1}\sum_{t=2}^{\tau}u_t \\
&= \left(\sum_{j=0}^{s-1}\tilde{\rho}_1^j\right)\frac{\tilde{\rho}_1 - \hat{\rho}_1}{1-\hat{\rho}_1}\frac{1}{\tau-1}\sum_{t=2}^{\tau}(x_t - x_{t-1}) - \left(\sum_{j=0}^{s-1}\tilde{\rho}_1^j\right)\frac{1}{\tau-1}\sum_{t=2}^{\tau}u_t \\
&= \left(\sum_{j=0}^{s-1}\tilde{\rho}_1^j\right)\frac{1+3\tilde{\rho}_1}{(1-\tilde{\rho}_1)\tau + (3+\tilde{\rho}_1)}\frac{x_\tau - x_1}{\tau-1} - \left(\sum_{j=0}^{s-1}\tilde{\rho}_1^j\right)\frac{1}{\tau-1}\sum_{t=2}^{\tau}u_t \\
&= \left(\sum_{j=0}^{s-1}\tilde{\rho}_1^j\right)\frac{(1+3\tilde{\rho}_1)(x_\tau - x_1)}{(1-\tilde{\rho}_1)\tau^2 + 2(1+\tilde{\rho}_1)\tau - (3+\tilde{\rho}_1)} - \left(\sum_{j=0}^{s-1}\tilde{\rho}_1^j\right)\frac{1}{\tau-1}\sum_{t=2}^{\tau}u_t,
\end{aligned}$$

in which the second line uses (E.2) to substitute out $\hat{\rho}_0 - \rho_0$ and (E.3) to substitute out $\tilde{\rho}_0$, the third line again imposes $\tilde{\rho}_1 = \rho_1$, the fourth line uses (E.1) to substitute out $\hat{\rho}_0$, the fifth line uses $\hat{\rho}_1 = ((\tau-1)\tilde{\rho}_1 - 1)/(\tau+2)$ to substitute out $\hat{\rho}_1$, and the sixth line rearranges terms. There is a τ^2 term in the denominator of the first term. Because $\tilde{\rho} < 1$, we set the first term to zero, yielding

$$-\left(\sum_{j=0}^{s-1}\tilde{\rho}_1^j\right)(\tilde{\rho}_0 - \rho_0) = -\left(\sum_{j=0}^{s-1}\tilde{\rho}_1^j\right)\frac{1}{\tau-1}\sum_{t=2}^{\tau}u_t,$$

and (E.5) becomes

$$x_{\tau+s} - \tilde{x}_{\tau+s} = -\left(\sum_{j=0}^{s-1}\tilde{\rho}_1^j\right)\left(\frac{1}{\tau-1}\sum_{t=2}^{\tau}u_t\right) + \sum_{j=0}^{s-1}\tilde{\rho}_1^j u_{\tau+s-j}. \tag{E.6}$$

Hence,

$$\begin{aligned}(x_{\tau+1} + \cdots + x_{\tau+h})/h &- (\tilde{x}_{\tau+1} + \cdots + \tilde{x}_{\tau+h})/h \\ &= h^{-1}[1 + (1+\tilde{\rho}_1) + \cdots + (1 + \tilde{\rho}_1 + \cdots + \tilde{\rho}_1^{h-1})]\left(\frac{1}{\tau-1}\sum_{t=2}^{\tau}u_t\right) \\ &\quad + h^{-1}[(1+\tilde{\rho}_1+\cdots+\tilde{\rho}_1^{h-1})u_{\tau+1} + (1+\tilde{\rho}_1+\cdots+\tilde{\rho}_1^{h-2})u_{\tau+2} + \cdots + u_{\tau+h}].\end{aligned} \tag{E.7}$$

For the first term on the right-hand side of (E.7), we use

$$h^{-1}[1 + (1+\tilde{\rho}_1) + \cdots + (1 + \tilde{\rho}_1 + \cdots + \tilde{\rho}_1^{h-1})]\left(\frac{1}{\tau-1}\sum_{t=2}^{\tau}u_t\right) \sim N(0, V_1) \tag{E.8}$$

in which $V_1 = [1 + (1+\tilde{\rho}_1) + \cdots + (1 + \tilde{\rho}_1 + \cdots + \tilde{\rho}_1^{h-1})]^2\sigma^2/((\tau-1)h^2)$.

