Central Bank Credibility and Inflation Expectations: A Microfounded Forecasting Approach – Online Appendix

João Victor Issler (Brazilian School of Economics and Finance FGV EPGE)
Ana Flávia Soares (Brazilian School of Economics and Finance FGV EPGE)

1 The core idea of Gaglianone and Issler (2021)

The following text draws heavily from Gaglianone and Issler (2021), which superseded previous versions. Indeed, we provide below a self-contained summary of the ideas in that paper. For a more complete understanding of the techniques used there, please refer to the original manuscript. The setup is pretty standard and obeys the stationary-ergodic environment with finite first and second moments.

Gaglianone and Issler follow the setup in Patton and Timmermann (2007), where each individual $i = 1, 2, \cdots, N$ chooses the optimal point forecast $f_{i,t}^{h}$ so that the conditional expected loss function (L^i) is minimized:

$$f_{i,t}^{h} = \arg\min_{f_i} \mathbb{E}\left[L^i(y_t; f_i) \,| \mathcal{F}_{i,t-h}\right]. \tag{1}$$

Here, $f_i \in \mathbb{R}$ represents all possible point-forecast choices for forecaster i.

As in Morris and Shin (2002), the conditioning information set $\mathcal{F}_{i,t-h}$ varies across forecasters (*heterogeneous information*) and it is based on both *public* and *private* information. The econometrician has no knowledge of the individual risk function used by forecaster i (*loss-function heterogeneity*), and the assumptions made by the agent on the conditional data generating process (DGP) of y_t used by her/him to compute $\mathbb{E}\left[L^i(y_t; f_i) \,| \mathcal{F}_{i,t-h}\right]$ and therefore to forecast y_t.

Assuming that the optimizing agent i employs the location-scale model with one covariate alone, different across i, forecaster's i model of the conditional quantile function of y_t, $Q_{y_t}(\tau \,| \mathcal{F}_{i,t-h})$ reads as:

$$Q_{y_t}(\tau \,| \mathcal{F}_{i,t-h}) = \alpha_{0,i}^{h}(\tau) + \alpha_{1,i}^{h}(\tau)\, x_{i,t-h}, \text{ for all } \tau \in [0,1], \tag{2}$$

where $x_{i,t-h}$ is the covariate used by agent i in the location-scale model, $\tau \in [0, 1]$ is a quantile of the conditional distribution of y_t, and $\alpha_{0,i}^{h}(\tau)$ and $\alpha_{1,i}^{h}(\tau)$ are parameters that vary across quantiles (τ), agents (i), and forecast horizons (h), but do not vary across time to preserve stationarity. To simplify notation, we drop the dependence of $\alpha_{0,i}^{h}(\tau)$ and $\alpha_{1,i}^{h}(\tau)$ on h in what follows.

They stress that equation (2) is an extreme case that works against *common* information, since each agent employs a *different* covariate $x_{i,t-h}$ in their respective location-scale model (*heterogeneity in information*). Despite that, these covariates $x_{i,t-h}$ are likely to be correlated. They follow Stock and Watson (2002) in assuming that:

$$\begin{aligned} y_t &= \beta_F' F_{t-h} + \beta_w' w_{t-h} + \xi_t, \\ x_{t-h} &= \Lambda F_{t-h} + e_{t-h}, \end{aligned} \tag{3}$$

where $x_{t-h} = (x_{1,t-h}, x_{2,t-h}, \cdots, x_{N,t-h})'$ stacks all the covariates $x_{i,t-h}$ used by the agents to forecast y_t; F_{t-h} stacks a reduced number of r ($r \ll N$) latent factors capturing the common components of these covariates and generating their cross-correlation; w_{t-h} stacks additional regressors used to forecast y_t, e.g., its lags; the Λ matrix and the vector β_F' store factor loadings; the vector β_w' stores the loadings of the additional regressors; and ξ_t and e_{t-h} are errors terms which are allowed to be both serially correlated and (weakly) cross-sectionally correlated.

