Measuring Uncertainty about Long-Run Predictions

Ulrich K. Müller and Mark W. Watson
Princeton University

Presentation at CBO

November 4, 2015

Set-up

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

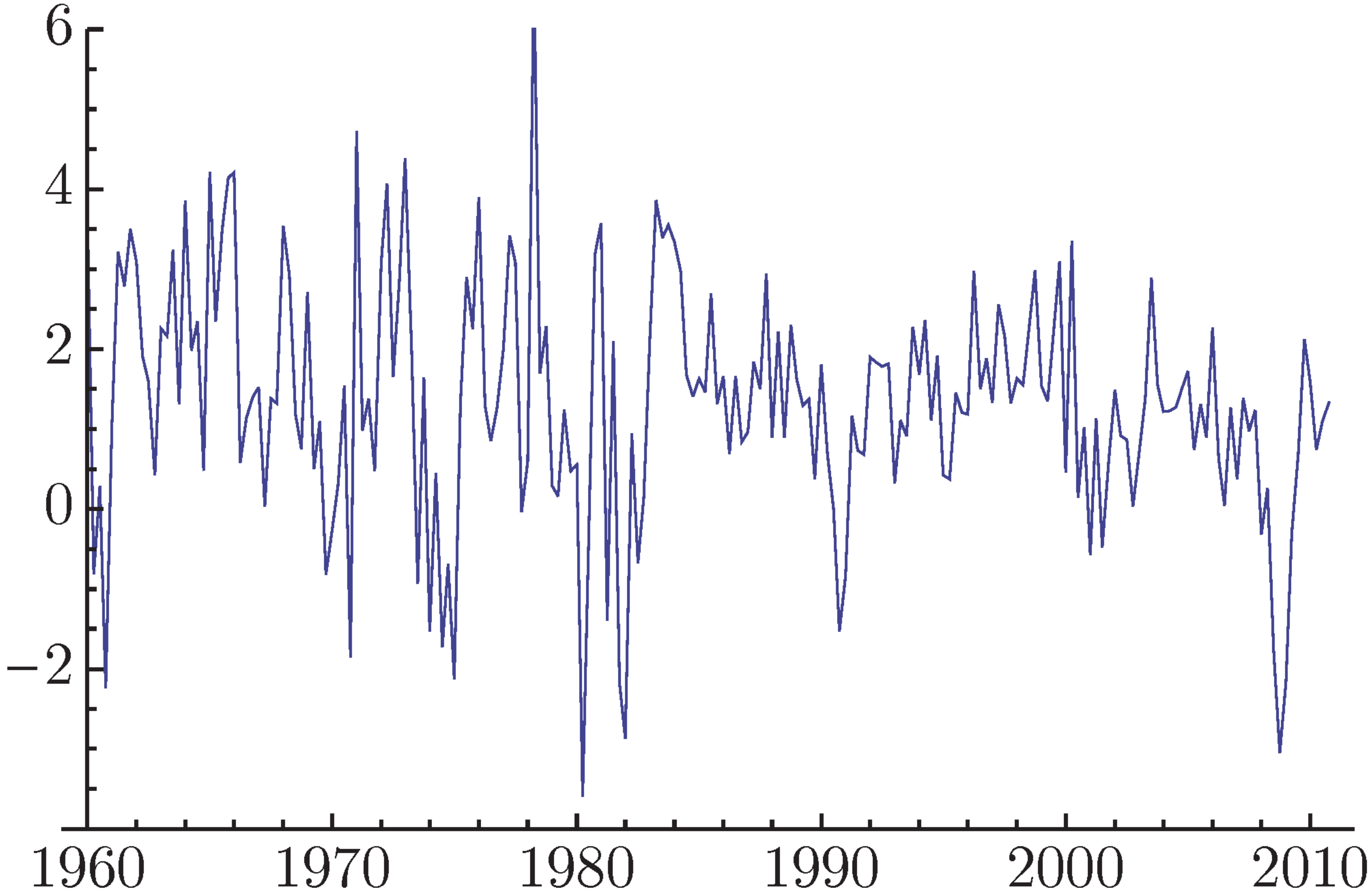
- Want to forecast *average* over the next $h = \lfloor rT \rfloor$ periods

$$\bar{x}_{T+1:hT} = h^{-1} \sum_{l=1}^{h} x_{T+l}$$

  where $r = 0.5$, say

- Aim: Construct interval from data $\{x_t\}_{t=1}^{T}$ that contains $\bar{x}_{T+1:hT}$ with 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

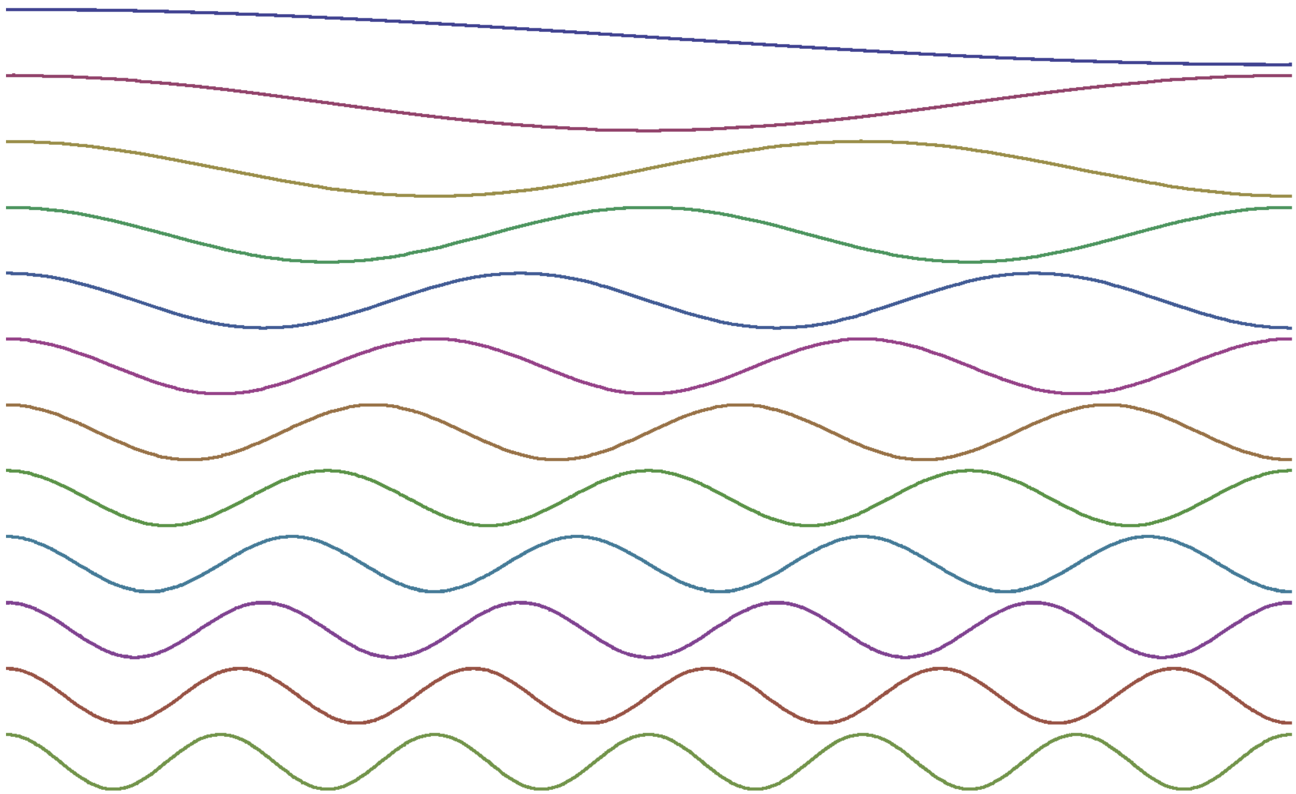
- Intuitively, question concerns low-frequency properties of x_t

