Measuring Uncertainty about Long-Run Predictions

Ulrich K. Müller and Mark W. Watson
Princeton University

Presentation at ECB

June 14, 2014
Set-up

- Observe data \( x_t, \ t = 1, \cdots, T \), such as growth rates of GDP, or inflation

- Want to forecast average over the next \( \lfloor \lambda T \rfloor \) periods

\[
f = [\lambda T]^{-1} \sum_{l=1}^{\lfloor \lambda T \rfloor} x_{T+l}
\]

where \( \lambda = 0.5 \), say.

- Aim: Construct interval from data \( \{x_t\}_{t=1}^{T} \) that contains \( f \) with, say, 90% probability in repeated samples
US Postwar GDP Per Capita
Focus and Challenges

• This paper: statistical (rather than "structural") univariate long-term forecasting

• Econometric challenges
  – Only limited sample information about long-term behavior
  – Set of plausible models of long-term behavior?
  – How to deal with model and/or parameter uncertainty?
Low-Frequency Transformations

- Intuitively, question concerns low-frequency properties of $x_t$.

- Extract relevant information by computing low-frequency transforms (Müller and Watson, 2008)

$$X_j = T^{-1} \sum_{t=1}^{T} \sqrt{2} \cos(\pi j t / T)x_t, \quad j = 1, \cdots, q$$

where $q$ is a number like $q = 12$, and treat $(X_1, \cdots, X_q)'$ and $\hat{\mu} = T^{-1} \sum_{t=1}^{T} x_t$ as only available data.
$q = 12$ LF Transforms for GDP
GDP LF Projection
Pros and Cons of LF Transforms

• Extract low-frequency information in \( \{x_t\} \)

• Avoids modelling and potential misspecification of higher frequency aspects

• Captures notion that relevant sample information about long-run forecasts limited

• But potential loss of efficiency (see paper)
**Standard I(0) Asymptotics for Time Series**

- Under a range of primitive conditions on the dependent and heterogeneous mean-zero process \( \{u_t\} \), a Central Limit Theorem holds for all fractions of the sample, i.e. for all \( 0 \leq r_1 < r_2 \leq s_1 < s_2 \),

\[
\left( \frac{1}{\sqrt{T}} \sum_{t=[r_1 T]+1}^{[r_2 T]} u_t \right) \Rightarrow \mathcal{N}\left(0, \begin{pmatrix} \sigma^2(r_2 - r_1) & 0 \\ 0 & \sigma^2(s_2 - s_1) \end{pmatrix} \right)
\]

- This (almost) implies the "Functional" Central Limit Theorem for nicely behaved I(0) processes

\[
T^{-1/2} \sum_{t=1}^{[T]} u_t \Rightarrow \sigma W(\cdot)
\]
Implications for Low-Frequency Transformations

- Suppose \( x_t = \mu + u_t \) and \( u_t \) is \( I(0) \) in the sense \( T^{-1/2} \sum_{t=1}^{\lfloor T \rfloor} u_t \Rightarrow \sigma W(\cdot) \).

- Cosine weights are orthogonal to constant:
  \[
  X_j = \frac{\sqrt{2}}{\sqrt{T}} \sum_{t=1}^{T} \cos(\pi j t / T) x_t = \frac{\sqrt{2}}{\sqrt{T}} \sum_{t=1}^{T} \cos(\pi j t / T) u_t
  \]

- With \( \hat{\mu} = T^{-1} \sum_{t=1}^{T} x_t \), \( f = [\lambda T]^{-1} \sum_{l=1}^{\lfloor \lambda T \rfloor} x_{T+l} \), \( X_0 = \sqrt{T}(\hat{\mu} - \mu) \), and \( Y = \sqrt{T}(f - \mu) \), we obtain
  \[
  (X_0, \cdots, X_q, Y)' \Rightarrow \mathcal{N}(0, \sigma^2 \Sigma)
  \]
  since weighted averages of Gaussian processes are multivariate Gaussian.