An important result from Patton and Timmermann (2007, Proposition 3, case b) regarding an optimal point forecast $f_{i,t}^{h}$ is that it is associated with a given quantile level of $Q_{y_t}(\tau | \mathcal{F}_{i,t-h})$, labelled here as τ_i^*, which may differ across agents, as follows:

$$f_{i,t}^{h} = Q_{y_t}(\tau_i^* | \mathcal{F}_{i,t-h}) = F_{i,t|t-h}^{-1}(\tau_i^*). \tag{4}$$

In words, the optimal point forecast $f_{i,t}^{h}$ of agent i is a specific conditional quantile of y_t (i.e., associated with quantile level $\tau_i^* \in [0,1]$), where $F_{i,t|t-h}(\cdot)$ is the conditional cumulative distribution function of y_t. Intuitively, one can think of specialization in forecasting: one agent is always pessimistic about future prospects of y_t, while another one is optimistic, and a third one is neutral, and so forth.

The key to the result in (4) is that each point forecast lies in the domain of the conditional cumulative distribution, so that the conditional probability of $y_t \leq f_{i,t}^{h}$ could be computed and $F_{i,t|t-h}(\cdot)$ inverted. Given (2) and (4), we obtain:

$$f_{i,t}^{h} = Q_{y_t}(\tau_i^* | \mathcal{F}_{i,t-h}) = \alpha_{0,i}(\tau_i^*) + \alpha_{1,i}(\tau_i^*) x_{i,t-h}. \tag{5}$$

Notice that this establishes an affine relationship between the covariate in the location-scale model and the optimal individual forecast. However, for every respondent, we can go one step further to relate the latter with the conditional mean of y_t, $\mathbb{E}(y_t | \mathcal{F}_{i,t-h}) \equiv \mathbb{E}_{i,t-h}(y_t)$, given all the information available to agent i – public and private. This could be accomplished by using an important result in Koenker (2005), which relates the conditional quantile function with the conditional mean $\mathbb{E}_{i,t-h}(y_t)$, as follows:

$$\begin{aligned}\mathbb{E}\left(y_t\,|\mathcal{F}_{i,t-h}\right) &\equiv \mathbb{E}_{i,t-h}(y_t) = \int_0^1 Q_{y_t}\left(\tau\,|\mathcal{F}_{i,t-h}\right)d\tau = \int_0^1 \left[\alpha_{0,i}\left(\tau\right)+\alpha_{1,i}\left(\tau\right)x_{i,t-h}\right]d\tau \\ &= \int_0^1 \alpha_{0,i}\left(\tau\right)d\tau + x_{i,t-h}\int_0^1 \alpha_{1,i}\left(\tau\right)d\tau = \overline{\alpha_{0,i}}+\overline{\alpha_{1,i}}x_{i,t-h},\end{aligned} \tag{6}$$

where $\overline{\alpha_{0,i}} = \int_0^1 \alpha_{0,i}\left(\tau\right)d\tau$ and $\overline{\alpha_{1,i}} = \int_0^1 \alpha_{1,i}\left(\tau\right)d\tau$. Combining (5) and (6), we arrive at:

$$f_{i,t}^h = \left(\alpha_{0,i}\left(\tau_i^*\right) - \frac{\alpha_{1,i}\left(\tau_i^*\right)\overline{\alpha_{0,i}}}{\overline{\alpha_{1,i}}}\right) + \frac{\alpha_{1,i}\left(\tau_i^*\right)}{\overline{\alpha_{1,i}}}\mathbb{E}_{i,t-h}(y_t), \tag{7}$$

which establishes an affine relationship between the optimal *individual* forecast $f_{i,t}^h$ and the individual conditional expectation $\mathbb{E}_{i,t-h}(y_t)$ of agent i.

They turn to the components of the individual information sets used by every agent. The information set $\mathcal{F}_{i,t-h}$ is partitioned here into two orthogonal components. The first is $\mathcal{F}_{t-h}$, which comprises *public* information available to all agents. The second is $\mathcal{F}_{i,t-h}^{priv}$, which includes idiosyncratic or *private* information available to each agent exclusively. Formally, Gaglianone and Issler impose that $\mathcal{F}_{t-h}\cup\mathcal{F}_{i,t-h}^{priv}=\mathcal{F}_{i,t-h}$ and that $\mathcal{F}_{t-h}\cap\mathcal{F}_{i,t-h}^{priv}=\emptyset$.