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

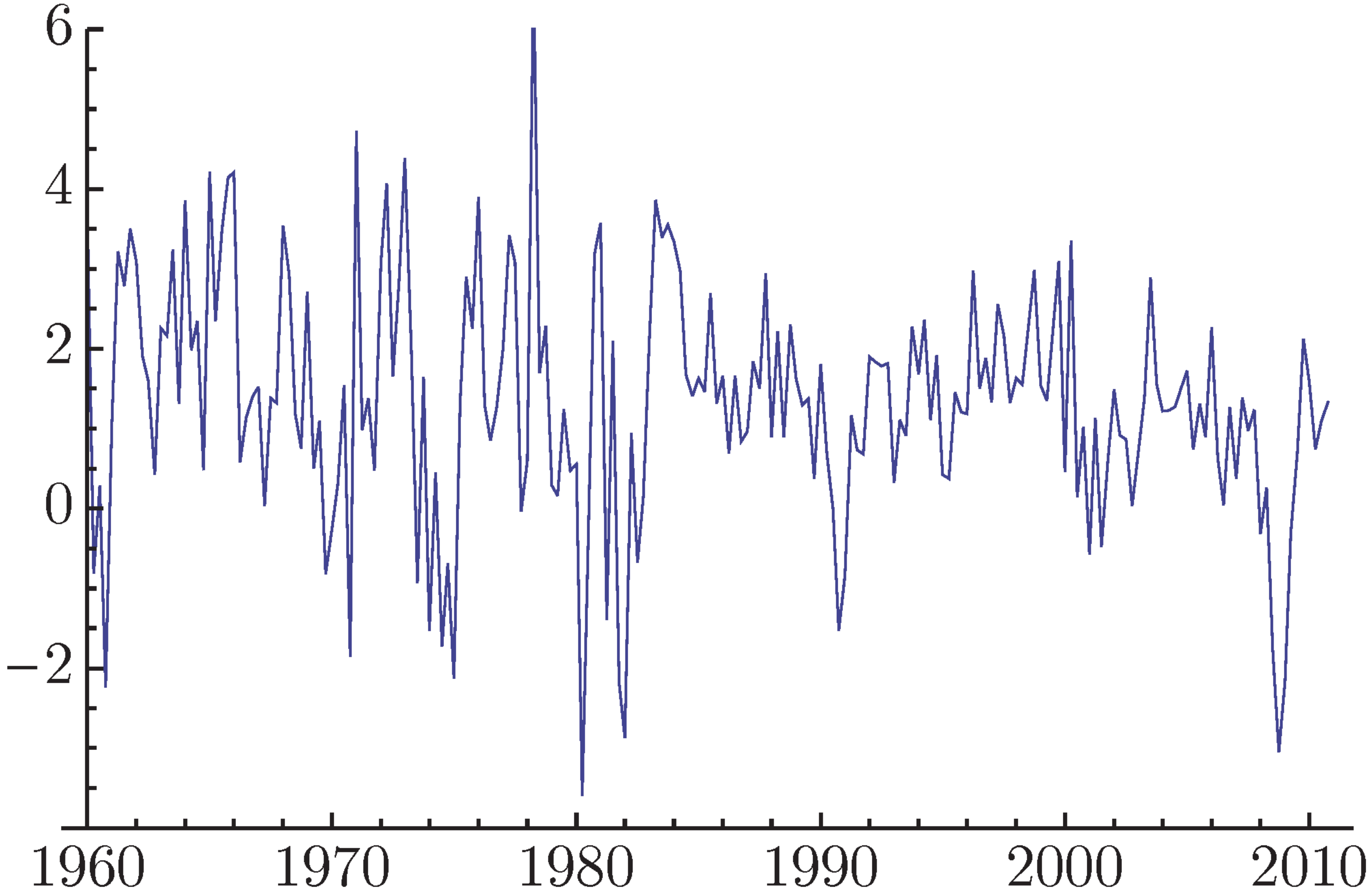
$$X_j = T^{-1} \sum_{t=1}^{T} \sqrt{2} \cos(\pi j t / T) x_t, \quad j = 1, \cdots, q$$

where q is a number like $q = 12$, and treat $(X_1, \cdots, X_q)'$ and $\bar{x}_{1:T} = T^{-1} \sum_{t=1}^{T} x_t$ as only available data

