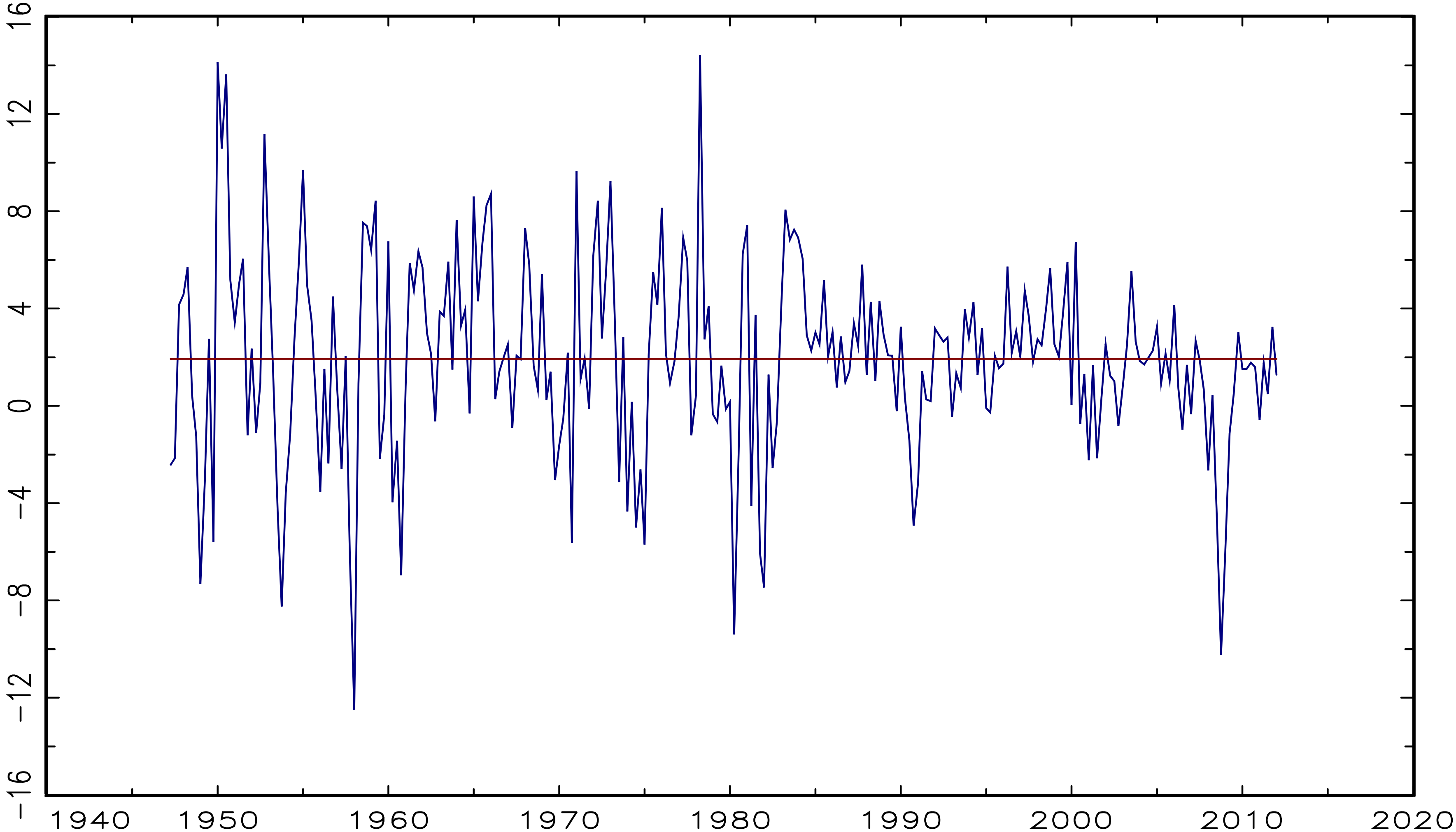
Measuring Uncertainty About Long-Run Forecasts

Ulrich Müller and Mark Watson
Princeton University

Society for Economic Dynamics
Seoul, Korea
June 27-29, 2013

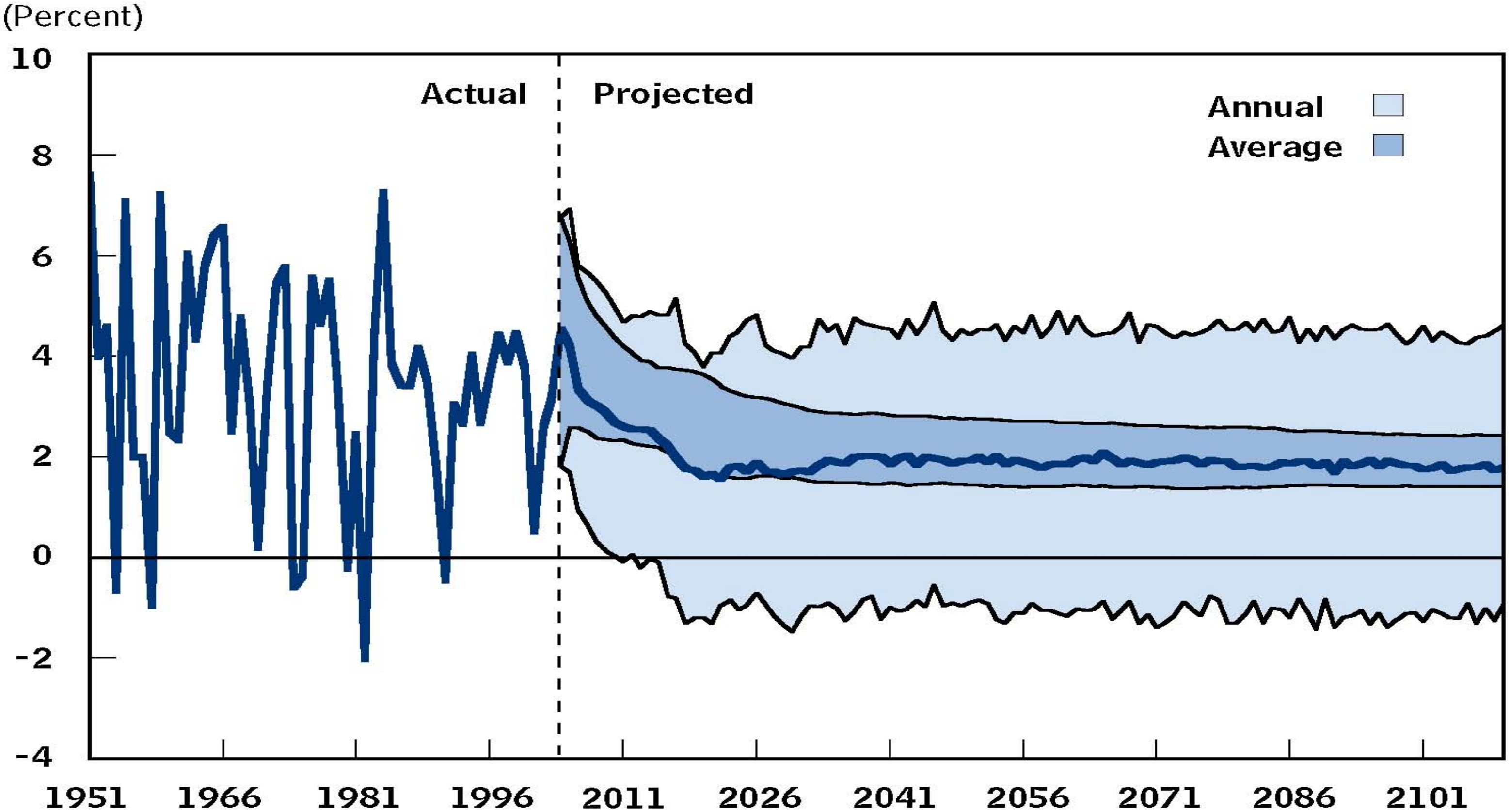
Motivation: Real U.S. Per Capita GDP



CBO 2005

Figure 14.