For the second term on the right-hand side (E.7), we define the new variables $v_{\tau+1} = (1 + \tilde{\rho}_1 + \cdots + \tilde{\rho}_1^{h-1})u_{\tau+1}/h$, $v_{\tau+2} = (1 + \tilde{\rho}_1 + \cdots + \tilde{\rho}_1^{h-2})u_{\tau+2}/h$, and so on. Hence, the second term on the right-hand side of (E.7) becomes $v_{\tau+1} + v_{\tau+2} + \cdots + v_{\tau+h}$, in which $v_{\tau+j}$ and $v_{\tau+i}$ are independent for $j \neq i$ with $E(v_{\tau+j}) = 0$ and $E(v_{\tau+j}^2) = (1 + \tilde{\rho}_1 + \cdots + \tilde{\rho}_1^{h-j})^2\sigma^2/h^2$. Then, we assume that $v_{\tau+j}$ for $j = 1, 2, \ldots$ satisfies Lindeberg's condition and that h is sufficiently large to yield

$$\frac{v_{\tau+1} + v_{\tau+2} + \cdots v_{\tau+h}}{\sqrt{\sum_{j=1}^{h}E(v_{\tau+j}^2)}} \sim N(0,1),$$

implying that

$$h^{-1}[(1+\tilde{\rho}_1+\cdots+\tilde{\rho}_1^{h-1})u_{\tau+1}+(1+\tilde{\rho}_1+\cdots+\tilde{\rho}_1^{h-2})u_{\tau+2}+\cdots+u_{\tau+h}]\sim N(0,V_2), \tag{E.9}$$

in which $V_2=[(1+\tilde{\rho}_1+\cdots+\tilde{\rho}_1^{h-1})^2+(1+\tilde{\rho}_1+\cdots+\tilde{\rho}_1^{h-2})^2+\cdots+1]\sigma^2/h^2$.

Note that the first and second terms on the right-hand side of (E.7) are based on the non-overlapping samples $\{u_2,\ldots,u_\tau\}$ and $\{u_{\tau+1},\ldots,u_{\tau+h}\}$. Because u_t is iid, these two terms are independent and we have

$$(x_{\tau+1}+\cdots+x_{\tau+h})/h-(\tilde{x}_{\tau+1}+\cdots+\tilde{x}_{\tau+h})/h\sim N(0,V), \tag{E.10}$$

in which

$$\begin{aligned} V=&[1+(1+\tilde{\rho}_1)+\cdots+(1+\tilde{\rho}_1+\cdots+\tilde{\rho}_1^{h-1})]^2\sigma^2/(h^2(\tau-1))\\ &+[(1+\tilde{\rho}_1+\cdots+\tilde{\rho}_1^{h-1})^2+(1+\tilde{\rho}_1+\cdots+\tilde{\rho}_1^{h-2})^2+\cdots+1]\sigma^2/h^2. \end{aligned} \tag{E.11}$$

Then, we use $\tilde{x}_{\tau+s}=\tilde{\rho}_0\sum_{j=0}^{s}\tilde{\rho}_1^j+\tilde{\rho}_1^s x_\tau=\tilde{\rho}_0/(1-\tilde{\rho}_1)+\tilde{\rho}_1^s(x_\tau-\tilde{\rho}_0/(1-\tilde{\rho}_1))$ so that we forecast $(x_{\tau+1}+\cdots+x_{\tau+h})/h$ with a Normal distribution with a mean of

$$\frac{\tilde{\rho}_0}{1-\tilde{\rho}_1}+\frac{1}{h}(\tilde{\rho}_1+\tilde{\rho}_1^2+\cdots+\tilde{\rho}_1^h)\left(x_\tau-\frac{\tilde{\rho}_0}{1-\tilde{\rho}_1}\right)$$

and a variance of V in Equation (E.11).