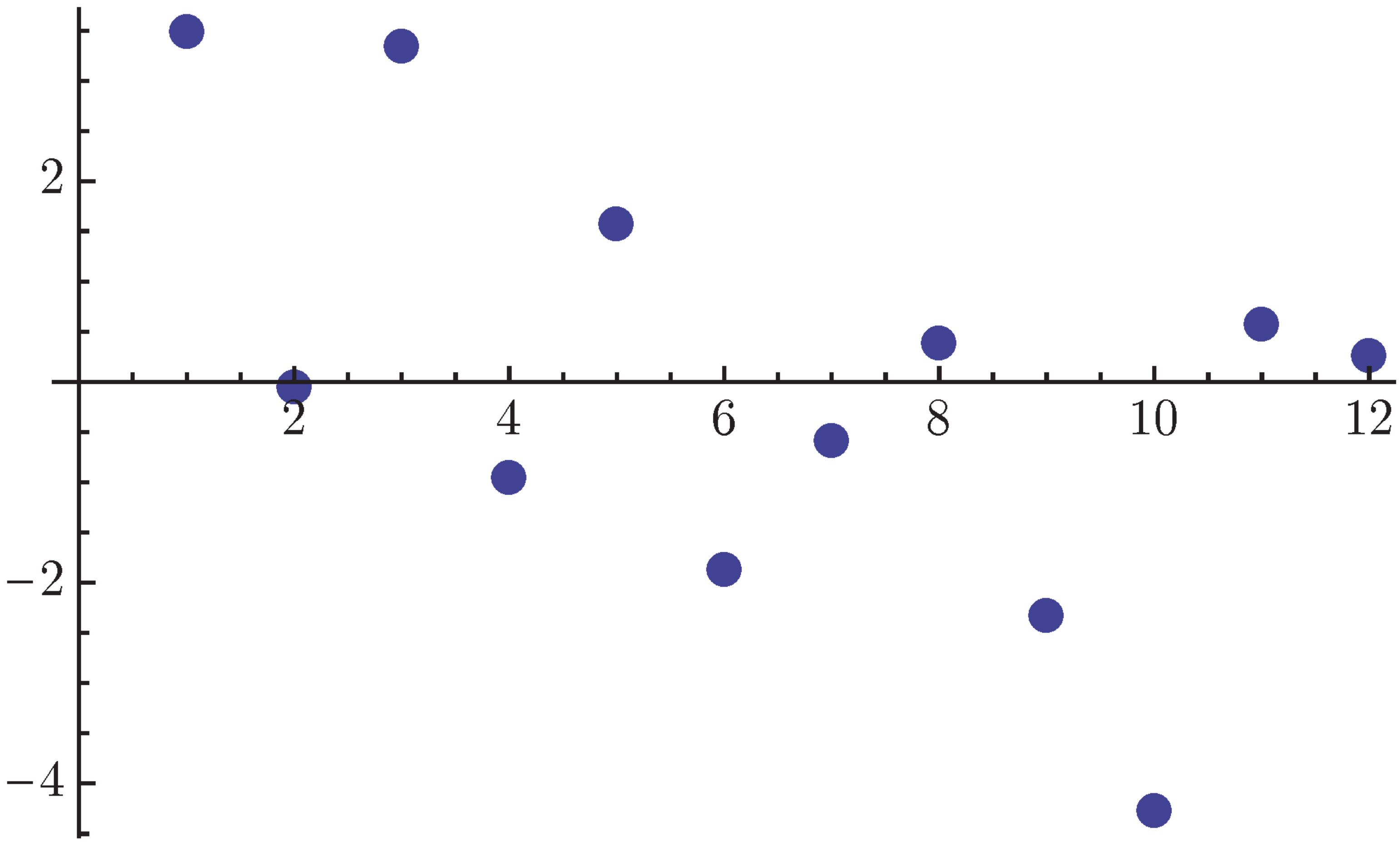
Cosine Weights



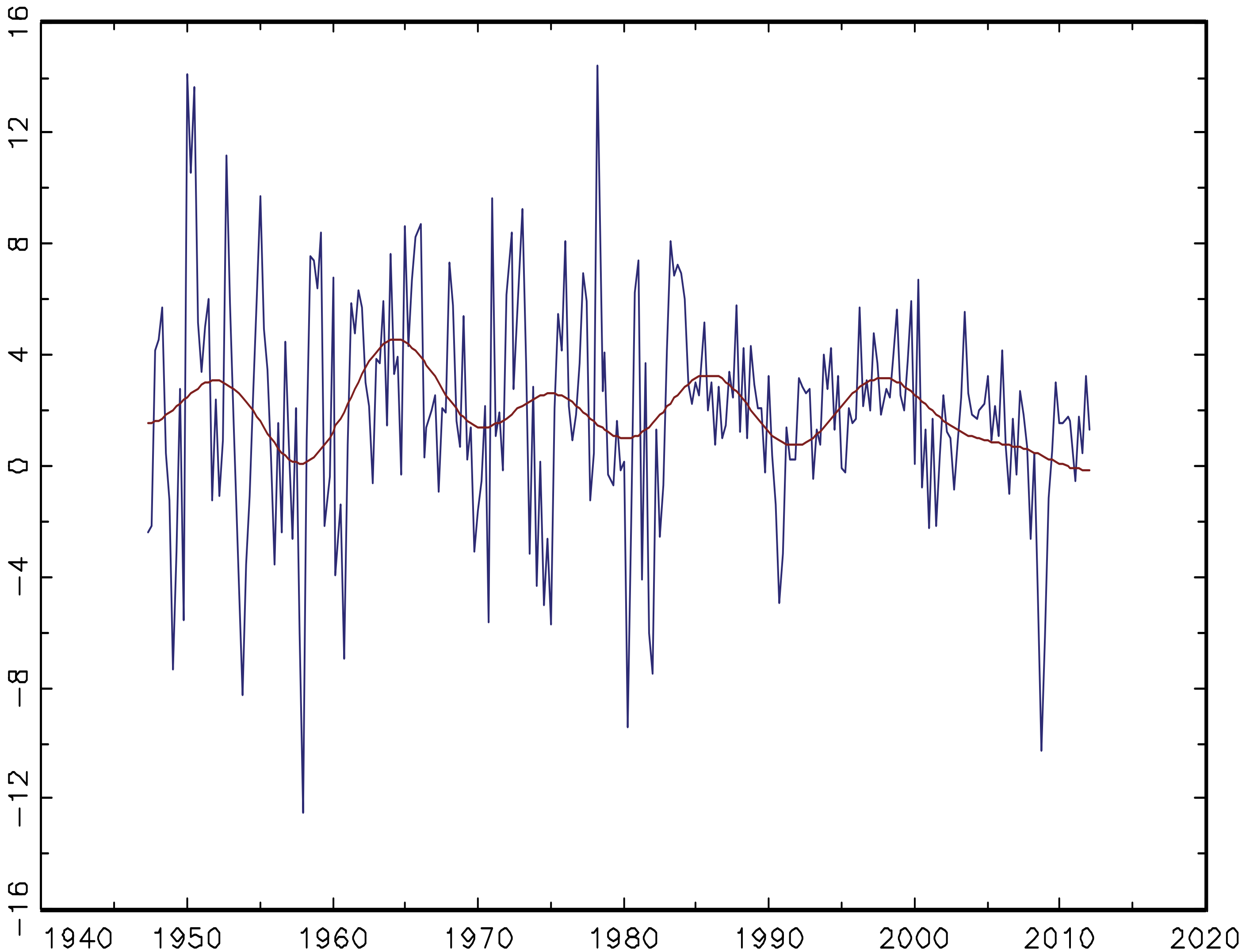
GDP



$q = 12$ 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

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 $0 \leq r_1 < r_2 \leq s_1 < s_2$,

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

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

$$T^{-1/2} \sum_{t=1}^{\lfloor \cdot T \rfloor} u_t \Rightarrow \sigma W(\cdot)$$

Implications for Low-Frequency Transformations

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

- 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 $F = \sqrt{T}(\bar{x}_{T+1:hT} - \mu)$, we obtain

$$(X_1, \cdots, X_q, F)' \Rightarrow \mathcal{N}(0, \sigma^2 \mathbf{\Sigma})$$

  since weighted averages of Gaussian processes are multivariate Gaussian

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

Invariance

- Impose scale and translation invariance:

$$\{x_t\}_{t=1}^T \mapsto \{m + cx_t\}_{t=1}^T \quad \text{for any } m \text{ and } c \neq 0$$

must lead to corresponding transformation of predictive set

- Can show: Under invariance, problem becomes construction of prediction set of

$$Y^s = \frac{Y}{s_X} \text{ given } X^s = \left(\frac{X_1}{s_X}, \cdots, \frac{X_q}{s_X}\right)'$$

where $Y = \sqrt{T}(\bar{x}_{T+1:hT} - \bar{x}_{1:T})$ and $s_X^2 = q^{-1} \sum_{j=1}^q X_j^2$

$\Rightarrow$ Prediction set for $\bar{x}_{T+1:hT}$ is this prediction set for Y^s scaled by $s_X/\sqrt{T}$ and shifted by $\bar{x}_{1:T}$