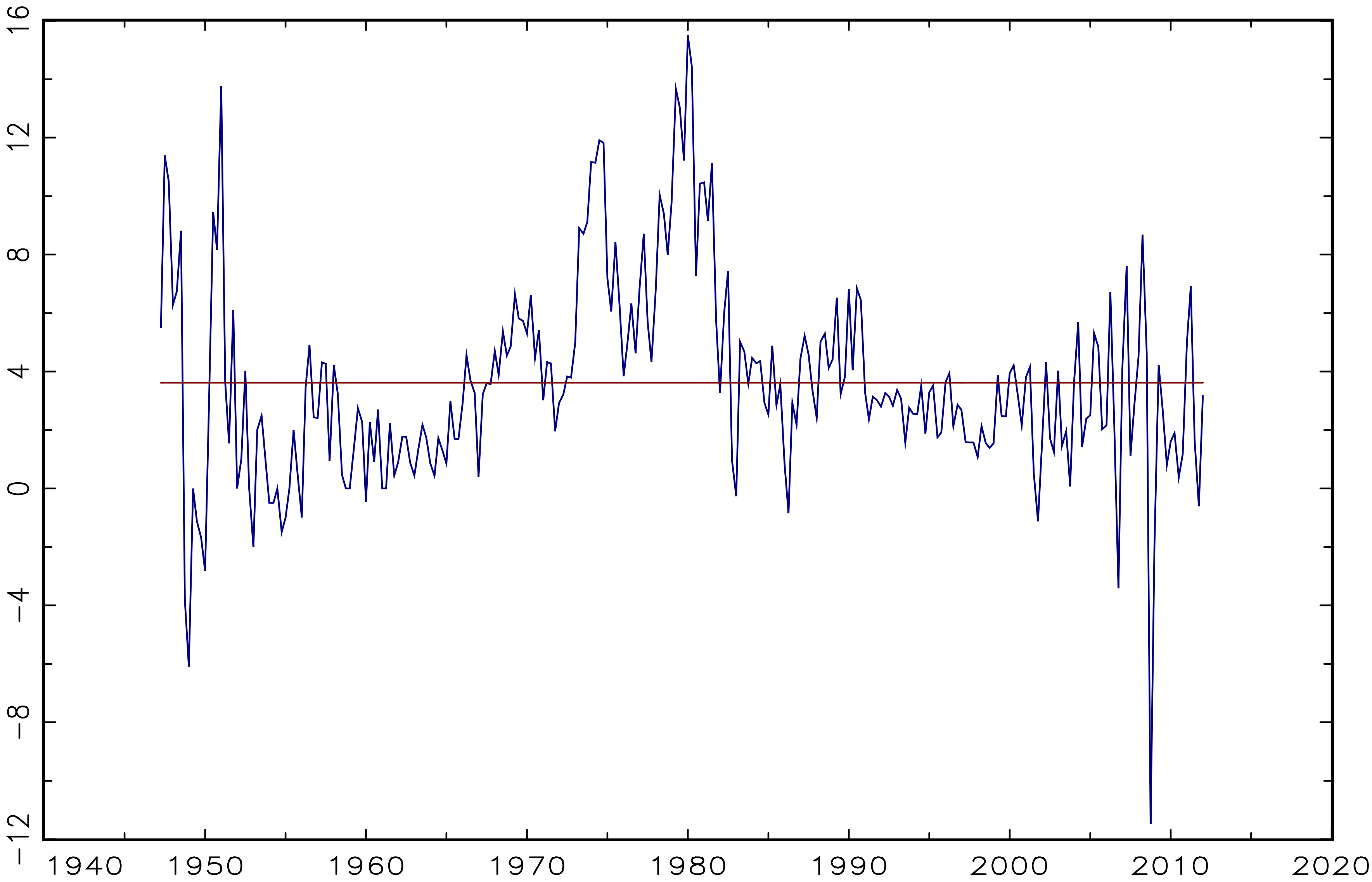
Uncertainty Bands for Real GDP Growth



Source: Congressional Budget Office.

Note: Annual uncertainty bands show the 80 percent confidence range for a given year. Average uncertainty bands show the 80 percent confidence range for the average of 2004 through a given year.

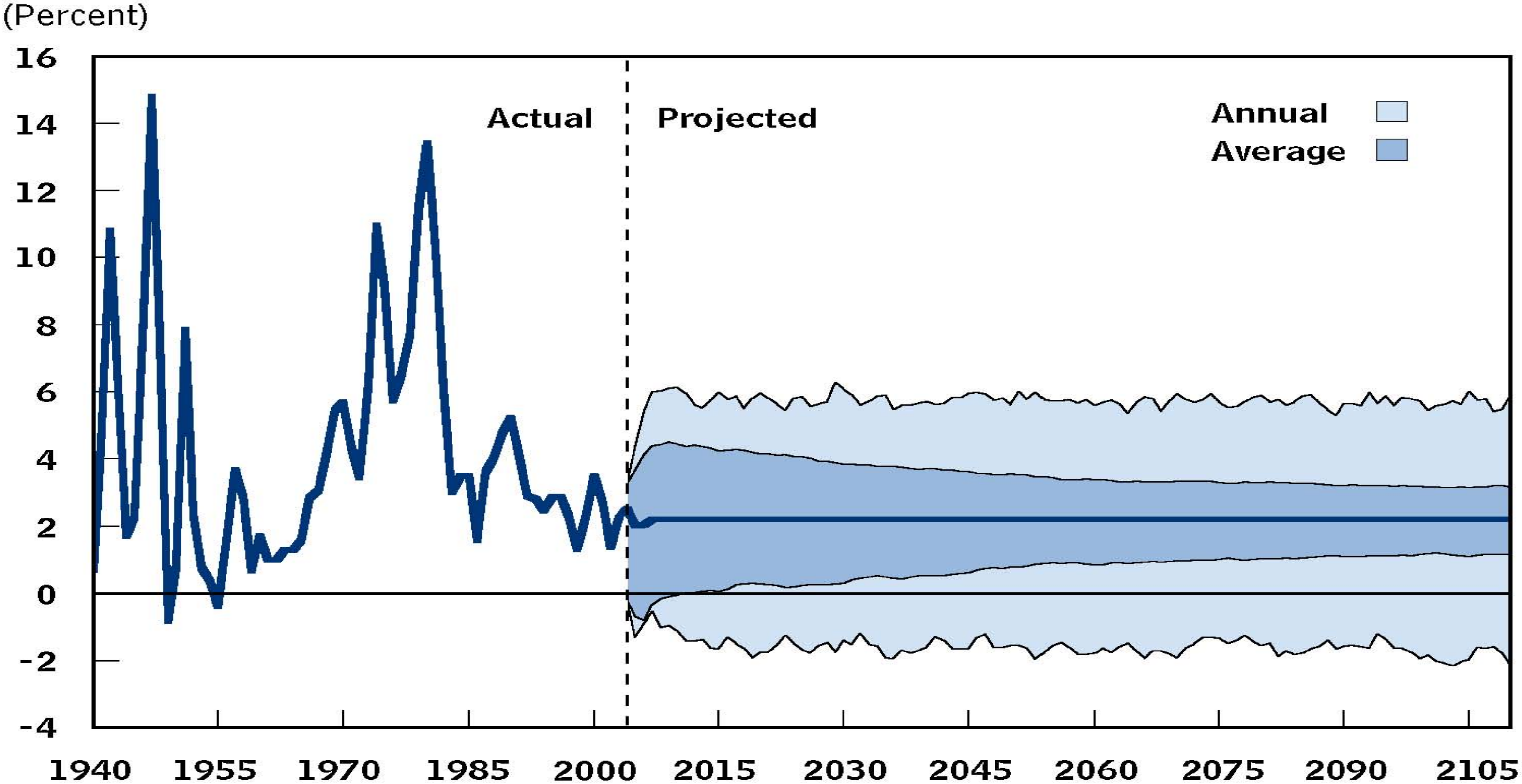
U.S. Inflation (CPI)



CBO 2005:

Figure 11.

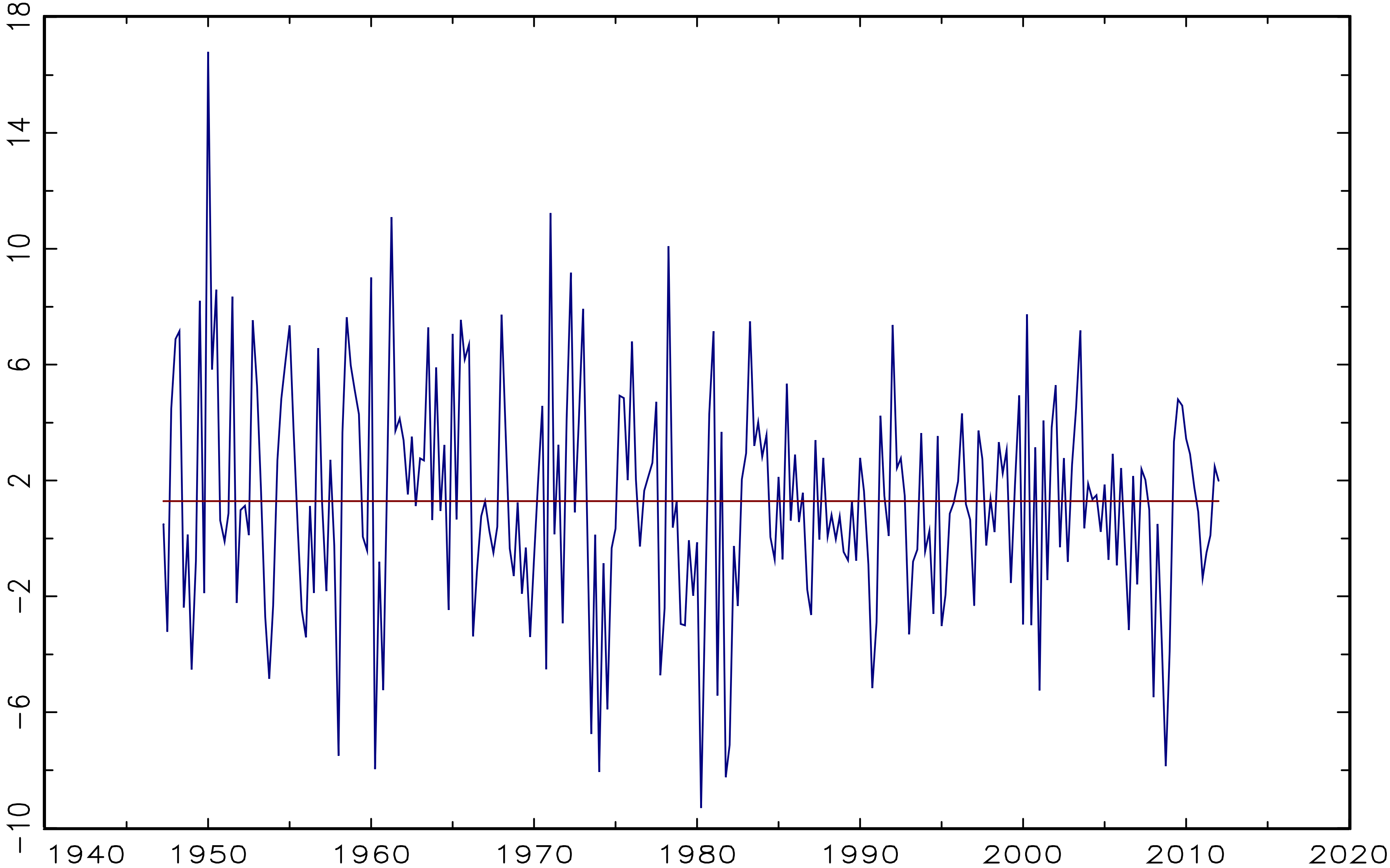
Uncertainty Bands for the Inflation Rate



Source: Congressional Budget Office.

Note: Annual uncertainty bands show the 80 percent confidence range for a given year. Average uncertainty bands show the 80 percent confidence range for the average of 2004 through a given year.