- If \( \mu \) and \( \sigma^2 \Sigma \) were known, then could simply report 90% set of the (suitably scaled and centered) conditionally normal distribution \( Y | \{X_j\}_{j=0}^{q} \).
Invariance

- Impose scale and translation invariance:
  \[ \{x_t\}_{t=1}^{T} \mapsto \{m + cx_t\}_{t=1}^{T} \] for any \( m \) and \( c \neq 0 \)
  must lead to corresponding transformation of predictive set

- Can show: Under invariance, asymptotic problem becomes construction of prediction set of
  
  \[ Y^s = \frac{Y - X_0}{s_X} \text{ given } X^s = \left( \frac{X_1}{s_X}, \ldots, \frac{X_q}{s_X} \right)' \]
  where \( s_X^2 = q^{-1} \sum_{j=1}^{q} X_j^2 \)
  \[ \Rightarrow \text{Invariance takes care of lack of knowledge of } \mu \text{ and } \sigma \text{ (but still need to know } \Sigma \text{ to compute the conditional distribution)} \]
Low-Frequency Forecasts—I(0) Model

- In I(0) model, it turns out that

\[ Y^s = \frac{Y - X_0}{s_X} \]

is scaled Student-t, scaled by \( \sqrt{1 + \lambda^{-1}} \)

- \( X^s \) is independent of \( Y^s \)

\[ \Rightarrow \] intervals for \( f \) are of the form \( \hat{\mu} \pm \) student-t quantiles multiplied by

\[ \frac{(1 + \lambda^{-1})^{1/2} s_X}{\sqrt{T}} \]
GDP 50% and 90% Intervals in I(0) Model
Beyond the I(0) Model

• Natural concern that I(0) model is “too stationary”

• Assume local-level model

\[ x_t = \mu + \frac{g}{T} \sum_{s=1}^{t} \eta_s + \varepsilon_t = \mu + u_t \]

where \( \{\varepsilon_t\} \) and \( \{\eta_t\} \) are I(0) with identical long-run variance \( \sigma^2 \), so that \( g \geq 0 \) measures extent of local mean variability

• Still implies

\[ T^{-1/2} \sum_{t=1}^{[\cdot T]} u_t \Rightarrow \sigma G(\cdot) \]  

(1)

for Gaussian process \( G \), so that \( (X_0, \cdots, X_q, Y)' \Rightarrow \mathcal{N}(0, \sigma^2 \Sigma) \), where now \( \Sigma = \Sigma(g) \)
GDP Predictive Densities, LLM, $\lambda = 0.2$
GDP LF Likelihood in LLM
GDP Bayes Predictive Densities

- Flat Prior
- Downward Sloping
- Upward Sloping
Beyond the Local-Level Model

- Approach generalizes to any model $x_t = \mu + u_t$ that satisfies

$$T^{-\alpha} \sum_{t=1}^{\lfloor T \rfloor} u_t \Rightarrow \sigma G(\cdot)$$

for some Gaussian process $G$ and $\alpha$ (for example: fractional model).

- Possible to derive predictive set that remains valid for arbitrary $G$? No, since $\Sigma$ then entirely unconstrained.

- Need some regularity of $x_t$ to be able to forecast.

- Consider covariance stationarity of $\Delta x_t$ (allowing mean growth rate to vary stochastically).
Local-To-Zero Spectrum

- Let \( s_T : [-\pi, \pi] \mapsto \mathbb{R}_+ \) be a sequence of (pseudo) spectral densities of \( \{x_t\} \), and define the local-to-zero spectrum \( S : \mathbb{R} \mapsto \mathbb{R} \) via

\[
S(\omega) = T^{-2\alpha+1} \lim_{T \to \infty} s_T(\omega/T),
\]

for suitable \( \alpha \).