$\Rightarrow$ Invariance takes care of lack of knowledge of μ and σ

Constructing Invariant Prediction Sets

- With $X = (X_1, \ldots, X_q)'$, under

$$\begin{pmatrix} X \\ Y \end{pmatrix} \Rightarrow \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \sigma^2 \begin{pmatrix} \Sigma_{XX} & \Sigma_{Xy} \\ \Sigma_{XY} & \Sigma_{YY} \end{pmatrix}\right)$$

a computation shows that conditional on $X^s = x^s$,

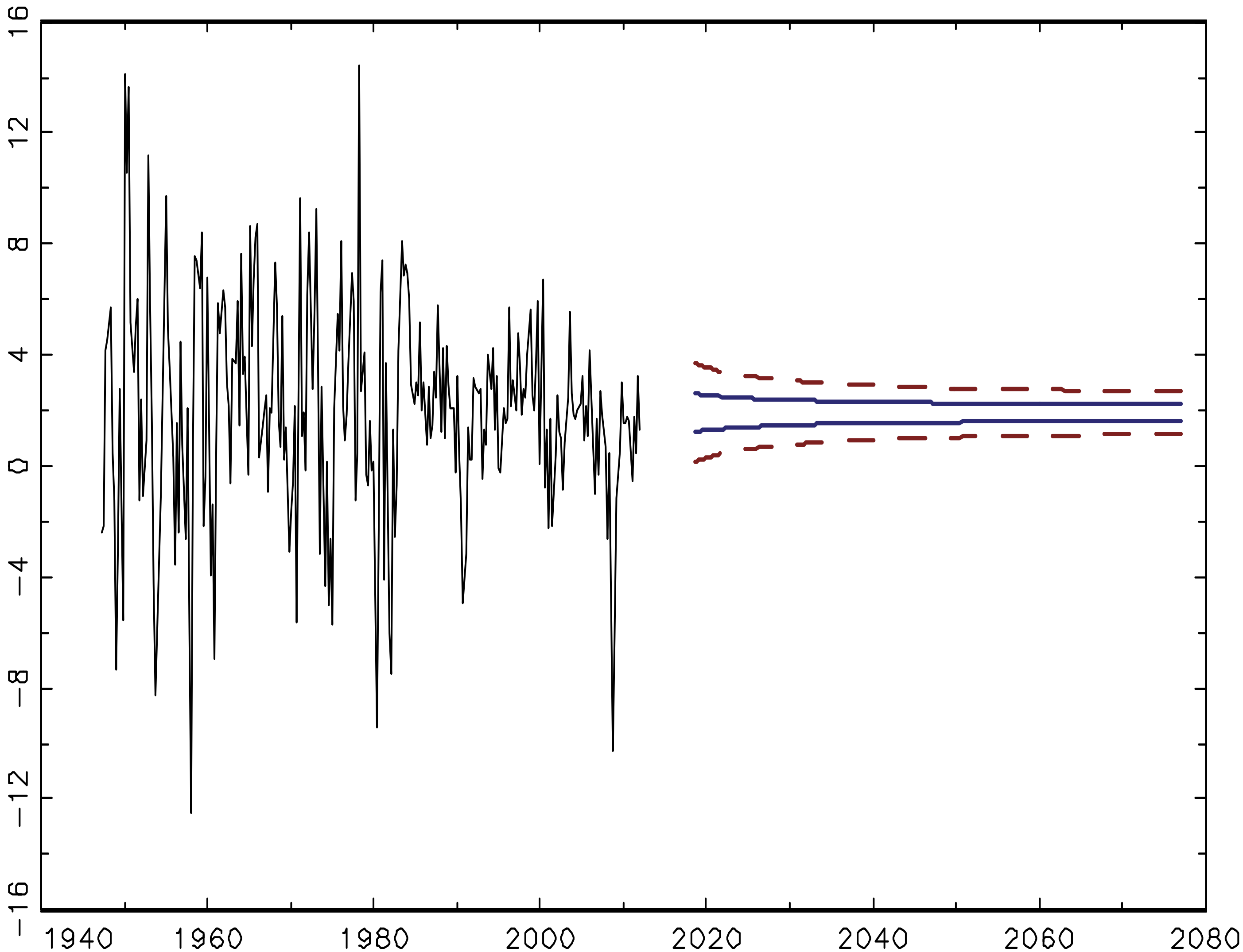
$$\frac{Y^s - \Sigma_{YX}\Sigma_{XX}^{-1}x^s}{\sqrt{\Sigma_{YY} - \Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}}\sqrt{x^{s\prime}\Sigma_{XX}^{-1}x^s/q}} \sim \text{Student-}t^q$$

$\Rightarrow$ Predictive densities are rescaled student-t densities

- I(0) model: $\Sigma_{XY} = 0$, $\Sigma_{YY} = 1 + r^{-1}$, so intervals for $\bar{x}_{T+1:hT}$ are of the form

$$\bar{x}_{1:T} \pm t_{0.95}^q \times (1 + r^{-1})^{1/2} s_X T^{-1/2}$$

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 σ^2, so that $g \geq 0$ measures extent of local mean variability

- Still implies

$$T^{-1/2} \sum_{t=1}^{\lfloor \cdot T \rfloor} u_t \Rightarrow \sigma G(\cdot)$$

for Gaussian process G, so that $(X', Y)' \Rightarrow \mathcal{N}(0, \sigma^2 \mathbf{\Sigma})$, where now $\mathbf{\Sigma} = \mathbf{\Sigma}(g)$