U.S. TFP

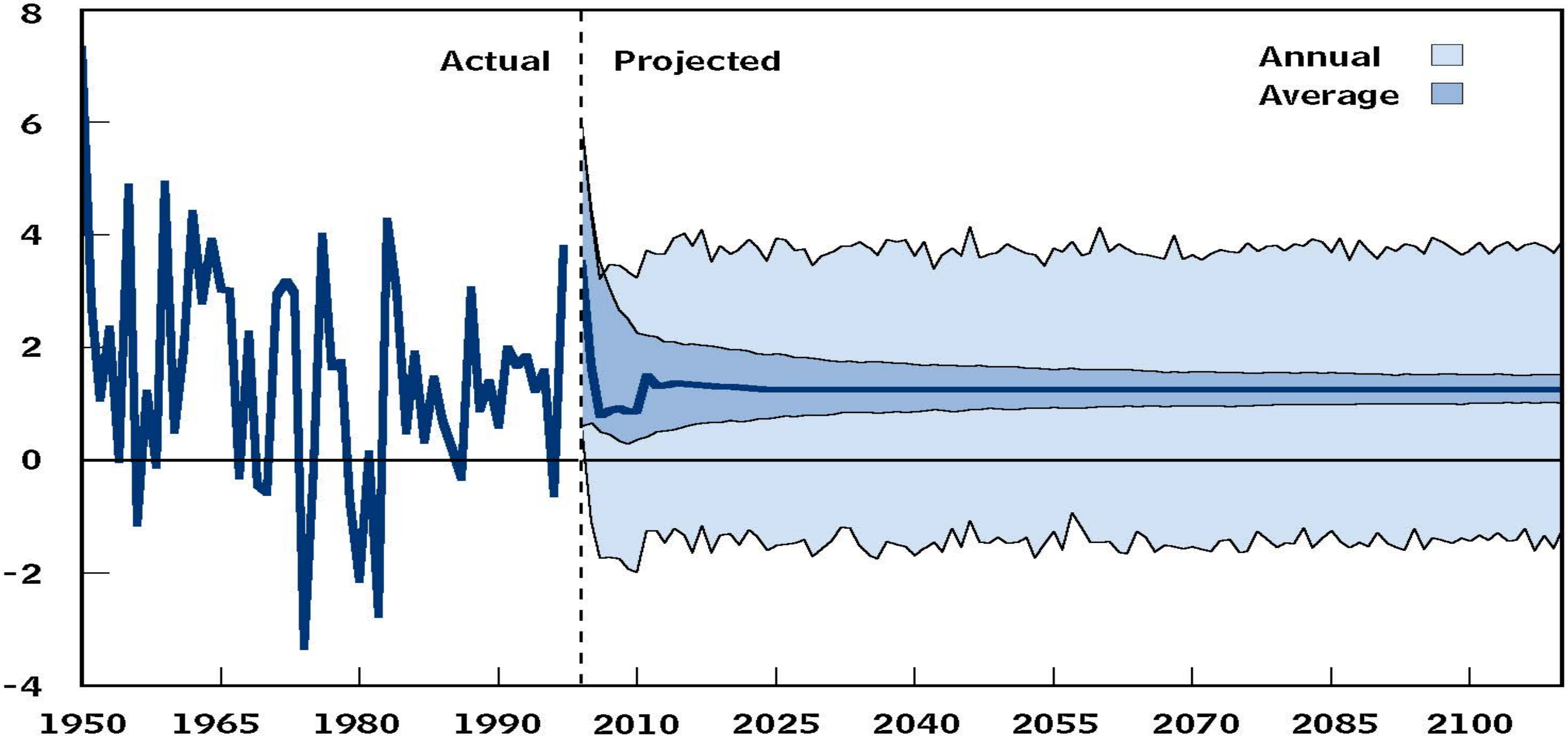


CBO 2005

Figure 8.

Uncertainty Bands for the Rate of Total Factor Productivity Growth

(Percent)



Source: Congressional Budget Office.

Note: Annual uncertainty bands show the 80 percent confidence range for a given year. Average uncertainty bands show the 80 percent confidence range for the average of 2004 through a given year.

Notation, etc.

- Sample Data: $\boldsymbol{x_t}$, $t = 1, \ldots, T$
  - o x denotes the growth rate of GDP, the inflation rate, etc.
  - o T is large, say 60 years of quarterly observations

- Variable of interest is average value of x over h periods in the future:

$$\bar{x}_{T+1:T+h} = \frac{1}{h}\sum_{i=1}^{h} x_{T+i}$$

  - o h is large, say 25 or 50 or 75 years of quarterly observations

- Goal: Construct Prediction Set, A_T with $P\left(\bar{x}_{T+1:T+h} \in A_T\right) = 1-\alpha$

  $A_T = A(\text{sample data})$

Suppose x_t is "I(0)" (or "covariance stationary") . Let μ be the mean of x.

$\overline{x}_{T+1:T+h}$ is the sample average of many future observations (h is large), so:

- $\overline{x}_{T+1:T+h} \sim N\left(\mu, \sigma_{LR}^2 / h\right)$
  - This is an approximation based on Central Limit Theorem
  - σ_{LR}^2 is the "long-run" variance.

- 95% Prediction Interval: $\mu \pm 1.96\sqrt{\sigma_{LR}^2 / h}$

I(0) Continued:

$$\text{95\% Prediction Interval: } \mu \pm 1.96\sqrt{\sigma_{LR}^2 / h}$$

- μ is unknown
  - o use sample mean $\bar{x}_{1:T}$ and incorporate additional source of uncertainty:
    - ▪ 95% interval: $\bar{x}_{1:T} \pm 1.96\sqrt{\sigma_{LR}^2 / h + \sigma_{LR}^2 / T}$

- σ^2 is unknown
  - o Use estimate (Newey-West or Müller)
    - ▪ $\bar{x}_{1:T} \pm t_{df,95\%}\sqrt{\hat{\sigma}_{LR}^2 / h + \hat{\sigma}_{LR}^2 / T}$

Some Numbers: What's wrong with the I(0) model?

I(0) Prediction Set: $\bar{x}_{1:T} \pm t_{df,95\%}\sqrt{\hat{\sigma}^2_{LR}/h + \hat{\sigma}^2_{LR}/T}$

Summary Statistics and I(0) Predictive Sets Selected Post-WWII Series

Series	$\bar{x}_{1:T}$	Predictive Sets: I(0) Model 25-Year Horizon		
		50%	80%	90%
GDP/Pop	1.9	(1.5, 2.3)	(1.1, 2.7)	(0.8, 3.0)
Inflation (CPI)	3.6	(2.8, 4.5)	(2.0, 5.3)	(1.5, 5.8)
Inflation (Japan, CPI)	3.2	(1.9, 4.5)	(0.7, 5.7)	(0.0, 6.5)
TFP	1.3	(1.0, 1.6)	(0.7, 1.9)	(0.5, 2.1)

Problem: Break, Persistence, etc … $\mu = \mu(t)$

…. and possibly $\sigma^2 = \sigma^2(t)$

Sources of uncertainty about the future:

(1) Future shocks: $\mu \pm 1.96\sqrt{\sigma_{LR}^2 / h}$

(2) Stochastic Process Describing *x*:

(a) Parameters: $\mu \pm 1.96\sqrt{\sigma_{LR}^2 / h}$

(b) "Breaks, shifts, persistence, etc." $\mu(t)$ and $\sigma^2(t)$

- Challenge: There is limited information in the sample relevant long-run properties of stochastic process.

    - 60 years of sample data
    - forecasting horizon is 25, 50, or 75 years

- How do we extract relevant long-run information from sample?

- Even after seeing the sample data there will be considerable uncertainty about long-run properties of stochastic process. How does this "model uncertainty" affect uncertainty about $\overline{x}_{T+1:T+h}$?

Approach Used Here

Summarize the "long-run information" in the sample data using a few "low-frequency" weighted averages. Call these averages X.

Prediction uncertainty:

- Shocks: (X, $\overline{x}_{T+1:T+h}$) are normally distributed (CLT)

- Stochastic Process: Specify a parametric model that is sufficiently flexible to capture many types of persistence, breaks, etc. (This is not too difficult as only low-frequency behavior matters.)

    - "Model uncertainty" becomes "parameter uncertainty", which can be measured using standard statistical methods (Bayes, Frequentist, etc.)

Key Steps in Our Approach / Outline of talk

(1) Summarize the sample data using low-frequency averages

(2) Characterize (large-sample) joint distribution of sample data and forecast variable. (This is a normal distribution with covariance matrix that depends on local-to-zero spectrum).

(3) Parameterize low-frequency spectrum

(4) Construct Prediction Sets

- Bayes
- Frequentist

(5) Empirical Results

Step 1: Summarize the sample data using low-frequency averages

$$x_t = \mu + u_t$$

Weighting function: $\Psi_j(s) = \sqrt{2}\cos(j\pi s)$

Weighted average of sample data:

$$X_T(j) = \int_0^1 \Psi_j(s) x_{[sT]+1} ds = l_{jT} T^{-1} \sum_{t=1}^{T} \Psi\left(\frac{t-0.5}{T}\right) x_t$$

where $l_{jT} = (2T/j\pi)\sin(j\pi/2T) \to 1$.

Non-singular transformation of sample data:

$(\bar{x}_{1:T}, X_T(1), X_T(2), \ldots X_T(T-1))$

Notes:

(1) $\bar{x}_{1:T}$ captures 0-frequency variability is series. $X_T(j)$ captures variability at frequency $j\pi/T$.

(2) $X_T(1)$ … $X_T(k)$ captures variability for periods longer than $2T/k$

(3) $\int_0^1 \Psi_j(s)ds = 0$, so $X_T(j)$ is location invariant (… is invariant to the value of μ).

Truncating information in sample data: We will use

$$\boldsymbol{X_{T,1:q} = [\ X_T(1)\ \ldots\ X_T(q)\ ]}$$

for q small and discard sample information $X_T(q+1) \ldots X_T(T-1)$.

- Empirical work uses $\boldsymbol{q = 12}$. (For T = 60 years, this uses periodicities greater than $2T/q$ = 10 years.)

- Benefits of truncating sample information
  - Robustness/Parameterization
  - Tractability

- Information loss
  - Prediction conditional on parameter values
  - Information about parameter values
  - Quantifying losses – Discussed below (not really – see paper)

Key Steps in Our Approach / Outline of talk

(1) ~~Summarize the sample data using low-frequency averages~~

(2) Characterize (large-sample) joint distribution of sample data and forecast variable. (This is a normal distribution with covariance matrix that depends on local-to-zero spectrum).