Next, they present an important result, widely used in the *Nowcasting* literature (e.g., Bańbura, Giannone, and Reichlin, 2011): the orthogonal-component decomposition[1] of the information set $\mathcal{F}_{i,t-h}$.

Lemma 1 (Gaglianone and Issler (2021)) *Let $(\Omega,\mathcal{F},P)$ be a probability space, where $\mathcal{F}_{i,t-h}\subseteq\mathcal{F}$ is a sub σ-algebra of $\mathcal{F}$. Assume that $\mathcal{F}_{t-h}$ and $\mathcal{F}_{i,t-h}^{priv}$ are orthogonal, closed and non-empty subspaces that form a partition of $\mathcal{F}_{i,t-h}$, $\mathcal{F}_{t-h}\cup\mathcal{F}_{i,t-h}^{priv}=\mathcal{F}_{i,t-h}$ and $\mathcal{F}_{t-h}\cap\mathcal{F}_{i,t-h}^{priv}=\emptyset$, where $\mathcal{F}_{t-h}$ only contains public (common) information, including a constant term, and $\mathcal{F}_{i,t-h}^{priv}$ only contains private idiosyncratic information available exclusively to agent i. Assume that y_t is in the Hilbert space of square-integrable real random variables. Then, one can decompose $\mathbb{E}\left(y_t\mid\mathcal{F}_{i,t-h}\right)$ as follows:*

$$(i)\ \mathbb{E}\left(y_t\mid\mathcal{F}_{i,t-h}\right)=\mathbb{E}\left(y_t\mid\mathcal{F}_{t-h}\right)+\mathbb{E}\left(y_t\mid\mathcal{F}_{i,t-h}^{priv}\right),$$

where the second term $\mathbb{E}\left(y_t\mid\mathcal{F}_{i,t-h}^{priv}\right)$ is orthogonal to the information used in the first, i.e.,

$$(ii)\ \mathbb{E}\left(\mathbb{E}\left(y_t\mid\mathcal{F}_{i,t-h}^{priv}\right)\mid\mathcal{F}_{t-h}\right)=0.$$

[1] A simple example of this decomposition relates to the linear regression model with Gaussian errors: $y=X\beta+\varepsilon$, where the OLS estimator decomposes y as: $y=\mathbb{E}\left(y\mid X\cup X^{\perp}\right)=\mathbb{E}\left(y\mid X\right)+\mathbb{E}\left(y\mid X^{\perp}\right)=X\widehat{\beta}+\widehat{\varepsilon}$ $=\underbrace{X\left(X'X\right)^{-1}X'y}_{\mathbb{E}(y|X)}+\underbrace{\left[I-X\left(X'X\right)^{-1}X'\right]y}_{\mathbb{E}(y|X^{\perp})}$, the first term being a projection of y onto the space of X and the second being a projection of y onto the orthogonal space of X, labelled here as $X^{\perp}$. Note that X must contain a constant term to impose a zero mean for $\widehat{\varepsilon}$.

Combining (7) and result (i) of Lemma 1, we arrive at:

$$f_{i,t}^{h}=\left(\alpha_{0,i}\left(\tau_{i}^{*}\right)-\frac{\alpha_{1,i}\left(\tau_{i}^{*}\right)\overline{\alpha_{0,i}}}{\overline{\alpha_{1,i}}}\right)+\frac{\alpha_{1,i}\left(\tau_{i}^{*}\right)}{\overline{\alpha_{1,i}}}\mathbb{E}_{t-h}(y_t)+\frac{\alpha_{1,i}\left(\tau_{i}^{*}\right)}{\overline{\alpha_{1,i}}}\mathbb{E}\left(y_t\mid\mathcal{F}_{i,t-h}^{priv}\right). \tag{8}$$