Predictive Density in LLM

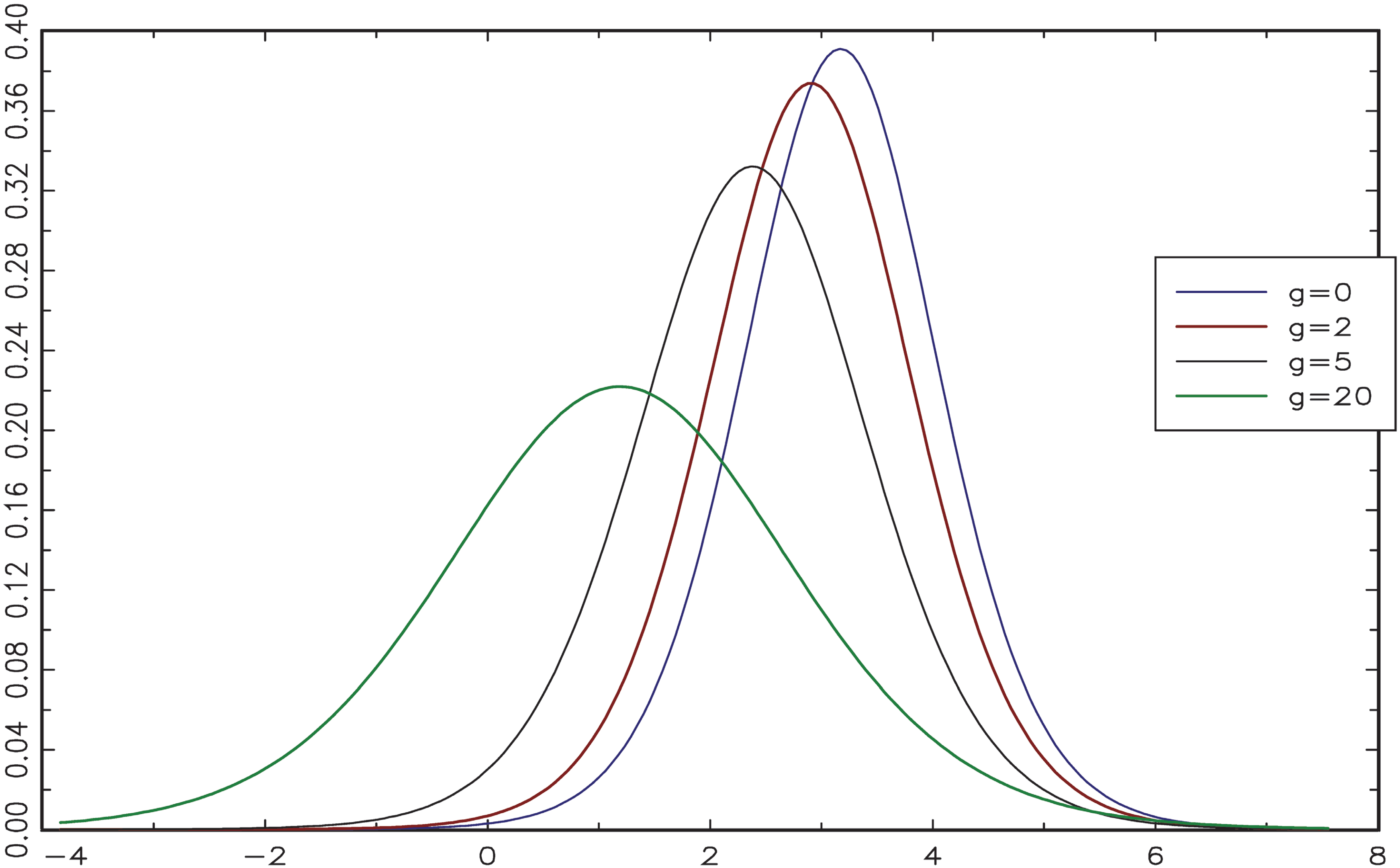
- As before, now using $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(g)$,

$$\frac{Y^s - \boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}x^s}{\sqrt{\boldsymbol{\Sigma}_{YY} - \boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{\Sigma}_{XY}}\sqrt{x^{s\prime}\boldsymbol{\Sigma}_{XX}^{-1}x^s/q}} \sim \text{Student-}t^q$$

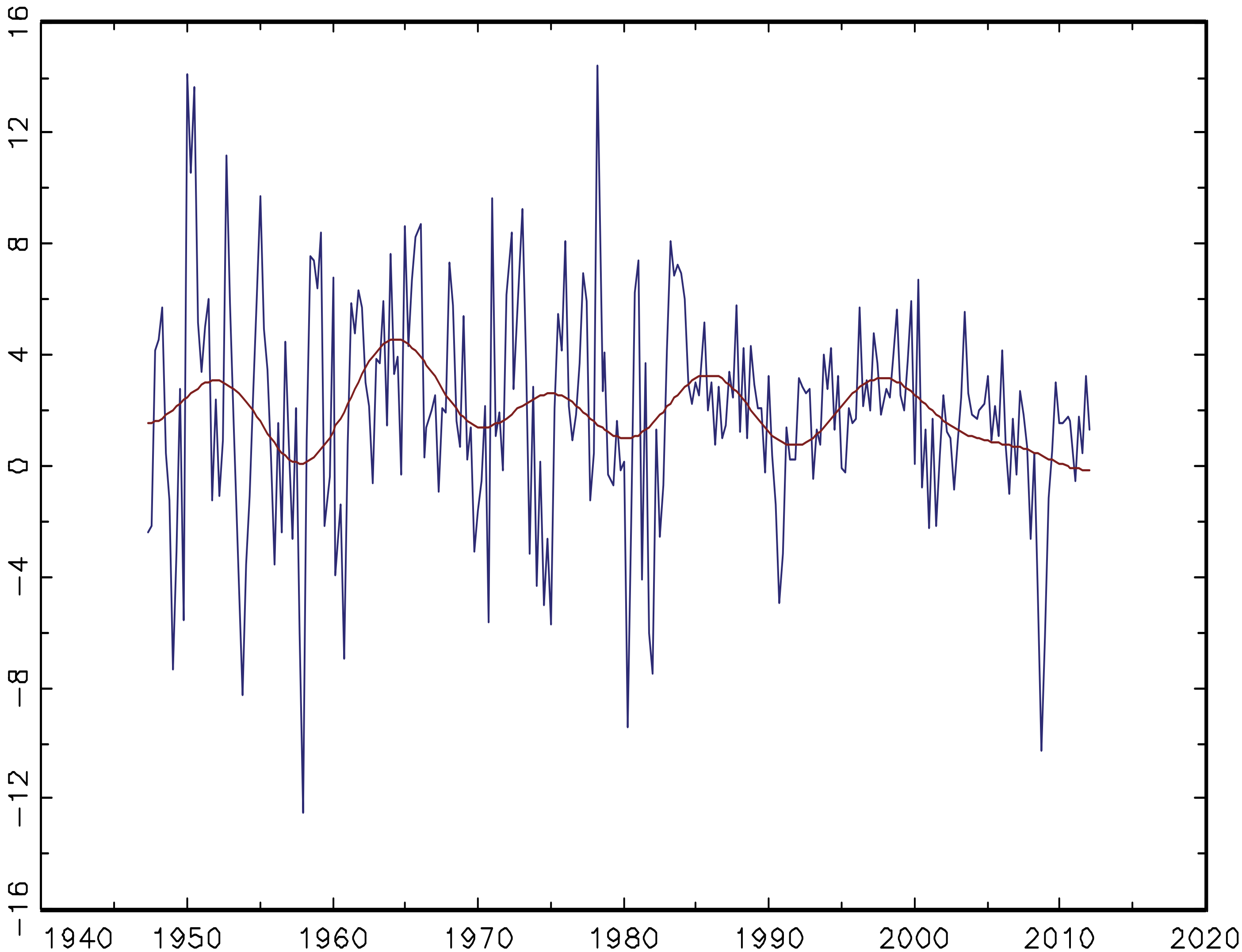
  so predictive density for given g is again rescaled student-t density

- Now $\boldsymbol{\Sigma}_{YX} \neq 0$ in general, so realization of X^s shifts mean forecast