(3) Parameterize low-frequency spectrum

(4) Construct Prediction Sets

- Bayes
- Frequentist

(5) Empirical Results

Step 2: Characterize (large-sample) joint distribution of sample data and forecast variable. (This is a normal distribution with covariance matrix that depends on local-to-zero spectrum).

Variables of interest: $\boldsymbol{X_{T,\,1:q}}$ and $\boldsymbol{Y_T} = \bar{x}_{T+1:T+h} - \bar{x}_{1:T}$

(location invariance)

CLT Assumptions: $\Delta x_t = c(\mathrm{L})\varepsilon_t$ ($\varepsilon_t \sim$ mds with $2+\delta$ moments, etc, $c(\mathrm{L})$ weights are square-summable, so Δx_t has spectrum, see paper.)

$$T^{-\alpha}\begin{bmatrix} X_{T,1:q} \\ Y_T \end{bmatrix} \Rightarrow \begin{bmatrix} X \\ Y \end{bmatrix} \sim N(0,\Sigma)$$

Charactering Σ

$$x_t = \mu + u_t$$

Each element of $X_{T,1:q}$ or Y_T can be written as $\eta_T = T^{-\alpha}\sum_{t=1}^{[(1+r)T]} g_t u_t$ for appropriate weights g_t and with $r = h/T$.

Suppose u_t has spectrum $R(\lambda)$ (appendix extends to Δu_t has spectrum)

m'th autocovariance is $\int_{-\pi}^{\pi} R(\lambda) e^{-i\lambda m} d\lambda$

$$\Sigma_{jk} = \mathrm{E}(\eta_{jT}, \eta_{kT}) = E\left(T^{-\alpha}\sum_{t=1}^{[(1+r)T]} g_{j,t}u_t\right)\left(T^{-\alpha}\sum_{s=1}^{[(1+r)T]} g_{k,s}u_s\right)$$

$$\Sigma_{jk} = T^{-2\alpha} \sum_{s,t=1}^{[(1+r)T]} \left(\int_{-\pi}^{\pi} e^{-i\lambda(t-s)} R(\lambda) d\lambda \right) g_{j,t} g_{k,s}$$

$$= T^{-2\alpha} \int_{-\pi}^{\pi} R(\lambda) \left(\sum_{s=1}^{[(1+r)T]} g_{k,s} e^{i\lambda s} \right) \left(\sum_{t=1}^{[(1+r)T]} g_{j,t} e^{-i\lambda t} \right) d\lambda$$

$$= T^{1-2\alpha} \int_{-T\pi}^{T\pi} R(\omega/T) \left(T^{-1} \sum_{s=1}^{[(1+r)T]} g_{k,s} e^{i\omega(s/T)} \right) \left(T^{-1} \sum_{t=1}^{[(1+r)T]} g_{j,t} e^{-i\omega(t/T)} \right) d\omega$$

$$\to \int_{-\infty}^{\infty} S(\omega) \left(\int_0^{1+r} g_k(s) e^{i\omega s} ds \right) \left(\int_0^{1+r} g_j(s) e^{-i\omega s} ds \right) d\omega$$

$$= \int_{-\infty}^{\infty} S(\omega) w_{jk}(\omega) d\omega$$

where $S(\omega) = \lim_{T\to\infty} T^{1-2\alpha} R(\omega/T)$ is the "local-to-zero spectrum" of u_t.

which is weighted by $\boldsymbol{w_{jk}(\omega)} = \left(\int_0^{1+r} g_k(s) e^{i\omega s} ds \right) \left(\int_0^{1+r} g_j(s) e^{-i\omega s} ds \right)$.

Key features of $w_{jk}(\omega)$ are given in Figure 1 in paper.

Key Steps in Our Approach / Outline of talk

(1) ~~Summarize the sample data using low-frequency averages~~

(2) ~~Characterize (large-sample) joint distribution of sample data and forecast variable. (This is a normal distribution with covariance matrix that depends on local-to-zero spectrum).~~

(3) Parameterize low-frequency spectrum

(4) Construct Prediction Sets

- Bayes
- Frequentist

(5) Empirical Results

Step 3: (3) Parameterize low-frequency spectrum – use "abd" model

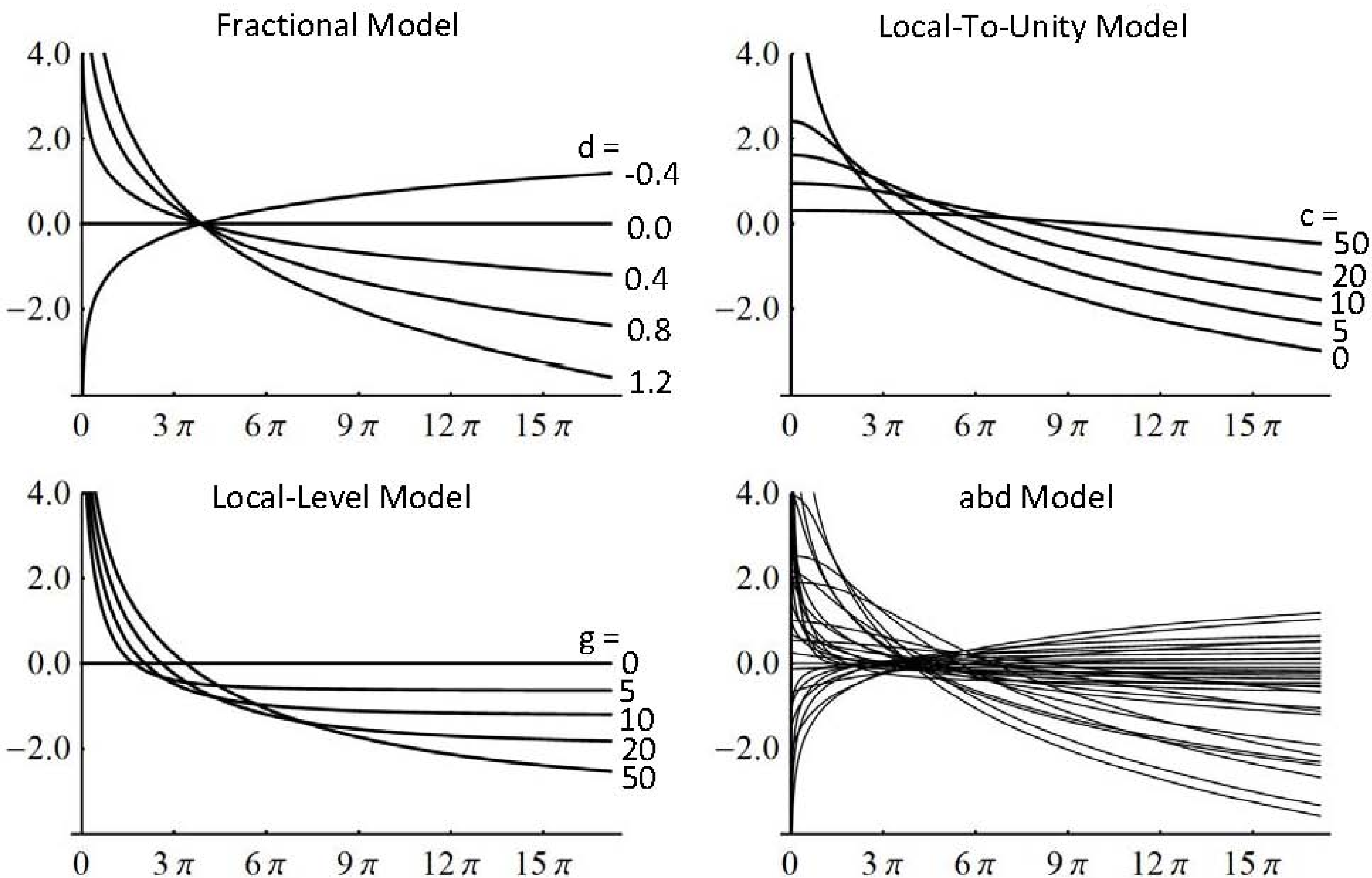
- Local Level: $\boldsymbol{x_t} = \boldsymbol{\mu_t} + \boldsymbol{u_t}$ (Permanent/I(1) + Transitory/I(0))
  - *GDP growth rates, Inflation*
  - $S(\omega) \propto b^2 + \omega^{-2}$

- Persistent AR: $\boldsymbol{x_t = \rho x_{t-1} + error}$ with ρ close to 1
  - *Interest Rates*
  - $S(\omega) \propto (\omega^2 + c^2)^{-1}$

- Long-Memory: $\mathbf{(1-L)}^{\boldsymbol{d}} \boldsymbol{x_t = error}$
  - I(0) $d = 0$; I(1) $d = 1$
  - $0 < d < 1$, persistent in between I(0) and I(1)
  - $d < 0$, "anti-persistence" (*GDP growth rates* ?)
  - $S(\omega) \propto |\omega|^{-2d}$

Because there is limited sample information it is hopeless to discriminate between these models (and/or others). Model uncertainty is an important part of forecast uncertainty.

- **"*abd*-model":** $S(\omega) \propto \left(|\omega|^{2d} + a^2\right)^{-1} + b^2$


Fractional Model
d =
-0.4
0.0
0.4
0.8
1.2
Local-To-Unity Model
c =
50
20
10
5
0
Local-Level Model
g =
0
5
10
20
50
abd Model
4.0
2.0
0.0
−2.0
0
3 π
6 π
9 π
12 π
15 π

Key Steps in Our Approach / Outline of talk

(1) ~~Summarize the sample data using low-frequency averages~~

(2) ~~Characterize (large-sample) joint distribution of sample data and forecast variable. (This is a normal distribution with covariance matrix that depends on local-to-zero spectrum).~~

(3) ~~Parameterize low-frequency spectrum~~

(4) Construct Prediction Sets

- Bayes
- Frequentist

(5) Empirical Results

(4) *Construct Prediction Sets (Bayes and Frequentist)*

Problem: $\min_A \boldsymbol{E}\left(vol\left(A(X^s)\right)\right)$ subject to $\boldsymbol{P}\left(Y^s \in A(X^s)\right) \geq 1-\alpha$

Bayes and Frequentist methods differ in the choice of distributions used for ***E*** and ***P*****.**

Problem: $\min_A E\left(vol\left(A(X^s)\right)\right)$ subject to $P\left(Y^s \in A(X^s)\right) \geq 1-\alpha$

Bayes Solution

Let $\theta = \{a,b,d\}$. Let Γ denote a prior for θ. Bayes solution uses predictive distribution for Y^s given X^s to compute both E and P.