Thus, the optimal forecasts $f_{i,t}^{h}$ takes the encompassing general form:

$$f_{i,t}^{h}=k_i^h+\beta_i^h\mathbb{E}_{t-h}(y_t)+\varepsilon_{i,t}^{h}, \tag{9}$$

where a full mapping from (8) to (9) is straightforward to obtain using the definitions below:

$$k_i^h \equiv \alpha_{0,i}\left(\tau_{i}^{*}\right)-\frac{\alpha_{1,i}\left(\tau_{i}^{*}\right)\overline{\alpha_{0,i}}}{\overline{\alpha_{1,i}}}, \tag{10}$$

$$\beta_i^h \equiv \frac{\alpha_{1,i}\left(\tau_{i}^{*}\right)}{\overline{\alpha_{1,i}}}, \tag{11}$$

$$\varepsilon_{i,t}^{h} \equiv \frac{\alpha_{1,i}\left(\tau_{i}^{*}\right)}{\overline{\alpha_{1,i}}}\mathbb{E}\left(y_t\mid\mathcal{F}_{i,t-h}^{priv}\right), \tag{12}$$

in which k_i^h and β_i^h are, respectively, intercept and slope bias-correction terms, and $\varepsilon_{i,t}^{h}$ is the idiosyncratic component of individual expectations.

Equation (9) delivers an affine factor model, with a single factor related to the common information set used by all forecasters in the survey, $\mathbb{E}_{t-h}\left(y_t\right)\equiv\mathbb{E}\left(y_t\mid\mathcal{F}_{t-h}\right)$. Of course, the individual response to common information is heterogeneous, since the affine model has an intercept and a slope that vary across i. Note that heterogeneity in expectations appears only in the last term $\varepsilon_{i,t}^{h}\equiv\frac{\alpha_{1,i}\left(\tau_{i}^{*}\right)}{\overline{\alpha_{1,i}}}\mathbb{E}\left(y_t\mid\mathcal{F}_{i,t-h}^{priv}\right)$ of the orthogonal decomposition. Since $\varepsilon_{i,t}^{h}$ and $\mathbb{E}_{t-h}(y_t)$ are orthogonal in (9), identification is easily achieved and, given a consistent estimate of $\mathbb{E}_{t-h}(y_t)$, we can estimate k_i^h, β_i^h, and $\varepsilon_{i,t}^{h}$ using least-square methods.

Next, Proposition 2 summarizes the core idea discussed above and provides a formal treatment on this subject.

Proposition 2 (Gaglianone and Issler (2021)) *Under assumptions A1-A5 in Gaglianone and Issler (2021): (i) the optimal forecast is an affine function of the conditional mean of y_t, using public (common) information, that is:*

$$f_{i,t}^{h}=k_i^h+\beta_i^h\mathbb{E}_{t-h}(y_t)+\varepsilon_{i,t}^{h},$$

(ii) in the absence of scale effects on the location-scale model of y_t, then $\beta_i^h=1$, for all i, and

$$f_{i,t}^{h}=k_i^h+\mathbb{E}_{t-h}(y_t)+\varepsilon_{i,t}^{h}.$$

1.1 Identification and GMM estimation

The basic approach to identify and estimate $\mathbb{E}_{t-h}(y_t)$ employs the generalized method of moments (GMM), relying on T asymptotics, which fits current surveys of expectations. However, the fact that $\mathbb{E}_{t-h}(y_t)$ is a latent variable is a drawback, since the moments used in GMM estimation must be a function of observables and parameters alone. However, one can use the decomposition:

$$y_t = \mathbb{E}_{t-h}(y_t) - \eta_t^h, \tag{13}$$

where, vis-a-vis public information, η_t^h is an unforecastable martingale-difference component, i.e., $\mathbb{E}_{t-h}(\eta_t^h) = 0$.

Combining (9) and (13) leads to:

$$\begin{aligned} f_{i,t}^h &= k_i^h + \beta_i^h(y_t + \eta_t^h) + \varepsilon_{i,t}^h \quad (14)\\ &= k_i^h + \beta_i^h y_t + v_{i,t}^h, \quad (15)\end{aligned}$$

where

$$v_{i,t}^h \equiv \beta_i^h \eta_t^h + \varepsilon_{i,t}^h$$

is a composite error term. Notice that, by construction, $\mathbb{E}_{t-h}\left(\eta_t^h\right) = 0$, so $\mathbb{E}_{t-h}\left(v_{i,t}^h\right) = 0$ if $\mathbb{E}_{t-h}(\varepsilon_{i,t}^h) = 0$. However, using the definition of $\varepsilon_{i,t}^h$:

$$\varepsilon_{i,t}^h \equiv \frac{\alpha_{1,i}\left(\tau_i^*\right)}{\overline{\alpha_{1,i}}}\mathbb{E}\left(y_t \mid \mathcal{F}_{i,t-h}^{priv}\right),$$

and the result (ii) from Lemma 1, it follows that $\varepsilon_{i,t}^h$ is orthogonal to public information. Then:

$$\begin{aligned} \mathbb{E}_{t-h}(\varepsilon_{i,t}^h) &= \mathbb{E}\left(\varepsilon_{i,t}^h \mid \mathcal{F}_{t-h}\right) = \frac{\alpha_{1,i}\left(\tau_i^*\right)}{\overline{\alpha_{1,i}}}\mathbb{E}\left(\mathbb{E}\left(y_t \mid \mathcal{F}_{i,t-h}^{priv}\right) \mid \mathcal{F}_{t-h}\right) = 0, \text{ implying that}\\ \mathbb{E}_{t-h}\left(v_{i,t}^h\right) &= 0, \mathbb{E}(\varepsilon_{i,t}^h) = 0, \text{ and finally that } \mathbb{E}\left(v_{i,t}^h\right) = 0. \quad (16)\end{aligned}$$

Equation (16) validates the use of *public information* dated $t-h$ as natural instruments in a GMM setup. Starting with (15) and (16), and using the law of iterated expectations and valid observable instruments z_{t-s}, where $s \geq h$, we obtain:

$$\mathbb{E}\left[\left(f_{i,t}^h - k_i^h - \beta_i^h y_t\right) \otimes z_{t-s}\right] = 0, \tag{17}$$

which is valid for all $i = 1, ..., N$, $t = 1, ..., T$, and $h = 1, ..., H$. The system (17) has $2NH$ parameters and (at least) $2NH$ moment conditions, provided that $\dim(z_{t-s}) > 2$, which is critical for over-identification. Despite that, one problem remains: if $N \to \infty$, the amount of parameters in (17) diverges, which goes against consistency.

To eliminate the *curse of dimensionality* problem, we take the cross-sectional averages of terms $f_{i,t}^h - k_i^h - \beta_i^h y_t$, leading to the following moment restrictions:

$$\mathbb{E}\left[\left(\overline{f_{\cdot,t}^h} - \overline{k^h} - \overline{\beta^h} y_t\right) \otimes z_{t-s}\right] = 0, \tag{18}$$

$t = 1, ..., T$, and $h = 1, ..., H$, where $\overline{f_{\cdot,t}^h} = \frac{1}{N}\sum_{i=1}^{N} f_{i,t}^h$, $\overline{k^h} = \frac{1}{N}\sum_{i=1}^{N} k_i^h$ and $\overline{\beta^h} = \frac{1}{N}\sum_{i=1}^{N} \beta_i^h$, represent cross-sectional averages for each h.

it is straightforward to show identification in a GMM context. For every instrument in z_{t-s}, say, $z_{j,t-s}$, $j = 1, 2, \cdots, k$, where $k > 2$, we can solve:

$$\mathbb{E}_{t-h}\left[\left(\overline{f_{\cdot,t}^h} - \overline{k^h} - \overline{\beta^h} y_t\right) \otimes z_{t-s}\right] = 0,$$

to obtain:

$$\underset{k\times 1}{0} = \mathbb{E}_{t-h}\left[\left(\begin{array}{c} \left(\overline{f_{\cdot,t}^h} - \overline{k^h} - \overline{\beta^h} y_t\right) \times z_{1,t-s} \\ \left(\overline{f_{\cdot,t}^h} - \overline{k^h} - \overline{\beta^h} y_t\right) \times z_{2,t-s} \\ \vdots \\ \left(\overline{f_{\cdot,t}^h} - \overline{k^h} - \overline{\beta^h} y_t\right) \times z_{k,t-s} \end{array}\right)\right] = \left[\left(\begin{array}{c} \left(\overline{f_{\cdot,t}^h} - \overline{k^h} - \overline{\beta^h}\mathbb{E}_{t-h}(y_t)\right) \times z_{1,t-s} \\ \left(\overline{f_{\cdot,t}^h} - \overline{k^h} - \overline{\beta^h}\mathbb{E}_{t-h}(y_t)\right) \times z_{2,t-s} \\ \vdots \\ \left(\overline{f_{\cdot,t}^h} - \overline{k^h} - \overline{\beta^h}\mathbb{E}_{t-h}(y_t)\right) \times z_{k,t-s} \end{array}\right)\right], \tag{19}$$