GDP Predictive Densities, LLM, $r = 0.2$



GDP LF Projection



GDP Bayes Predictive Densities

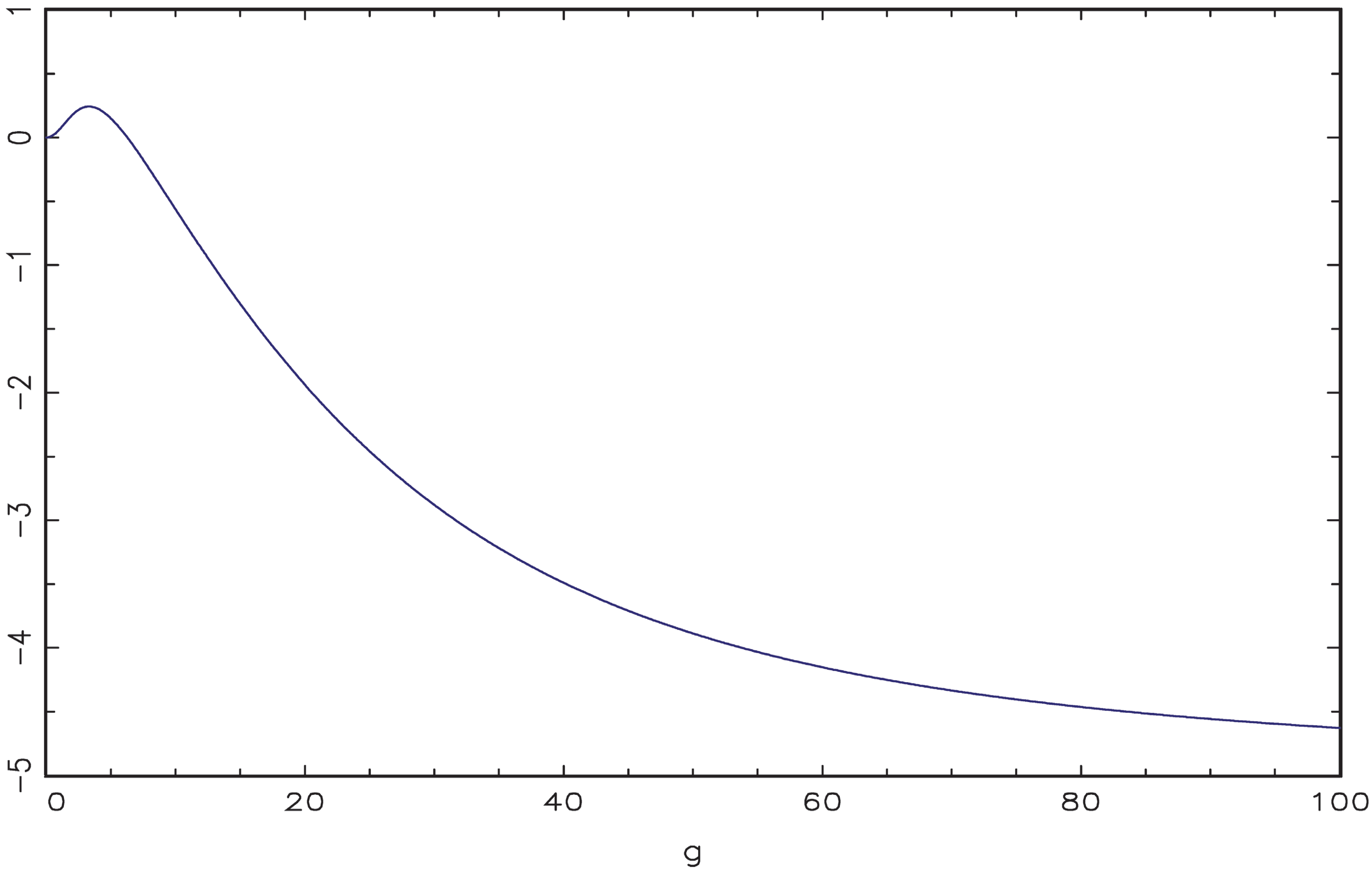
- Posit prior with density $\pi(g)$

- Form posterior about g given observation $X^s = X/\sqrt{X'X}$

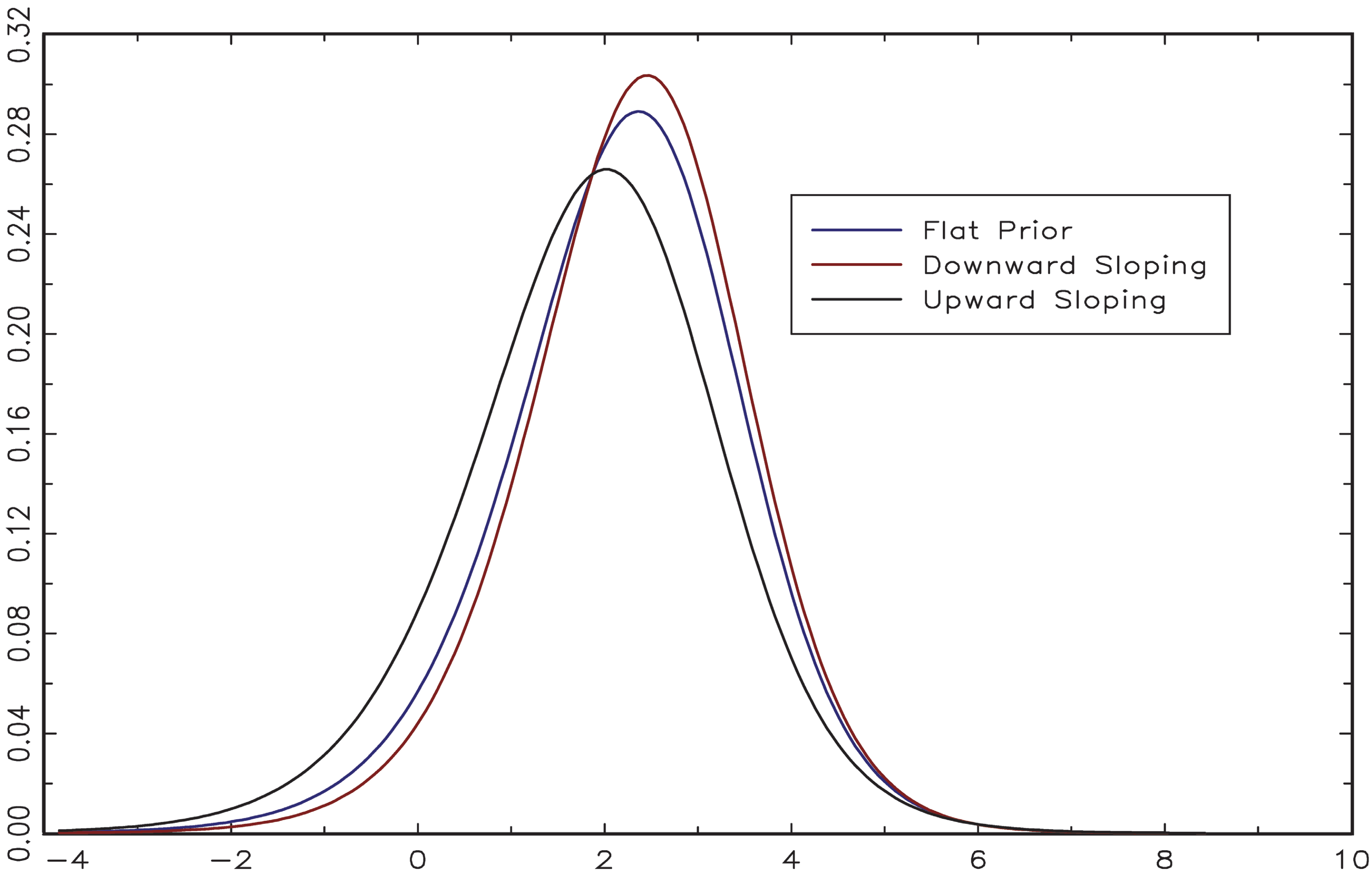
  $\Rightarrow$ Posterior $p(g|x^s)$ proportional to $\pi(g) f_{X^s|g}(x^s)$