F Details for the Müller and Watson (2016) Models

F.1 Covariance Approximations

The MW forecasting approach uses $\hat{\beta}_{\tau,0}=\tau^{-1}\sum_{t=1}^{\tau}x_t$ and $\hat{\beta}_{\tau,j}=\tau^{-1}\sum_{t=1}^{\tau}\sqrt{2}\cos(\pi j(t-1/2)/\tau)x_t$ for $j=1,\ldots,q$, in which q is much smaller than τ. Write $\hat{\beta}_{\tau,1:q}=[\hat{\beta}_{\tau,1},\ldots,\hat{\beta}_{\tau,q}]'$ as a $(q\times 1)$ vector and $y_{\tau,h}=(x_{\tau+1}+\cdots+x_{\tau+h})/h-\hat{\beta}_{\tau,0}$ as a scalar. Then, $\hat{\beta}_{\tau,1:q}$ and $y_{\tau,h}$ are jointly normally distributed as in Equation (C.1). MW's forecasting approach relies on knowing the form of Σ in Equation (C.1). For the MW0 model, MW provide analytical values for every element of Σ. However, for the MW1 and MWd models, we use numerical approximations from Section 3.2 of

Müller and Watson (2020). To start, let $r = h/\tau$ be the ratio of the forecast horizon to the sample size. We use $N = 1000$ and compute the integer $H = \text{round}(rN)$. Using the notation $\psi_{j,t} = \sqrt{2}\cos(\pi j(t-1/2)/N)$, we write the $(N \times q)$ matrix

$$\Psi = \begin{bmatrix} \psi_{1,1} & \psi_{2,1} & \cdots & \psi_{q,1} \\ \psi_{1,2} & \psi_{2,2} & \cdots & \psi_{q,2} \\ \vdots & \vdots & & \vdots \\ \psi_{1,N} & \psi_{2,N} & \cdots & \psi_{q,N} \end{bmatrix}.$$

Then, we write the $((N+H) \times (q+1))$ matrix

$$\Xi = \begin{bmatrix} \Psi & -\mathbf{1}_{N\times 1} \\ \mathbf{0}_{H\times q} & (N/H)\mathbf{1}_{H\times 1}, \end{bmatrix}$$

in which $\mathbf{1}_{m\times n}$ denotes an $(m \times n)$ matrix of ones and $\mathbf{0}_{m\times n}$ denotes an $(m \times n)$ matrix of zeros. Next, let L be a lower-triangular $((N+H) \times (N+H))$ matrix with ones on and below the main diagonal. Then, we approximate Σ for the MW1 model with

$$\Sigma = \sigma^2_{lrv}(\Xi' L L' \Xi)/N^3,$$

in which σ^2_{lrv} denotes the long-run variance of Δu_t. The distribution of $(x_{\tau+1} + \cdots + x_{\tau+h})/h$ is generalized Student-t with q degrees of freedom and has a location parameter of $\hat{\beta}_{\tau,0} + \Sigma_{y\beta}\Sigma^{-1}_{\beta\beta}\hat{\beta}_{\tau,1:q}$ and a scale parameter of $\sqrt{[\Sigma_{yy} - \Sigma_{y\beta}\Sigma^{-1}_{\beta\beta}\Sigma_{\beta y}](\hat{\beta}'_{\tau,1:q}\Sigma^{-1}_{\beta\beta}\hat{\beta}_{\tau,1:q}/q)}$. Any value of $\sigma^2_{lrv} > 0$ cancels out of both the location and scale parameters; hence, we set $\sigma^2_{lrv} = 1$ and compute

$$\Sigma = (\Xi' L L' \Xi)/N^3, \tag{F.1}$$

for the MW1 model.

For the MWd model, if the value of d is such that $-0.5 < d < 0.5$, define a $((N+H) \times (N+H))$ matrix Λ in which the (i, j) element is given by

$$\lambda_{i,j} = \frac{\Gamma(k+d)\Gamma(1-2d)}{\Gamma(k+1-d)\Gamma(1-d)\Gamma(d)},$$

in which $k = |i-j|$ and $\Gamma(\cdot)$ denotes the gamma function. Then, we set $\sigma^2_{lrv} = 1$ as in the MW1 model[2] and compute

$$\Sigma = (\Xi' \Lambda \Xi)/N^{1+2d}. \tag{F.2}$$

If the value of d is such that $0.5 < d < 1.5$, compute $\tilde{d} = d - 1$ and define a $((N+H) \times (N+H))$ matrix Λ in which the (i,j) element is given by

$$\lambda_{i,j} = \frac{\Gamma(k+\tilde{d})\Gamma(1-2\tilde{d})}{\Gamma(k+1-\tilde{d})\Gamma(1-\tilde{d})\Gamma(\tilde{d})},$$

in which $k = |i-j|$ and $\Gamma(\cdot)$ denotes the gamma function. Then, we set $\sigma^2_{lrv} = 1$ and compute