since the average forecast, $\overline{f_{\cdot,t}^h}$, and all the instruments $z_{j,t-s}$, $j = 1, 2, \cdots, k$, are measurable regarding information in $\mathcal{F}_{t-h}$. If we divide every equation in the system (19) by its respective instrument, and then solve $\mathbb{E}_{t-h}(y_t)$, we obtain the same result for all of them:

$$\mathbb{E}_{t-h}(y_t) = \frac{\overline{f_{\cdot,t}^h} - \overline{k^h}}{\overline{\beta^h}} = \frac{1}{N}\sum_{i=1}^{N} \frac{f_{i,t}^h - \overline{k^h}}{\overline{\beta^h}},$$

which shows identification after assuming standard conditions in the literature. We now discuss consistent GMM estimation based on (18).

Proposition 3 (Gaglianone and Issler (2021)) *Under assumptions A1-A9 in Gaglianone and Issler (2021), the Extended BCAF (Bias Corrected Average Forecast)* $\frac{1}{N}\sum_{i=1}^{N} \frac{f_{i,t}^h - \widehat{\overline{k^h}}}{\widehat{\overline{\beta^h}}}$, *based on T-consistent GMM estimates* $\widehat{\overline{\theta^h}} = \left[\widehat{\overline{k^h}}; \widehat{\overline{\beta^h}}\right]'$, *obeys* $\underset{T\to\infty}{plim}\left(\frac{1}{N}\sum_{i=1}^{N} \frac{f_{i,t}^h - \widehat{\overline{k^h}}}{\widehat{\overline{\beta^h}}}\right) = \mathbb{E}_{t-h}(y_t)$, *where we let $T \to \infty$, with N fixed. Convergence to $\mathbb{E}_{t-h}(y_t)$ also happens when we let first $T \to \infty$ and then let $N \to \infty$, that is,* $\underset{(T,N\to\infty)_{\text{seq}}}{plim}\left(\frac{1}{N}\sum_{i=1}^{N} \frac{f_{i,t}^h - \widehat{\overline{k^h}}}{\widehat{\overline{\beta^h}}}\right) = \mathbb{E}_{t-h}(y_t)$, *where* $(T, N \to \infty)_{\text{seq}}$ *denotes the sequential asymptotic approach of Phillips and Moon (1999).*

Based on this result, a T-consistent estimate of $\mathbb{E}_{t-h}(y_t)$ is given by:

$$\widehat{\mathbb{E}}_{t-h}(y_t) = \frac{1}{N}\sum_{i=1}^{N}\frac{f_{i,t}^h - \widehat{\overline{k}^h}}{\widehat{\overline{\beta}^h}}.$$

2 A Robust (HAC) estimation for the variance of $\widehat{\mathbb{E}}_{t-h}(y_t)$

This is one of the original contributions of this paper. As shown above we can count on consistent GMM estimation based on (18). In our paper, we explain at some length how to obtain a robust heteroskedasticity and autocorrelation consistent (HAC) estimator for the variance of $\widehat{\mathbb{E}}_{t-h}(y_t)$. Here, we summarize that discussion. Our strategy was to first show how to get an estimate of the long-run variance of the population moments used in GMM estimation. Based on that, we apply the Delta Method to get a robust HAC estimator for the variance of $\widehat{\mathbb{E}}_{t-h}(y_t)$.