- Bayes predictive density is mixture of predictive densities given g, weighted by $p(g|x^s)$

GDP Log-Likelihood in LLM



GDP Bayes Predictive Densities



Beyond the Local-Level Model

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

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

  for some Gaussian process G and α (for example: fractional model)

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

- Need some regularity of x_t to be able to forecast

bcd-Model

- Assume $x_t = \mu + u_t$, where

$$
\begin{aligned}
u_t &= \varepsilon_{1t} + (bT)^{-d}\eta_t \\
(1-\rho T)^d \eta_t &= \varepsilon_{2t}
\end{aligned}
$$

  with $\rho = \rho_T = 1 - c/T$, $d \in [-1/2, 3/2]$ and $(\varepsilon_{1T}, \varepsilon_{2T})$ uncorrelated I(0) with long-run variance σ^2

  $\Rightarrow$ "bcd-model": Nests local-level model, fractional model and local-to-unity AR(1) model as special cases

  $\Rightarrow$ With x_t growth rate, allows for stochastically trending mean growth, slow mean reversions, anti-persistence ($d < 0$)

- Under some regularity conditions, we show that in bcd-model

$$
(X', Y)' \Rightarrow \mathcal{N}(0, \sigma^2 \Sigma(\theta))
$$

  with $\theta = (b, c, d)$

Parameter Uncertainty: Bayes Approach

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

- As in LLM, with prior Γ on θ, posterior density of y^s is posterior mixture of Student-t densities, now computed from $(X', Y)' \Rightarrow \mathcal{N}(0, \sigma^2 \Sigma(\theta))$

  $\Rightarrow$ But set $\Psi^\Gamma(x^s)$ that contains $1 - \alpha$ posterior predictive mass for all x^s does not necessarily satisfy $P_\theta(Y^s \in \Psi^\Gamma(X^s)) \geq 1 - \alpha$ all θ

  (by construction, only $\int P_\theta(Y^s \in \Psi^\Gamma(X^s)) d\Gamma(\theta) = 1 - \alpha$)

Parameter Uncertainty: Frequentist Approach

- Frequentist might insist on $P_\theta(Y^s \in \Psi(X^s)) \geq 1 - \alpha$ all θ, and prefer short Ψ

- Consider solution Ψ^* to program

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

$\Rightarrow$ Can be written as Lagrangian, with Lagrange multipliers λ_θ associated with constraints $P_\theta(Y^s \in \Psi(X^s)) \geq 1 - \alpha$

$\Rightarrow$ Given Lagrange multipliers, optimal Ψ^* straightforward

- Apply numerical technique developed in Elliott, Müller and Watson (2015) to approximate λ_θ

Parameter Uncertainty: Conditional Properties

- Potential problem: $\Psi^*(X^s)$ could be empty for some X^s, and have otherwise unreasonable conditional properties

  $\Rightarrow$ Intuitively, it might be optimal to report the empty set for very uninformative draws X^s, as there is a high cost in terms of length to report the full degree of uncertainty

  $\Rightarrow$ see Müller and Norets (2012) for additional examples

- Solution: Impose that $\Psi^*(x^s)$ contains the $1 - a$ credible set relative to some prior Γ for all x^s

  $\Rightarrow$ Minor change in numerical approach to determining Ψ^*

Implementation

- Set $q = 12$

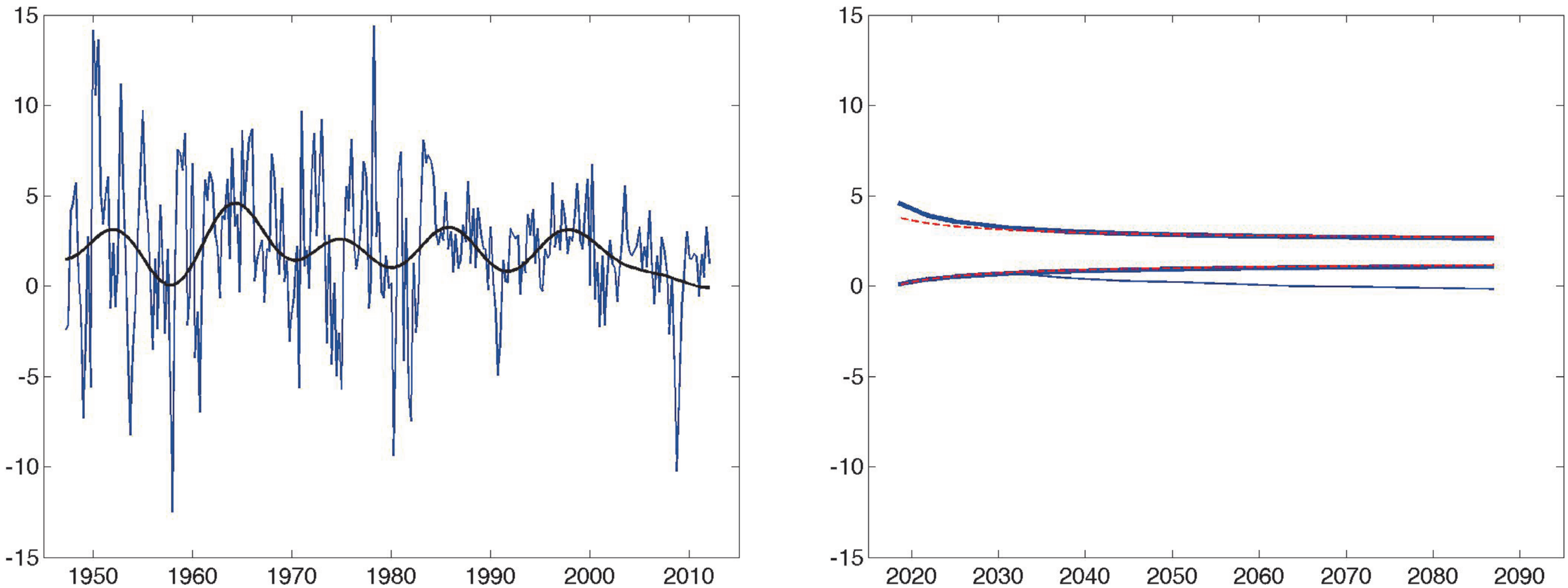
- Choose W and Γ to be uniformly distributed on $d \in [-0.4, 1.0]$ in fractional model (and $b = c = 0$ under W and Γ)

  $\Rightarrow$ 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 bcd-model with $d \in [-0.4, 1.0]$, and (b, c) unconstrained