$$\Sigma = (\Xi' L \Lambda L' \Xi)/N^{1+2d}. \tag{F.3}$$

F.2 The Distribution of the MWd Model

For the MWd model, we treat $d \in (-0.5, 1.5)$ as unknown and use the Bayesian approach in MW. We allow d to take values in a discrete grid, $\mathcal{G} = \{d_1, d_2, \ldots, d_N\}$, and use prior weights, $\{\omega_1, \omega_2, \ldots, \omega_N\}$ subject to $\omega_n \in (0,1)$ for $n = 1, \ldots, N$ and $\sum_{n=1}^{N} \omega_N = 1$. As in MW, we choose $\mathcal{G} = \{-0.4, -0.2, 0, 0.2, 0.4, 0.6, 0.8, 1.0\}$ and our weights are $\omega_n = 1/8$ for $n = 1, \ldots, 8$.

MW redistrict the model so that the forecast densities are invariant to location and scale. This means using $\hat{\beta}^s_{\tau,1:q} = \hat{\beta}_{\tau,1:q}/\sqrt{\hat{\beta}'_{\tau,1:q}\hat{\beta}_{\tau,1:q}}$ to construct the forecast densities instead of just using $\hat{\beta}_{\tau,1:q}$. It also means that the model is set up to initially predict $y^s_{\tau,h} = y_{\tau,h}/\sqrt{\hat{\beta}'_{\tau,1:q}\hat{\beta}_{\tau,1:q}}$ before then making predictions about $y_{\tau,h}$. That is, the Bayes predictive density is constructed to predict $y^s_{\tau,h}$ conditional on $\hat{\beta}^s_{\tau,1:q}$:

$$f^{bayes}(y^s_{\tau,h}|\hat{\beta}^s_{\tau,1:q}) = \frac{\sum_{n=1}^{N} f_{d_n}(\hat{\beta}^s_{\tau,1:q}, y^s_{\tau,h})\omega_n}{\sum_{n=1}^{N} f_{d_n}(\hat{\beta}^s_{\tau,1:q})\omega_n}, \tag{F.4}$$

in which $f_{d_n}(\hat{\beta}^s_{\tau,1:q}, y^s_{\tau,h})$ is the joint density of $\hat{\beta}^s_{\tau,1:q}$ and $y^s_{\tau,h}$ with a covariance matrix associated with fractional integration parameter d_n, $\Sigma(d_n)$, and $f_{d_n}(\hat{\beta}^s_{\tau,1:q})$ is the marginal density of $\hat{\beta}^s_{\tau,1:q}$ implied by $f_{d_n}(\hat{\beta}^s_{\tau,1:q}, y^s_{\tau,h})$. To ease notation going forward, we write $\Sigma_n = \Sigma(d_n)$ for indexing the

[2]For the MWd model, σ^2_{lrv} denotes the long-run variance of $(1-B)^d u_t$ with B being the backshift or lag operator.

covariance matrices for the different values of d in $\mathcal{G}$. Then, the joint density of $\hat{\beta}^s_{\tau,1:q}$ and $y^s_{\tau,h}$ is

$$f_{d_n}(\hat{\beta}^s_{\tau,1:q}, y^s_{\tau,h}) = \frac{1}{2}\pi^{-(q+1)/2}|\Sigma_n|^{-1/2}\Gamma((q+1)/2)\left(\begin{bmatrix}\hat{\beta}^{s\prime}_{\tau,1:q} & y^s_{\tau,h}\end{bmatrix}\Sigma_n^{-1}\begin{bmatrix}\hat{\beta}^s_{\tau,1:q} \\ y^s_{\tau,h}\end{bmatrix}\right)^{-(q+1)/2}, \tag{F.5}$$

in which Γ denotes the gamma function. We write the submatrices of Σ_n as $\Sigma_{n,\beta\beta}$, $\Sigma_{n,y\beta} = \Sigma'_{n,\beta y}$, and $\Sigma_{n,yy}$. Then, the implied marginal density of $\hat{\beta}^s_{\tau,1:q}$ is

$$f_{d_n}(\hat{\beta}^s_{\tau,1:q}) = \frac{1}{2}\pi^{-q/2}|\Sigma_{n,\beta\beta}|^{-1/2}\Gamma(q/2)\left(\hat{\beta}^{s\prime}_{\tau,1:q}\Sigma^{-1}_{n,\beta\beta}\hat{\beta}^s_{\tau,1:q}\right)^{-q/2}. \tag{F.6}$$

We can then compute the maximum likelihood value of d by checking which value of d in $\mathcal{G}$ maximizes $f_{d_n}(\hat{\beta}^s_{\tau,1:q})$ in (F.6).