- Under some linear process conditions on \( \Delta x_t \) and (1), we show

\[
(X_0, \ldots, X_q, Y)' \Rightarrow \mathcal{N}(0, \sigma^2 \Sigma)
\]

where \( \Sigma \) is a function of \( S \).

\( \Rightarrow \) Long-run forecasting uncertainty is fully determined by (pseudo) spectral shape close to origin.
Local Spectra

- Fractional model:

\[ S_d(\omega) \propto |\omega|^{-2d}, \quad d \in (-0.5, 1.5) \]

- Local-to-Unity model (AR(1) with \( \rho_T = 1 - c/T \)):

\[ S_c(\omega) \propto 1/(\omega^2 + c^2) \]

- Local-Level model \( x_t = \mu + b\varepsilon_t + \frac{1}{T} \sum_{s=1}^{t} \eta_s \):

\[ S_b(\omega) \propto 1/\omega^2 + b^2 \]

⇒ All these local spectra are in bcd-family

\[ S_\theta(\omega) \propto \left( \frac{1}{\omega^2 + c^2} \right)^{2d} + b^2, \quad \theta = (b, c, d) \]
Local Log-Spectra

Fractional Model

Local-Level Model

Local-To-Unity Model

bcd Model
Parameter Uncertainty

• Local spectrum depends on $\theta = (b, c, d)$, which cannot be estimated consistently by fixed number $q$ of cosine transforms.

• Recall that via invariance, (asymptotic) problem is to forecast $Y^s = \frac{Y - X_0}{s_X}$ by $X^s = \left(\frac{X_1}{s_X}, \ldots, \frac{X_q}{s_X}\right)'$, where $s_X^2 = q^{-1} \sum_{j=1}^{q} X_j^2$.

• Let $\Psi(X^s)$ be a predictive interval of level $1 - \alpha$. Determine $\Psi^*$ that minimizes weighted average expected length over $\theta$, subject to coverage constraint for all values of $\theta$:

$$\min_{\Psi} \int w(\theta) E_{\theta}[\text{length}(\Psi(X^s))] d\theta \quad \text{s.t.} \quad P_{\theta}(Y^s \in \Psi(X^s)) \geq 1 - \alpha \quad \forall \theta \in \Theta$$

$\Rightarrow$ Almost same problem as in Müller, Elliott and Watson (2013)
Parameter Uncertainty: Conditional Properties

- Potential problem: $\Psi^*(X^s)$ could be empty for some $X^s$, and have otherwise unreasonable conditional properties
  
  $\Rightarrow$ generic potential problem of descriptions of uncertainty in nonstandard problems with sets that (only) satisfy confidence type property

  $\Rightarrow$ see Müller and Norets (2012)

- Solution: Impose that $\Psi^*(X^s)$ contains the $1-a$ credible set relative to the prior $w$. 
Implementation

- Set $q = 12$.

- Choose weighting function $w$ uniformly distributed on $d \in [-0.4, 1.4]$ in fractional model

  $\Rightarrow$ seek to minimizes expected length on average with data drawn from fractional model, subject to including the $1 - \alpha$ credible set with that prior and model

- Impose coverage $P_\theta(Y^s \in \Psi(X^s)) \geq 1 - \alpha$ in larger class with local-to-zero spectrum

  $$S_\theta(\omega) \propto \left(\frac{1}{\omega^2 + c^2}\right)^{2d} + b^2$$

  with $d \in [-0.4, 1.4]$ and $b, c$ arbitrary.

  $\Rightarrow$ Frequentist robustification of Bayes credible set
GDP 90% Intervals
US Postwar PCE Inflation
Inflation 50% Intervals
Labor Productivity 50% Interval
Labor Productivity 90% Interval
Conclusions

• Formalization of uncertainty of statistical long-term predictions
  – Low-frequency transformations to yield robustness.
  – Need regularity. Express regularity via shapes of local-to-zero spectrum.
  – Parameter uncertainty resolved by length minimizing robustification of Bayes credible sets.

• Extension to multivariate problem computationally difficult