  $\Rightarrow$ Frequentist robustification of Bayes credible set

Average GDP Growth 90% Forecast Intervals

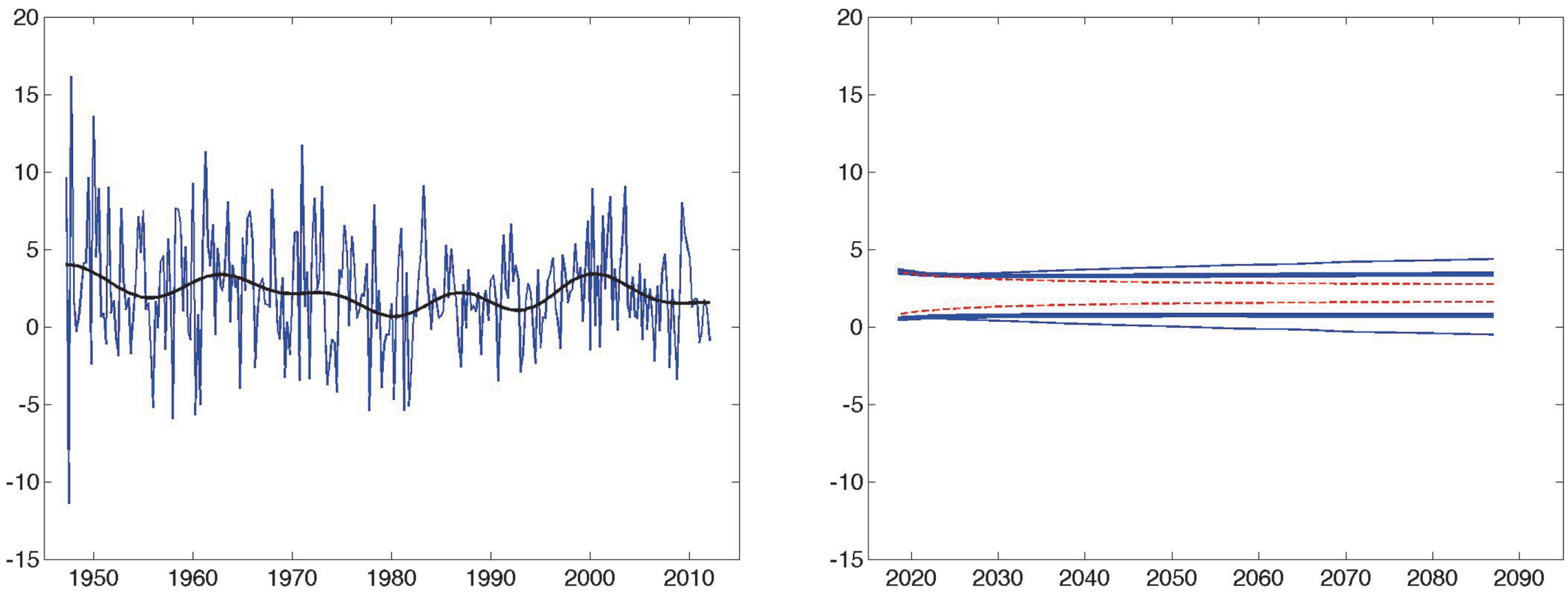


Dashed: I(0)

Thick: Bayes

Thin: 90% Coverage

Average Labor Prod. Growth 90% Forecast Intervals

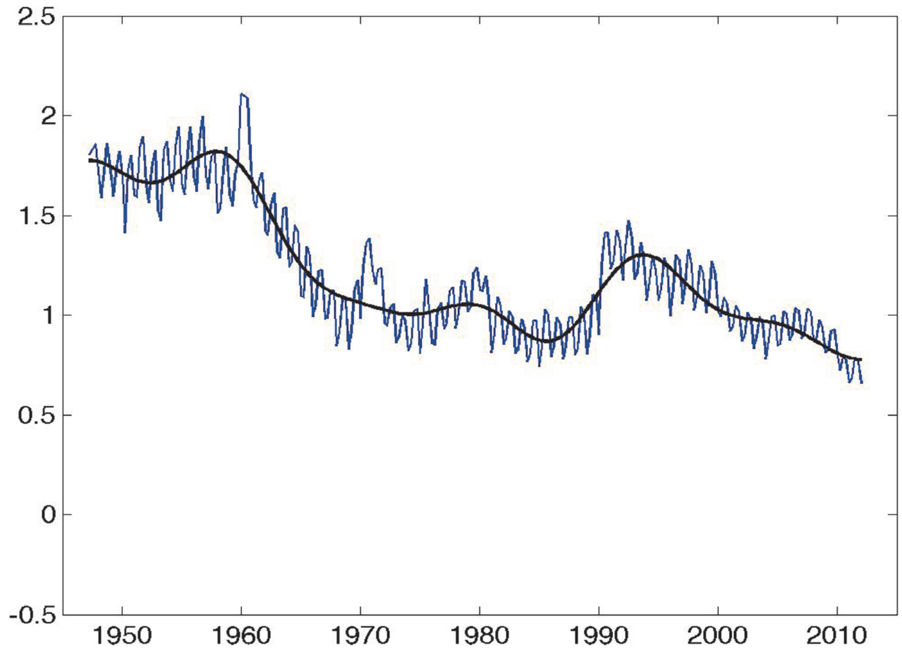


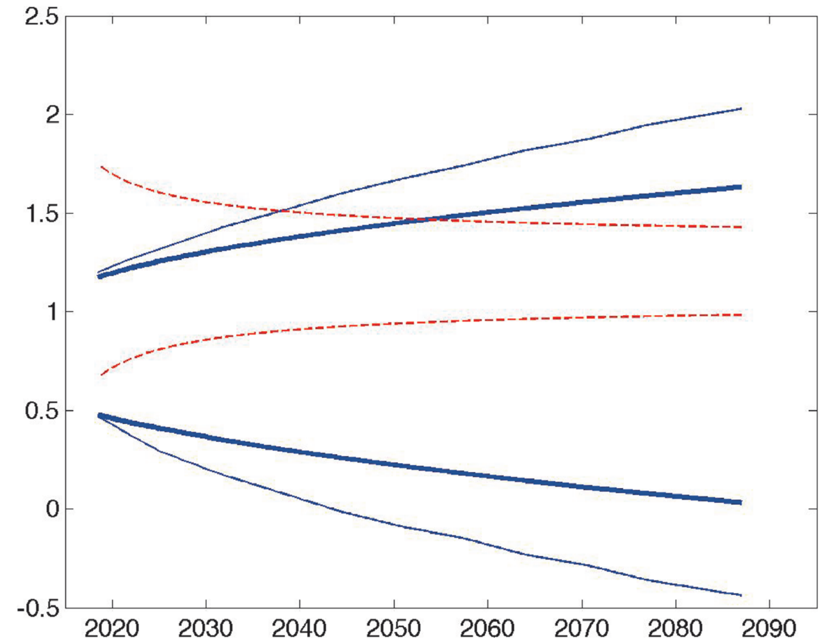
Dashed: I(0)

Thick: Bayes

Thin: 90% Coverage

Average Population Growth 90% Forecast Intervals





Dashed: I(0)

Thick: Bayes

Thin: 90% Coverage

Conclusions

- Formalization of uncertainty of statistical long-term predictions
  - Low-frequency transformations to yield robustness
  - Need regularity: Flexible 3 parameter model of low-frequency variability
  - Parameter uncertainty resolved by length minimizing robustification of Bayes credible sets
- Extension to multivariate problem computationally difficult