We re-write Equation (F.4) as

$$f^{bayes}(y^s_{\tau,h}|\hat{\beta}^s_{\tau,1:q}) = \sum_{n=1}^{N}\frac{f_{d_n}(\hat{\beta}^s_{\tau,1:q}, y^s_{\tau,h})}{f_{d_n}(\hat{\beta}^s_{\tau,1:q})}\frac{f_{d_n}(\hat{\beta}^s_{\tau,1:q})\omega_n}{\sum_{k=1}^{N} f_{d_k}(\hat{\beta}^s_{\tau,1:q})\omega_k}.$$

We then use

$$\begin{aligned}|\Sigma_n| &= |\Sigma_{n,\beta\beta}||\Sigma_{n,yy} - \Sigma_{n,y\beta}\Sigma^{-1}_{n,\beta\beta}\Sigma_{n,\beta y}| \\ &= |\Sigma_{n,\beta\beta}|(\Sigma_{n,yy} - \Sigma_{n,y\beta}\Sigma^{-1}_{n,\beta\beta}\Sigma_{n,\beta y}),\end{aligned}$$

where the second line follows because $\Sigma_{n,yy} - \Sigma_{n,y\beta}\Sigma^{-1}_{n,\beta\beta}\Sigma_{n,\beta y}$ is scalar, and

$$\begin{bmatrix}\Sigma_{n,\beta\beta} & \Sigma_{n,\beta y} \\ \Sigma_{n,y\beta} & \Sigma_{n,yy}\end{bmatrix}^{-1} = \begin{bmatrix}\Sigma^{-1}_{n,\beta\beta} + \Sigma^{-1}_{n,\beta\beta}\Sigma_{n,\beta y}\Sigma_{n,y\beta}\Sigma^{-1}_{n,\beta\beta}\nu_n^{-1} & \Sigma^{-1}_{n,\beta\beta}\Sigma_{n,\beta y}\nu_n^{-1} \\ \nu_n^{-1}\Sigma_{n,y\beta}\Sigma^{-1}_{n,\beta\beta} & \nu_n^{-1}\end{bmatrix},$$

in which $\nu_n = \Sigma_{n,yy} - \Sigma_{n,y\beta}\Sigma^{-1}_{n,\beta\beta}\Sigma_{n,\beta y}$. Defining

$$m_n(\hat{\beta}^s_{\tau,1:q}) = \Sigma_{n,y\beta}\Sigma^{-1}_{n,\beta\beta}\hat{\beta}^s_{\tau,1:q}$$

and

$$s_n^2(\hat{\beta}^s_{\tau,1:q}) = (\Sigma_{n,yy} - \Sigma_{n,y\beta}\Sigma^{-1}_{n,\beta\beta}\Sigma_{n,\beta y})(\hat{\beta}^{s\prime}_{\tau,1:q}\Sigma^{-1}_{n,\beta\beta}\hat{\beta}^s_{\tau,1:q})/q,$$

we have

$$\frac{f_{d_n}(\hat{\beta}^s_{\tau,1:q}, y^s_{\tau,h})}{f_{d_n}(\hat{\beta}^s_{\tau,1:q})} = \frac{1}{\sqrt{s^2_n(\hat{\beta}^s_{\tau,1:q})}} \frac{1}{\sqrt{\pi q}} \frac{\Gamma((q+1)/2)}{\Gamma(q/2)} \left(1 + \frac{1}{q} \frac{(y^s_{\tau,h} - m_n(\hat{\beta}^s_{\tau,1:q}))^2}{s^2_n(\hat{\beta}^s_{\tau,1:q})}\right)^{-(q+1)/2}.$$