The population moment condition is given by:

$$0 = \mathbb{E}\left[h_t\left(\overline{\theta}_0^h, W_t\right)\right] = \mathbb{E}\left[\left(\overline{f}_{\cdot,t}^h - \overline{k}^h - \overline{\beta}^h y_t\right) \otimes z_{t-s}\right] \tag{20}$$

for each horizon $h = 1, 2, \cdots, H$,where $\overline{\theta}_0^h = (\overline{k_0^h}, \overline{\beta_0^h})$ is the true parameter value and W_t stacks observables $\left(\overline{f}_{\cdot,t}^h, y_t\right)'$. We use the well-known result regarding long-run variances (LRV):

$$\frac{1}{\sqrt{T}}\sum_{i=1}^{T} h_t\left(\overline{\theta}_0^h, W_t\right) \xrightarrow{d} \mathcal{N}[0, \mathrm{LRV}] = \mathcal{N}\left[0, S^h\right].$$

The sample mean counterpart of (20) is:

$$\frac{1}{T}\sum_{t=1}^{T} h_t\left(\widehat{\overline{\theta}^h}, W_t\right) = \frac{1}{T}\sum_{t=1}^{T} h_t\left(\widehat{\overline{k}}^h, \widehat{\overline{\beta}}^h, W_t\right).$$

The sample counterpart HAC covariance estimator of S^h is:

$$\widehat{S^h} = \widehat{\Gamma}_0(\widehat{\overline{\theta}^h}) + \sum_{j=1}^{l}\kappa(j,l)\left[\widehat{\Gamma}_j(\widehat{\overline{\theta}^h}) + \widehat{\Gamma}_j'(\widehat{\overline{\theta}^h})\right],$$

where $\widehat{\Gamma}_j(\widehat{\overline{\theta}}^h) = T^{-1}\sum_{t=j+1}^{T} h_t(\widehat{\overline{\theta}^h})h_{t-j}'(\widehat{\overline{\theta}^h})$ is the j-th sample autocorrelation of $h_t(\cdot)$, where $\kappa(j,l)$ is the kernel function weight and l is the bandwidth parameter. Using the asymptotic results for the LRV, we obtain the HAC estimate for variance-covariance matrix of $\widehat{\overline{\theta}^h} = \left[\widehat{\overline{k}^h}, \widehat{\overline{\beta}^h}\right]'$:

$$\sqrt{T}\left(\widehat{\overline{\theta}}^h - \overline{\theta}_0^h\right) \xrightarrow{d} \mathcal{N}[0, V_1],$$

where:

$$V_1 = \left[\frac{\partial h_t(\overline{\theta^h})}{\partial \overline{\theta^h}}' \left(S^h\right)^{-1} \frac{\partial h_t(\overline{\theta^h})}{\partial \overline{\theta^h}} \right]^{-1}. \tag{21}$$

Note that,

$$\widehat{\mathbb{E}}_{t-h}\left(y_t\right) = f(\widehat{\overline{k^h}}, \widehat{\overline{\beta^h}}) = \frac{1}{N}\sum_{i=1}^{N} \frac{f_{i,t}^h - \widehat{\overline{k^h}}}{\widehat{\overline{\beta^h}}},$$

is a continuous function of $\widehat{\overline{\theta^h}} = \left[\widehat{\overline{k^h}}, \widehat{\overline{\beta^h}}\right]'$. The Delta method can be applied to find the asymptotic variance of $\widehat{\mathbb{E}}_{t-h}\left(y_t\right)$. Indeed, we have:

$$\begin{aligned} &\sqrt{T}(\widehat{\mathbb{E}}_{t-h}\left(y_t\right) - \mathbb{E}_{t-h}\left(y_t\right)) \xrightarrow{d} \mathcal{N}\left[0, V\right], \text{ or,} \\ &\widehat{\mathbb{E}}_{t-h}\left(y_t\right) \overset{\text{asy}}{\sim} \mathcal{N}\left[\mathbb{E}_{t-h}\left(y_t\right), \frac{V}{T}\right], \end{aligned}$$

with $V = D(f(\overline{\theta^h}))'V_1 D(f(\overline{\theta^h}))$, where $D(f(\overline{\theta^h}))$ is the Jacobian of $f(\overline{\theta^h})$ and V_1 is defined as in equation (21).