Bayes Predictive Set:

$$A_\Gamma^{Bayes}(x^s) = \left\{ y^s : \frac{f^\Gamma_{(Y^s,X^s)}(y^s,x^s)}{f^\Gamma_{X^s}(x^s)} > cv(x^s) \right\}$$

$cv(x^s)$ is "critical value" that depends on x^s because coverage constraint uses predictive distribution of $Y^s|X^s = x^s$.

Notes on Bayes Solution:

(1) Coverage constraint satisfied for each value of x^s.

(2) Coverage constraint satisfied "on average" over $\theta \sim \Gamma$

(3) $I(0)$ special case: Y^s is independent of X^s (with t-distribution) and Prediction set for $\bar{x}_{T+1:T+h}$ is

$$\boxed{\bar{x}_{1:T} \pm t^q_{(1-\alpha/2)}(1+r^{-1})^{1/2}T^{-1/2}s_{LR}}$$

where $s^2_{LR} = (T/q)X_{T,1:q}{}'X_{T,1:q}$ is an estimate of the long-run variance of x.

Frequentist Solution

Coverage Constraint: $Inf_{\theta \in \Theta} P_{\theta}\left(Y^{s} \in A(X^{s})\right) \geq 1-\alpha$

(Coverage constraint must hold over all values of **"*a,b,d*"…** that is, over all the models parameterized by *a*,*b*,*d*.)

Details for constructing frequentist prediction set is in paper …

Here we modify (enlarge) the Bayes sets so that they satisfy coverage constraint. (Optimality: Müller and Norets (2012), computational methods for "Approximate Least Favorable Distribution" in Elliott, Müller, and Watson (2012))

Key Steps in Our Approach / Outline of talk

(1) ~~Summarize the sample data using low-frequency averages~~

(2) ~~Characterize (large-sample) joint distribution of sample data and forecast variable. (This is a normal distribution with covariance matrix that depends on local-to-zero spectrum).~~

(3) ~~Parameterize low-frequency spectrum~~

(4) ~~Construct Prediction Sets~~

- ~~Bayes~~
- ~~Frequentist~~

(5) Empirical Results

Step 5: Empirical Results

- $q = 12$

- Bayes set uses *prior*: $d \sim U[-0.4, 1.4]$, $a = b = 0$

- Frequentist set uses approximate "least favorable distribution" that is computed numerically for $-0.4 \leq d \leq 1.4$ and all values of a and b. (So coverage is uniform over these values.)

Series:

- Post WWII Quarterly observation on growth rates of real per capita GDP, consumption, productivity (TFP, Labor Prod); inflation rates (US and Japan); stock returns.

- Longer-run annual series beginning in early 20th century and many of the same series

- *h* forecast horizons of 10-75 years.

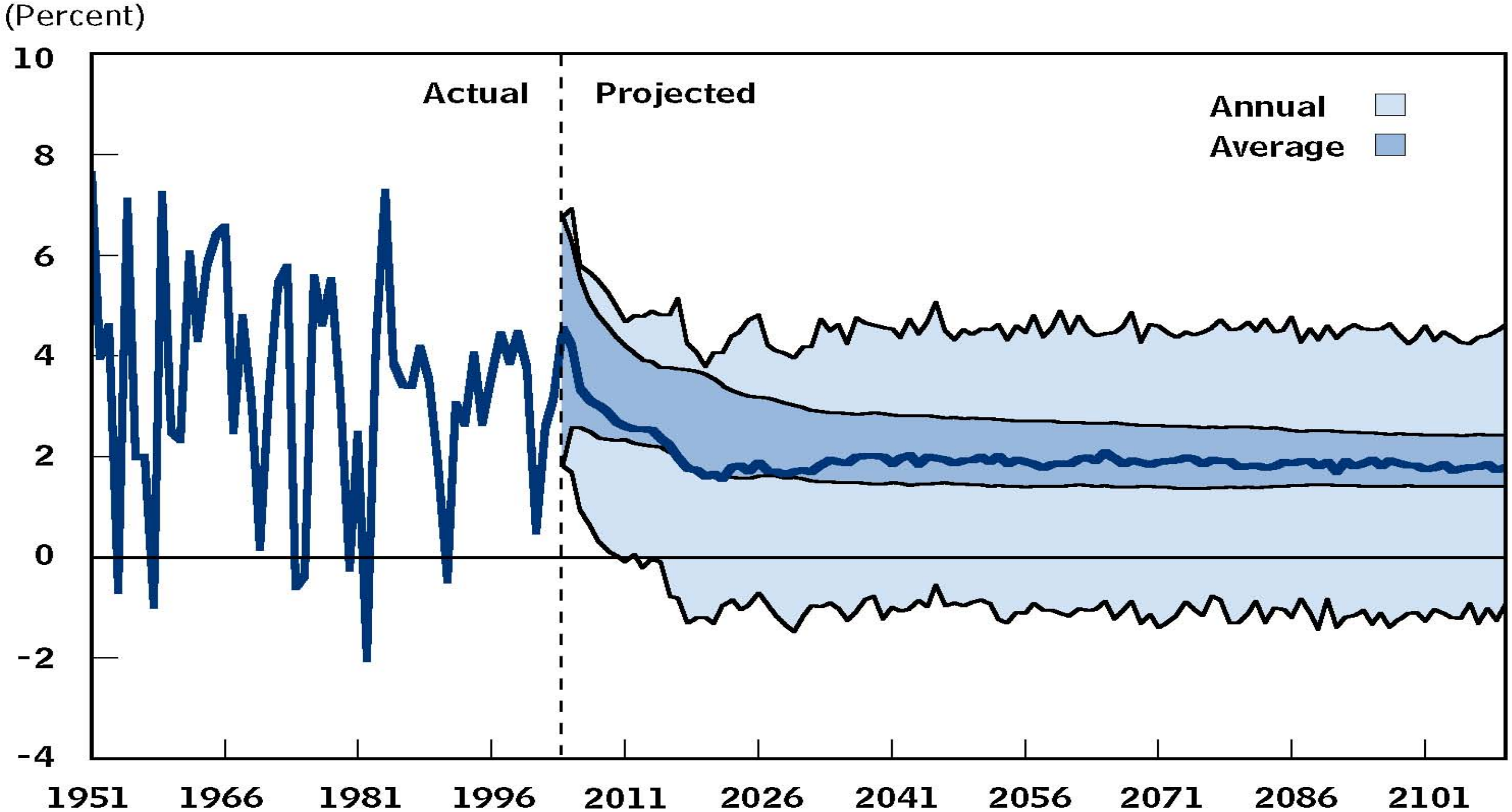
Summary Statistics and I(0) Predictive Sets Selected Post-WWII Series (See Table 2 for all Series)

Series	$\overline{x}_{1:T}$	s_{LR}	Predictive Sets: I(0) Model 25-Year Horizon		
			50%	80%	90%
GDP/Pop	1.9	5.1	(1.5, 2.3)	(1.1, 2.7)	(0.8, 3.0)
TF Prod	1.3	4.0	(1.0, 1.6)	(0.7, 1.9)	(0.5, 2.1)
Inflation (CPI)	3.6	10.4	(2.8, 4.5)	(2.0, 5.3)	(1.5, 5.8)

CBO 2005

Figure 14.

Uncertainty Bands for Real GDP Growth

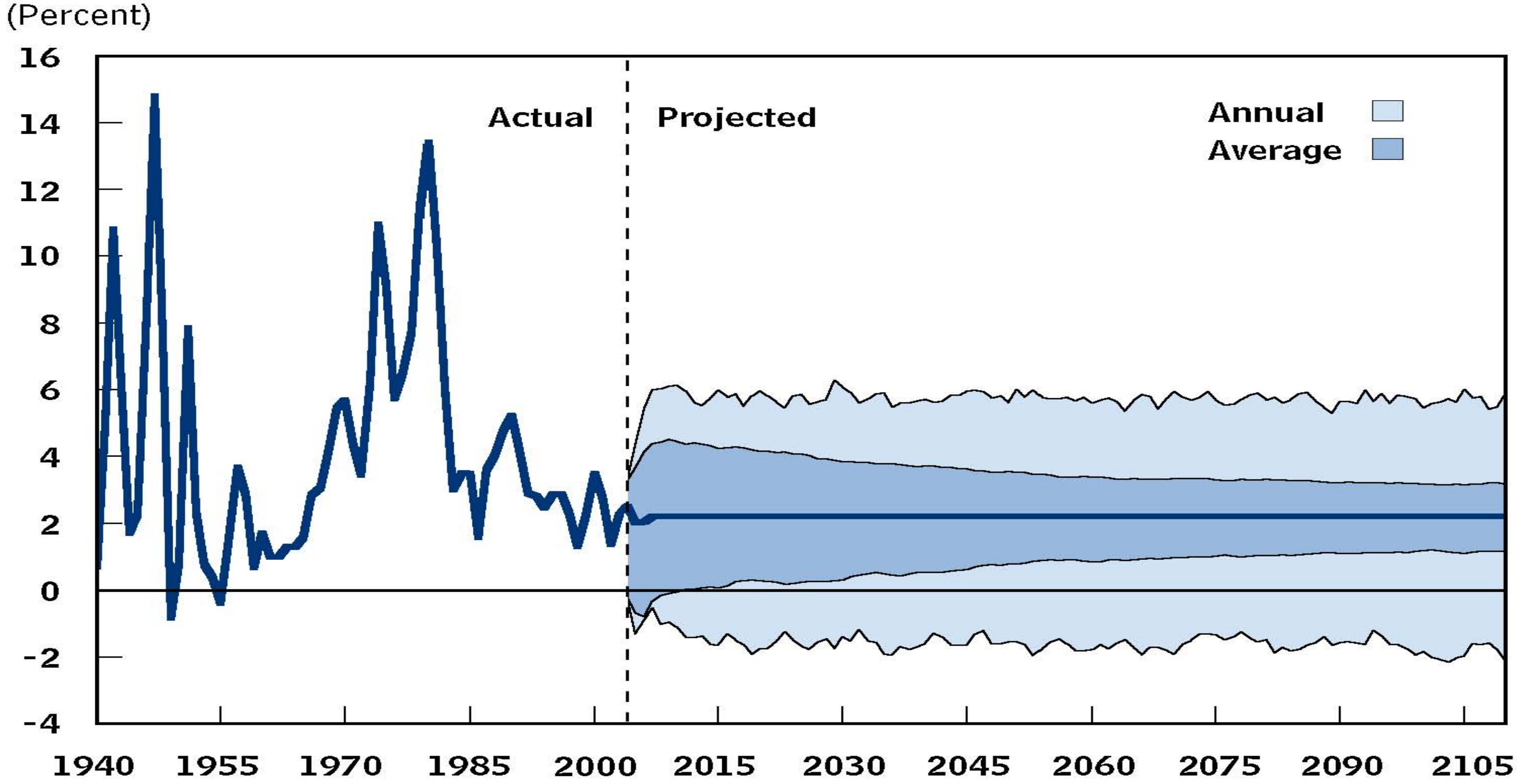


Source: Congressional Budget Office.