Hence, $f_{d_n}(\hat{\beta}^s_{\tau,1:q}, y^s_{\tau,h})/f_{d_n}(\hat{\beta}^s_{\tau,1:q})$ is a generalized Student-t density with q degrees of freedom, a location parameter of $m_n(\hat{\beta}^s_{\tau,1:q})$ and a scale parameter of $\sqrt{s^2_n(\hat{\beta}^s_{\tau,1:q})}$. This result then implies that $f^{bayes}(y^s_{\tau,h}|\hat{\beta}^s_{\tau,1:q})$ is a weighted average of generalized Student-t^q densities with weights given by $f_{d_n}(\hat{\beta}^s_{\tau,1:q})\omega_n/(\sum_{k=1}^N f_{d_k}(\hat{\beta}^s_{\tau,1:q})\omega_k)$. Using $\hat{\beta}^s_{\tau,1:q} = \hat{\beta}_{\tau,1:q}/\sqrt{\hat{\beta}'_{\tau,1:q}\hat{\beta}_{\tau,1:q}}$ and $y^s_{\tau,h} = y_{\tau,h}/\sqrt{\hat{\beta}'_{\tau,1:q}\hat{\beta}_{\tau,1:q}}$, we can push the above result further and write

$$\frac{f_{d_n}(\hat{\beta}^s_{\tau,1:q}, y^s_{\tau,h})}{f_{d_n}(\hat{\beta}^s_{\tau,1:q})} = \frac{\sqrt{\hat{\beta}'_{\tau,1:q}\hat{\beta}_{\tau,1:q}}}{\sqrt{s^2_n(\hat{\beta}_{\tau,1:q})}} \frac{1}{\sqrt{\pi q}} \frac{\Gamma((q+1)/2)}{\Gamma(q/2)} \left(1 + \frac{1}{q} \frac{(y_{\tau,h} - m_n(\hat{\beta}_{\tau,1:q}))^2}{s^2_n(\hat{\beta}_{\tau,1:q})}\right)^{-(q+1)/2},$$

so that $f_{d_n}(\hat{\beta}^s_{\tau,1:q}, y^s_{\tau,h})/f_{d_n}(\hat{\beta}^s_{\tau,1:q})$ can be written in terms of $y_{\tau,h}$ and $\hat{\beta}_{\tau,1:q}$. Let $t(y_{\tau,h}, m, s^2, q)$ be the generalized Student-t density with location m, scale s, and degrees of freedom q. Then,

$$\begin{aligned}\frac{f_{d_n}(\hat{\beta}^s_{\tau,1:q}, y^s_{\tau,h})}{f_{d_n}(\hat{\beta}^s_{\tau,1:q})} &= \sqrt{\hat{\beta}'_{\tau,1:q}\hat{\beta}_{\tau,1:q}}\; t(y_{\tau,h}, m_n(\hat{\beta}_{\tau,1:q}), s^2_n(\hat{\beta}_{\tau,1:q}), q) \\ &= \sqrt{\hat{\beta}'_{\tau,1:q}\hat{\beta}_{\tau,1:q}}\; t((x_{\tau+1} + \cdots + x_{\tau+h})/h, \hat{\beta}_{\tau,0} + m_n(\hat{\beta}_{\tau,1:q}), s^2_n(\hat{\beta}_{\tau,1:q}), q),\end{aligned}$$

and we define $f^{bayes}((x_{\tau+1} + \cdots + x_{\tau+h})/h|\hat{\beta}_{\tau,0}, \hat{\beta}_{\tau,1:q})$ as

$$\begin{aligned}&f^{bayes}((x_{\tau+1} + \cdots + x_{\tau+h})/h|\hat{\beta}_{\tau,0}, \hat{\beta}_{\tau,1:q}) \\ &\quad = \frac{1}{\sqrt{\hat{\beta}'_{\tau,1:q}\hat{\beta}_{\tau,1:q}}} f^{bayes}(y^s_{\tau,h}|\hat{\beta}^s_{\tau,1:q}) \\ &\quad = \sum_{n=1}^N t((x_{\tau+1} + \cdots + x_{\tau+h})/h, \hat{\beta}_{\tau,0} + m_n(\hat{\beta}_{\tau,1:q}), s^2_n(\hat{\beta}_{\tau,1:q}), q) \frac{f_{d_n}(\hat{\beta}_{\tau,1:q})\omega_n}{\sum_{k=1}^N f_{d_k}(\hat{\beta}_{\tau,1:q})\omega_k}.\end{aligned} \tag{F.7}$$

Hence, the density of $(x_{\tau+1} + \cdots + x_{\tau+h})/h$ conditional on $\hat{\beta}_{\tau,0}$ and $\hat{\beta}_{\tau,1:q}$ is a weighted average of generalized Student-t^q densities in which the weights are functions of the prior weights and the likelihoods of the values of d, determined by the marginal density in Equation (F.6).