Our final goal now is to get an estimate of V. This implies constructing directly $D(f(\widehat{\overline{\theta^h}}))'\widehat{V}D(f(\widehat{\overline{\theta^h}}))$, computing the Jacobian and then evaluating the whole expression using $\widehat{\overline{\theta^h}} = \left[\widehat{\overline{k^h}}; \widehat{\overline{\beta^h}}\right]'$ as follows:

$$\widehat{V} = \left[\begin{array}{cc} -\frac{1}{\widehat{\overline{\beta^h}}} & \frac{1}{N}\sum_{i=1}^{N} \frac{-f_{i,t}^h + \widehat{\overline{k^h}}}{\left(\widehat{\overline{\beta^h}}\right)^2} \end{array}\right] \left[\frac{\partial h_t(\widehat{\overline{\theta^h}})}{\partial \widehat{\overline{\theta^h}}}' \left(\widehat{S^h}\right)^{-1} \frac{\partial h_t(\widehat{\overline{\theta^h}})}{\partial \widehat{\overline{\theta^h}}} \right]^{-1} \left[\begin{array}{c} -\frac{1}{\widehat{\overline{\beta^h}}} \\ \frac{1}{N}\sum_{i=1}^{N} \frac{-f_{i,t}^h + \widehat{\overline{k^h}}}{\left(\widehat{\overline{\beta^h}}\right)^2} \end{array}\right].$$

Once we obtain $\widehat{V}$, we can test credibility of central banks using asymptotic confidence intervals for $\widehat{\mathbb{E}}_{t-h}\left(y_t\right)$. A Central Bank is credible if the fixed Inflation Target π_t^* obeys:

$$\pi_t^* \in \left[\widehat{\mathbb{E}}_{t-h}\left(y_t\right) - 1.96 \times \left(\frac{\widehat{V}}{T}\right)^{1/2}, \quad \widehat{\mathbb{E}}_{t-h}\left(y_t\right) + 1.96 \times \left(\frac{\widehat{V}}{T}\right)^{1/2}\right],$$

i.e., a central bank is credible if the 95% confidence interval around $\widehat{\mathbb{E}}_{t-h}\left(y_t\right)$ (expected beliefs) contains the target.

We now turn to how to construct a credibility index for central banks. The cumulative distribution of $\widehat{\mathbb{E}}_{t-h}\left(y_t\right)$, defined as $F(x_t) = \Phi\left(x_t\right)$, where $\Phi\left(\cdot\right)$ is the CDF of a Normal distribution. In our case, we employ the $\mathcal{N}\left(\mathbb{E}_{t-h}\left(y_t\right), \frac{\widehat{V}}{T}\right)$ distribution and use it to construct a credibility index (CI_t). It will inherit all the advantages of our measure of credibility. Our proposed credibility index (CI_t) is as follows:

$$CI_t = 1 - \frac{\left| F\left(\widehat{\mathbb{E}}_{t-h}\left(y_t\right)\right) - F\left(\pi_t^*\right)\right|}{1/2}, \qquad -\infty < \pi_t^* < \infty.$$

A plot of the Normal density with the credibility index shows how it works in practice. Targets equal to $\widehat{\mathbb{E}}_{t-h}\left(y_t\right)$ generate full credibility, i.e., $CI_t = 1.0$. The further away π_t^* is from $\widehat{\mathbb{E}}_{t-h}\left(y_t\right)$ – no matter to which side of the density – the smaller the credibility index will be. In the limit – whatever the side – we will have $CI_t = 0$.

References

[1] Bańbura, M., Giannone, D., Reichlin, L., 2011. *Nowcasting*, Chapter 7 in *The Oxford Handbook of Economic Forecasting*, Edited by Michael P. Clements and David F. Hendry, pp. 193-224.

[2] Gaglianone, W.P., Issler, J.V., 2021. Microfounded Forecasting.

[3] Koenker, R., 2005. *Quantile Regression*. Cambridge University Press.

[4] Morris, S. and Shin, H.S., 2002. Social Value of Public Information, *American Economic Review*, Vol. 92(5), pp. 1521-1534.

[5] Patton, A.J., Timmermann, A., 2007. Testing Forecast Optimality under Unknown Loss. *Journal of the American Statistical Association* 102 (480), 1172-1184.

[6] Stock, J., Watson, M., 2002. Forecasting Using Principal Components from a Large Number of Predictors. *Journal of the American Statistical Association* 97 (460), 1167-1179.