Note: Annual uncertainty bands show the 80 percent confidence range for a given year. Average uncertainty bands show the 80 percent confidence range for the average of 2004 through a given year.

Figure 11.

Uncertainty Bands for the Inflation Rate



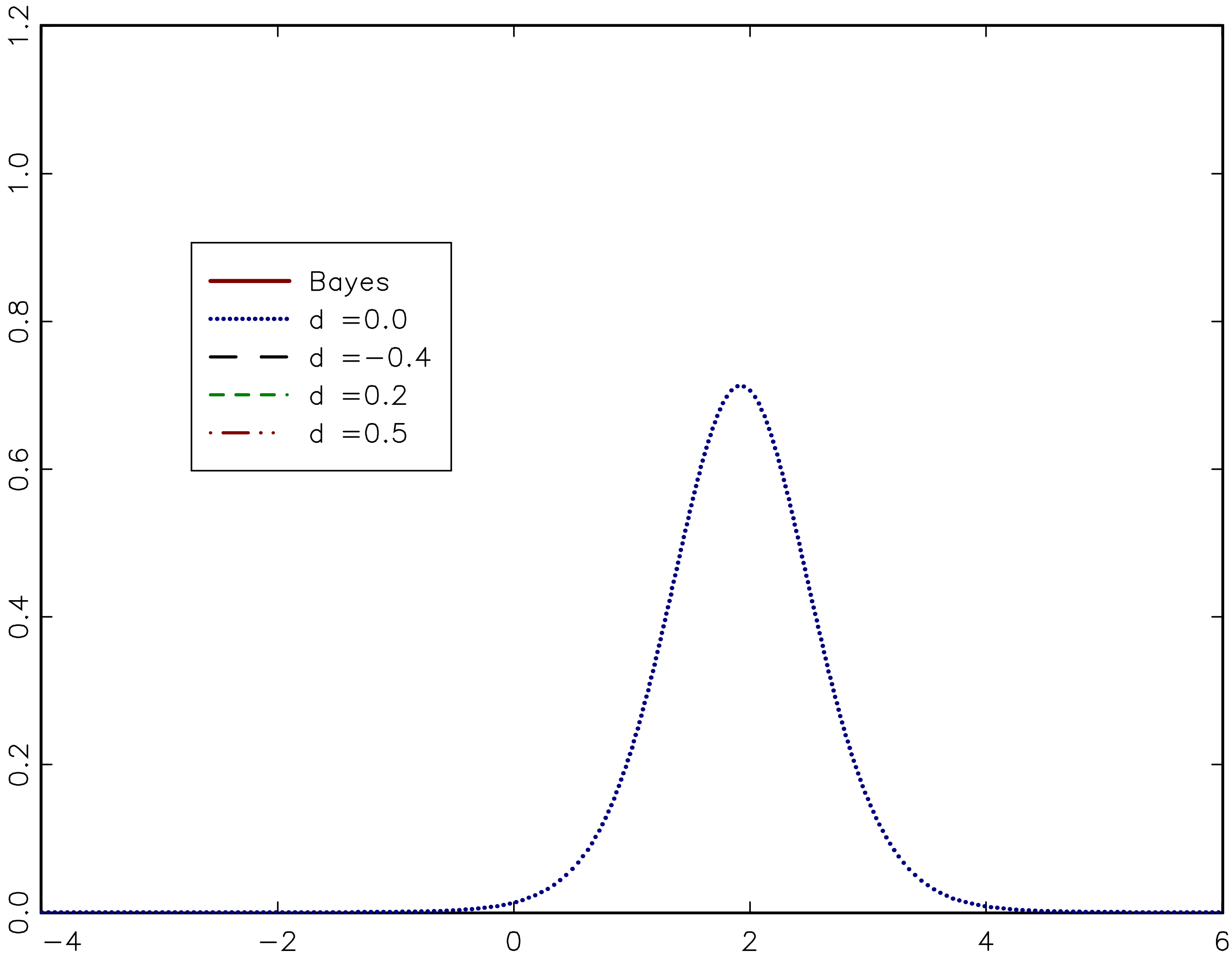
Source: Congressional Budget Office.

Note: Annual uncertainty bands show the 80 percent confidence range for a given year. Average uncertainty bands show the 80 percent confidence range for the average of 2004 through a given year.

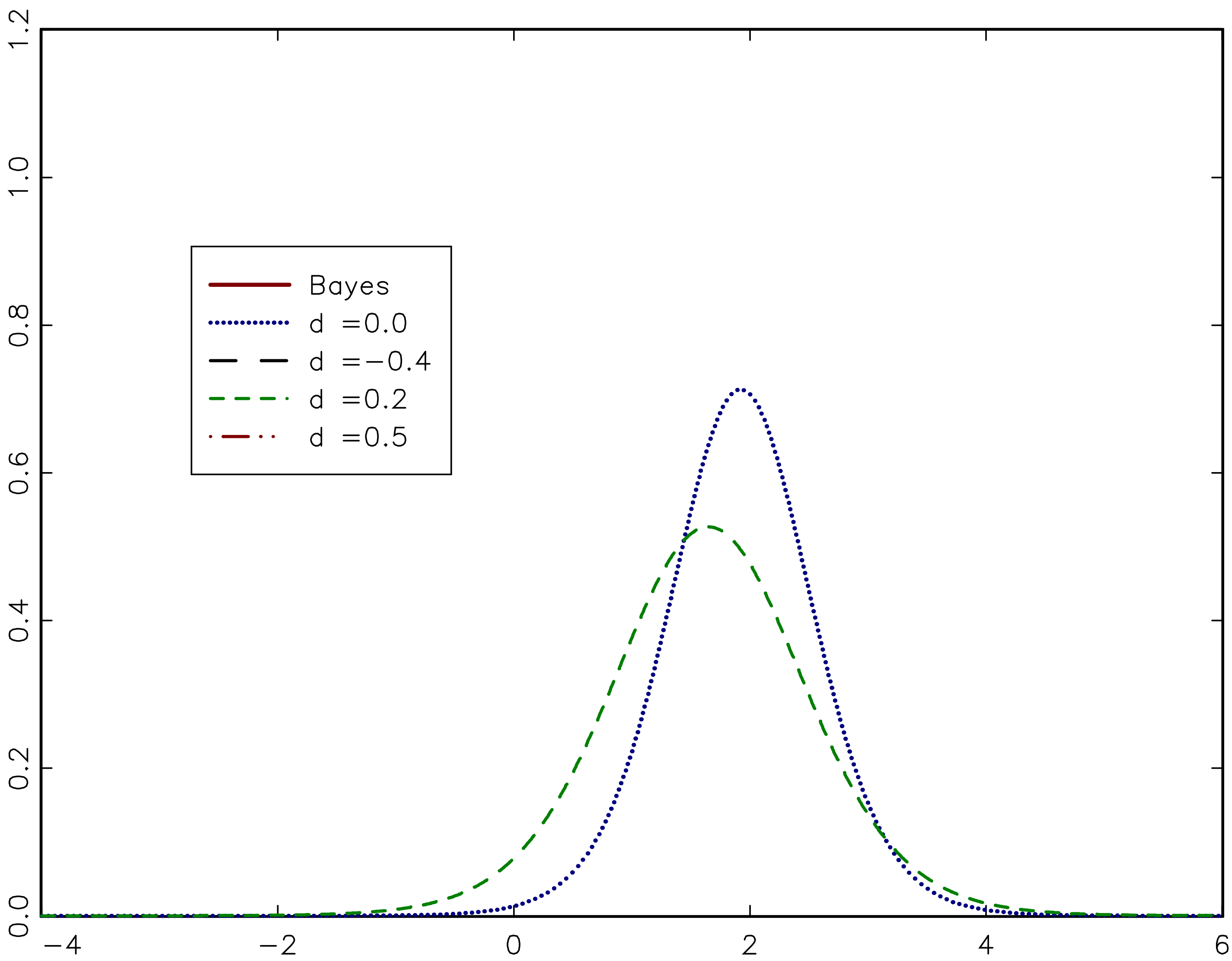
Table 3: Log Likelihood Values for I(*d*) model

Series	*d*									
	-0.4	**-0.2**	**0.0**	**0.2**	**0.4**	**0.6**	**0.8**	**1.0**	**1.2**	**1.4**
GDP/Pop	0.5	0.4	0.0	-0.6	-1.6	-2.8	-4.3	-6.0	-8.1	-10.6
TFP	-2.6	-1.1	0.0	0.6	0.7	0.1	-0.9	-2.4	-4.2	-6.6
Inflation (CPI)	-3.3	-1.5	0.0	1.1	1.7	1.9	1.5	0.7	-0.5	-2.3
Stock Returns	-1.6	-0.6	0.0	0.2	0.0	-0.6	-1.5	-2.8	-4.4	-6.4