To compute the point forecast, which is the expectation of $(x_{\tau+1} + \cdots + x_{\tau+h})/h$ over $f((x_{\tau+1} +$

$\dots + x_{\tau+h})/h|\hat{\beta}_{\tau,0}, \hat{\beta}_{\tau,1:q})^{bayes}$, we first note that the expectation of $t((x_{\tau+1} + \dots + x_{\tau+h})/h, \hat{\beta}_{\tau,0} + m(\hat{\beta}_{\tau,1:q}, d_n), s^2(\hat{\beta}_{\tau,1:q}, d_n), q)$ is $\hat{\beta}_{\tau,0} + m(\hat{\beta}_{\tau,1:q}, d_n)$. Then, we have

$$\hat{f}_{\tau,h}^{MWd} = \sum_{n=1}^{N} (\hat{\beta}_{\tau,0} + \Sigma_{n,y\beta}\Sigma_{n,\beta\beta}^{-1}\hat{\beta}_{\tau,1:q}) \frac{f_{d_n}(\hat{\beta}_{\tau,1:q})\omega_n}{\sum_{k=1}^{N} f_{d_k}(\hat{\beta}_{\tau,1:q})\omega_k}. \tag{F.8}$$

We compute medians and equal-tailed forecast intervals using the cumulative distribution function (CDF) that corresponds with Equation (F.7). The CDF is

$$\begin{aligned} &F^{bayes}((x_{\tau+1} + \dots + x_{\tau+h})/h|\hat{\beta}_{\tau,0}, \hat{\beta}_{\tau,1:q}) \\ &\quad = \sum_{n=1}^{N} T\left(\frac{(x_{\tau+1} + \dots + x_{\tau+h})/h - \hat{\beta}_{\tau,0} - m_n(\hat{\beta}_{\tau,1:q})}{\sqrt{s_n^2(\hat{\beta}_{\tau,1:q})}}, q \right) \frac{f_{d_n}(\hat{\beta}_{\tau,1:q})\omega_n}{\sum_{k=1}^{N} f_{d_k}(\hat{\beta}_{\tau,1:q})\omega_k}, \end{aligned} \tag{F.9}$$

in which $T(\cdot, q)$ is the CDF for a standard Student-t distribution with q degrees of freedom. Taking $\hat{\beta}_{\tau,0}$ and $\hat{\beta}_{\tau,1:q}$ as given, we use the method of bisection to solve for the values of $(x_{\tau+1}+\dots+x_{\tau+h})/h$ that yield $F^{bayes} = 0.16$, $F^{bayes} = 0.0.5$, and $F^{bayes} = 0.84$.

References

Gneiting, Tilmann. 2011. "Making and Evaluating Point Forecasts." *Journal of the American Statistical Association* 106 (494):746–762. URL https://doi.org/10.1198/jasa.2011.r10138.

Gneiting, Tilmann and Adrian E. Raftery. 2007. "Strictly Proper Scoring Rules, Prediction, and Estimation." *Journal of the American Statistical Association* 102 (477):359–378. URL https://doi.org/10.1198/016214506000001437.

Hamilton, James D. 1994. *Time Series Analysis*. Princeton University Press.

Jordan, Alexander, Fabian Krüger, and Sebastian Lerch. 2019. "Evaluating Probabilistic Forecasts with scoringRules." *Journal of Statistical Software* 90 (12):1–37. URL https://doi.org/10.18637/jss.v090.i12.

Müller, Ulrich K. and Mark W. Watson. 2015. "Low-Frequency Econometrics." NBER Working Paper 21564, URL https://doi.org/10.3386/w21564.

———. 2016. "Measuring Uncertainty about Long-Run Predictions." *Review of Economic Studies* 83 (4):1711–1740. URL https://doi.org/10.1093/restud/rdw003.

———. 2020. "Low-Frequency Analysis of Economic Time Series." Working Paper, Princeton University.

Yamamoto, Taku and Naoto Kunitomo. 1984. "Asymptotic Bias of the Least Squares Estimator for Multivariate Autoregressive Models." *Annals of the Institute of Statistical Mathematics* 36:419–430. URL https://doi.org/10.1007/BF02481980.