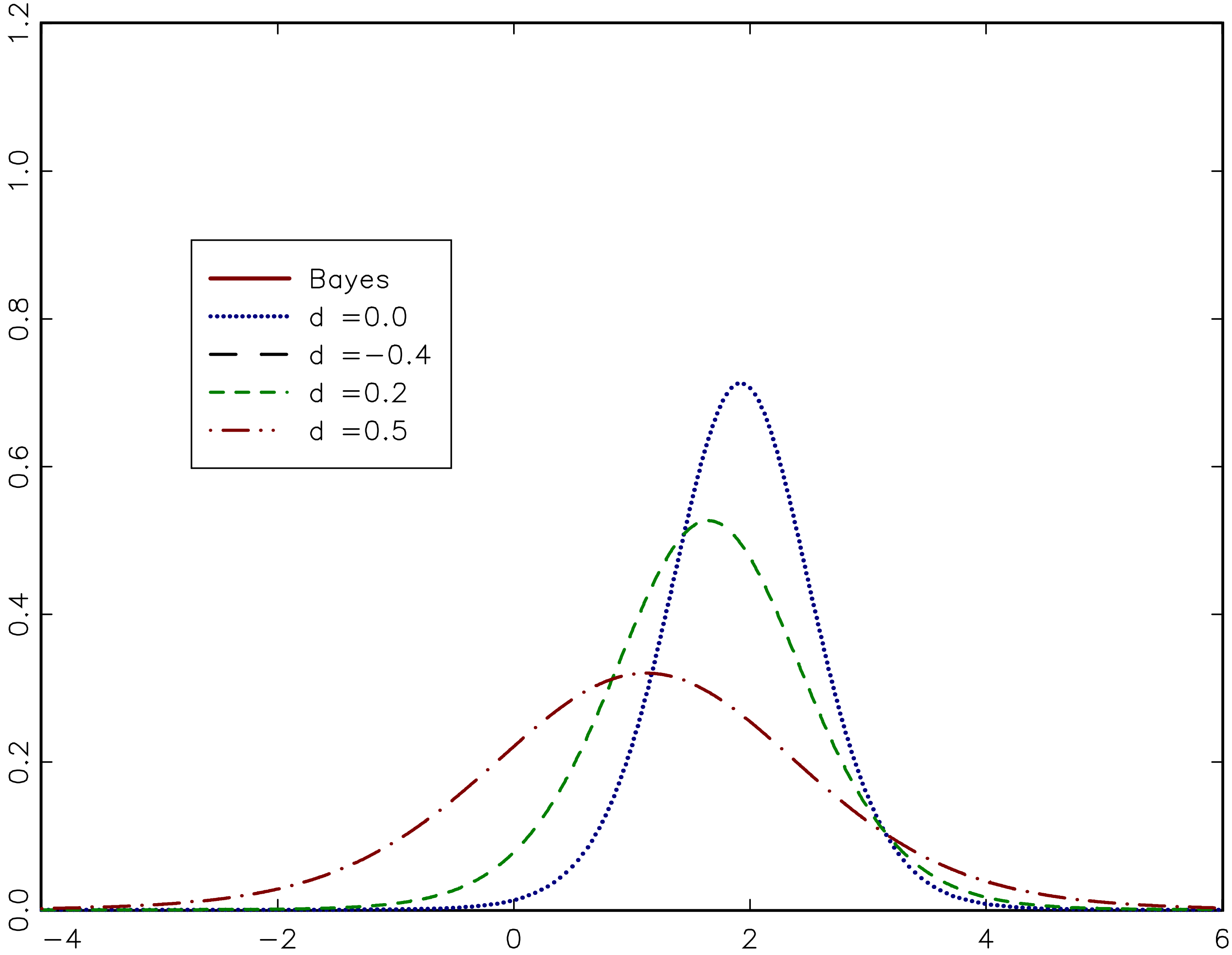
Predictive Density
Real per capita GDP – 25 year horizon



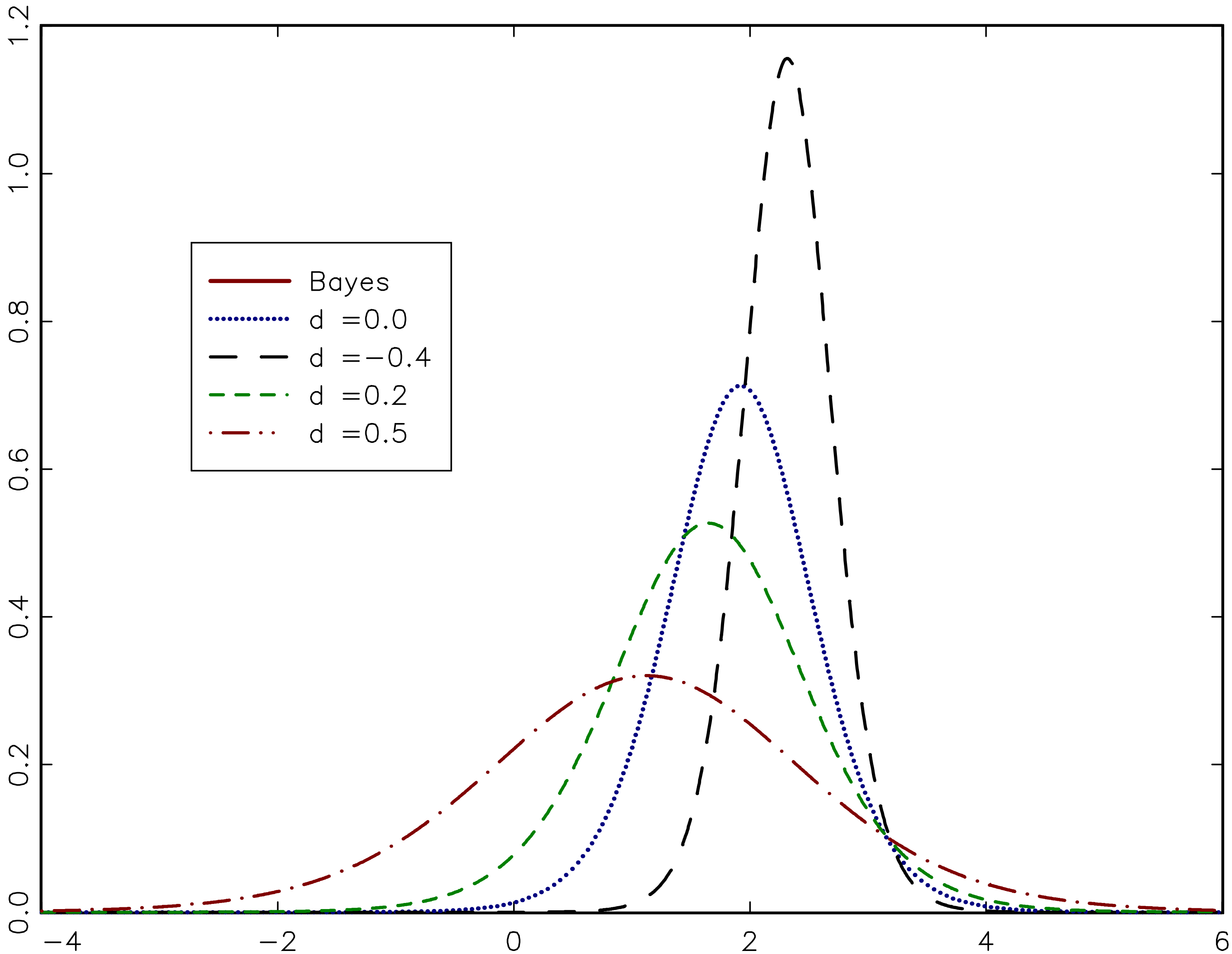
Predictive Density
Real per capita GDP – 25 year horizon



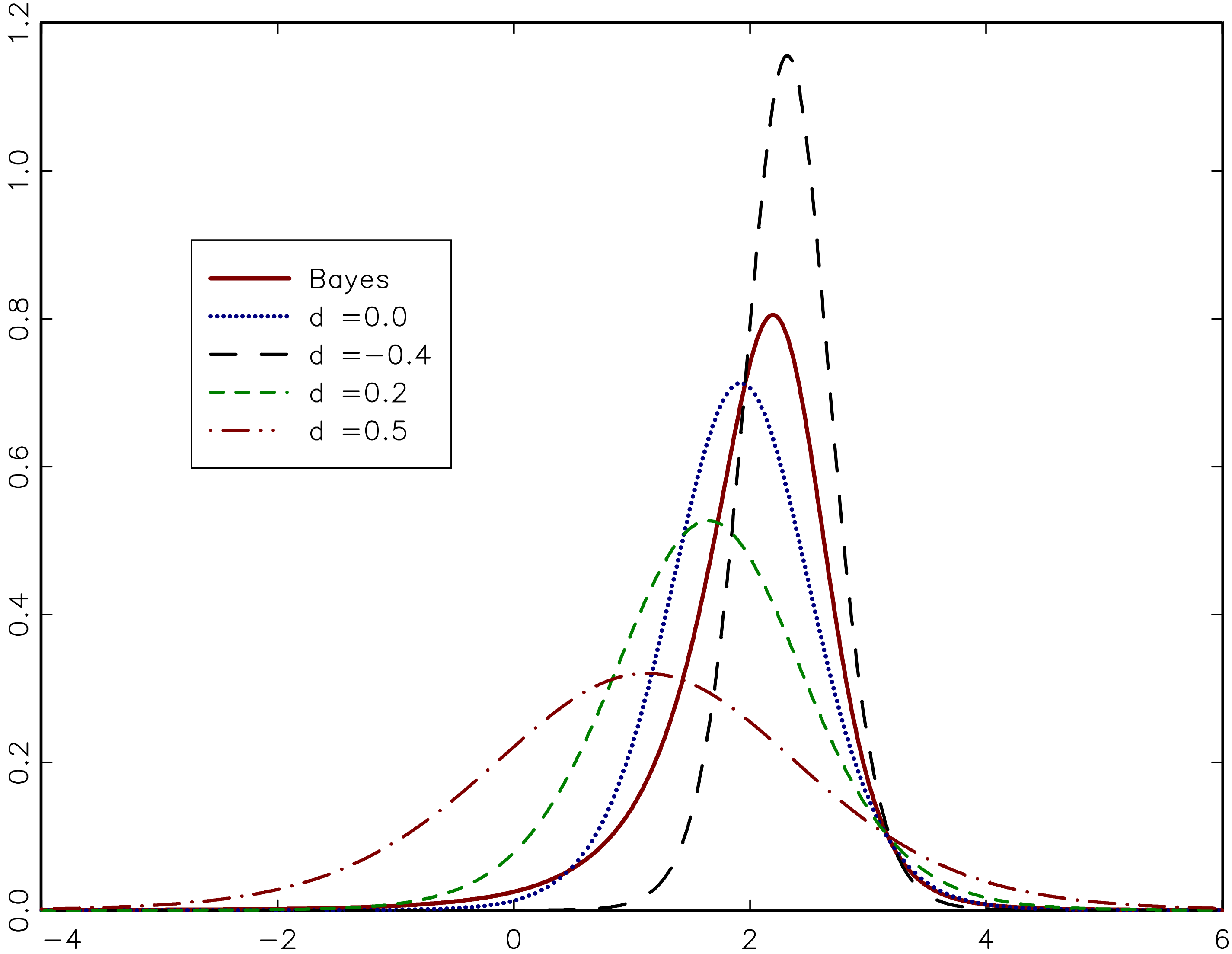
Predictive Density
Real per capita GDP – 25 year horizon



Predictive Density
Real per capita GDP – 25 year horizon

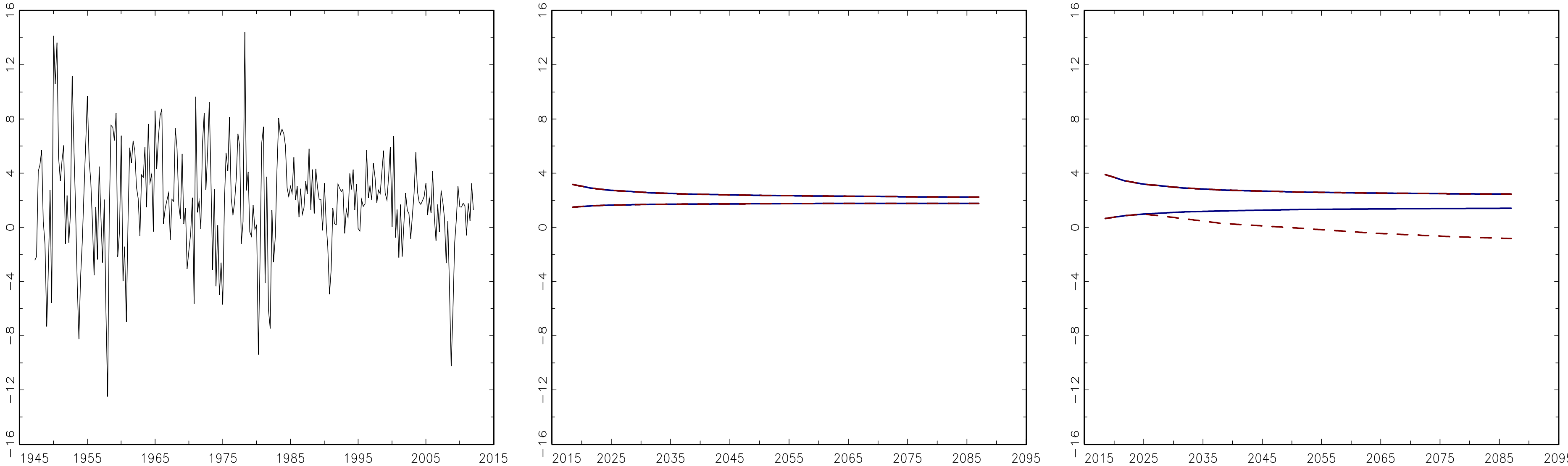


Predictive Density
Real per capita GDP – 25 year horizon



Bayes and Frequentist Sets: Real Per-Capita GDP

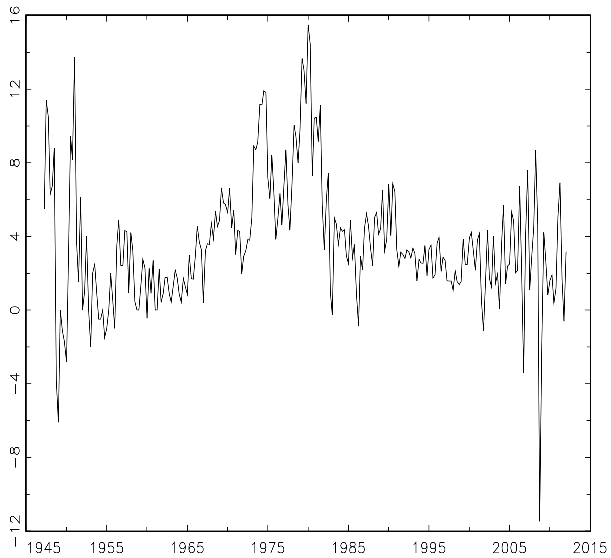
(i) Data (ii) 50% (iii) 80%



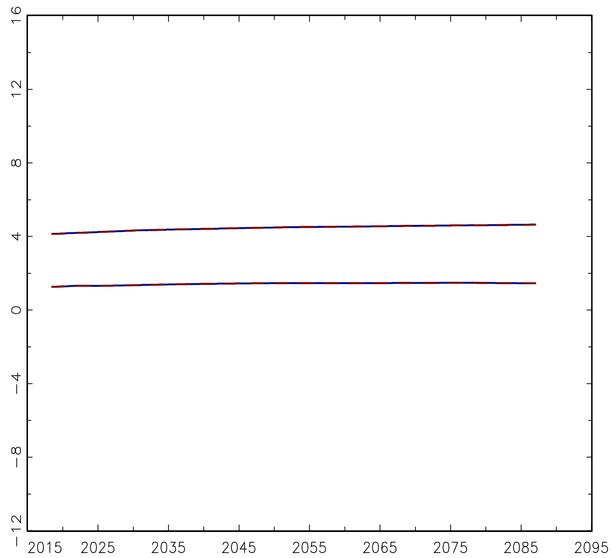
Bayes: Solid
MN-Frequentist: Dashed

Bayes and Frequentist Sets: CPI Inflation

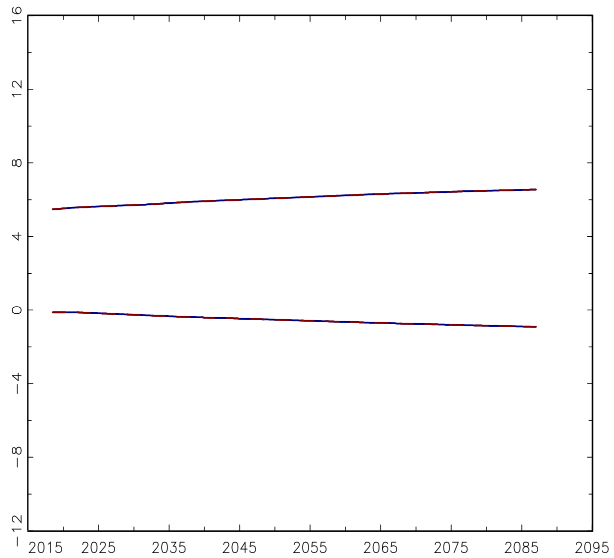
(i) Data



(ii) 50%



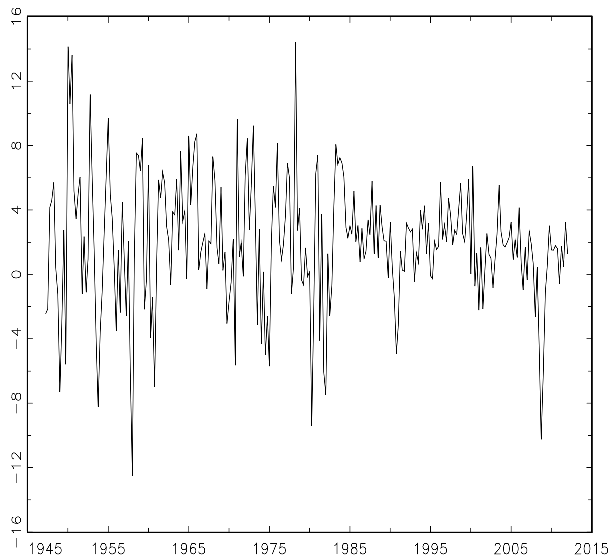
(iii) 80%



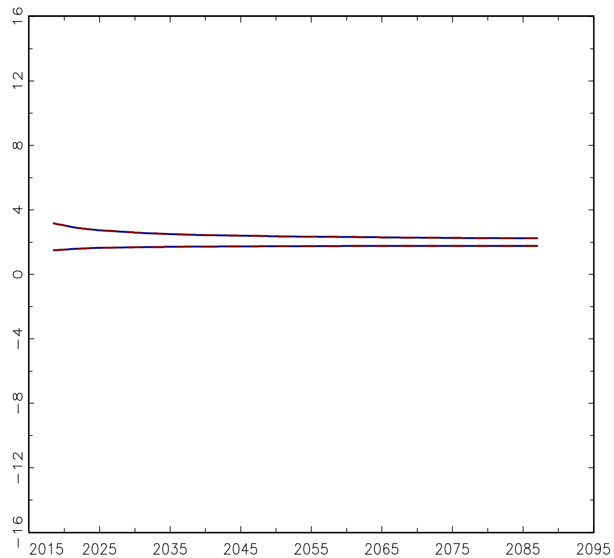
Bayes: Solid
MN-Frequentist: Dashed

Bayes and Frequentist Prediction Sets: Real Per-Capita GDP

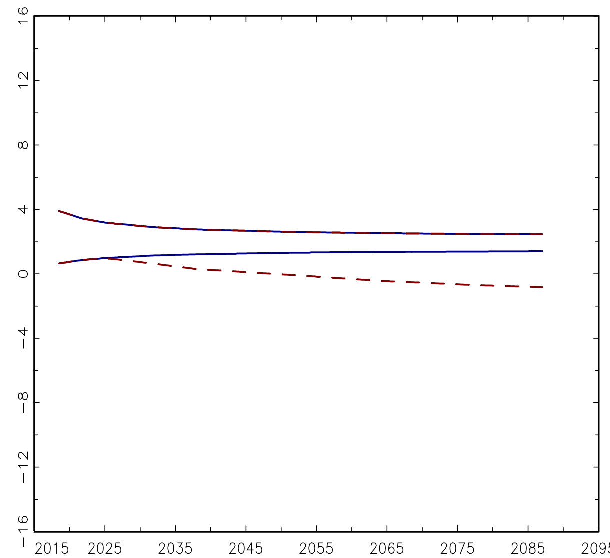
(i) Data



(ii) 50%



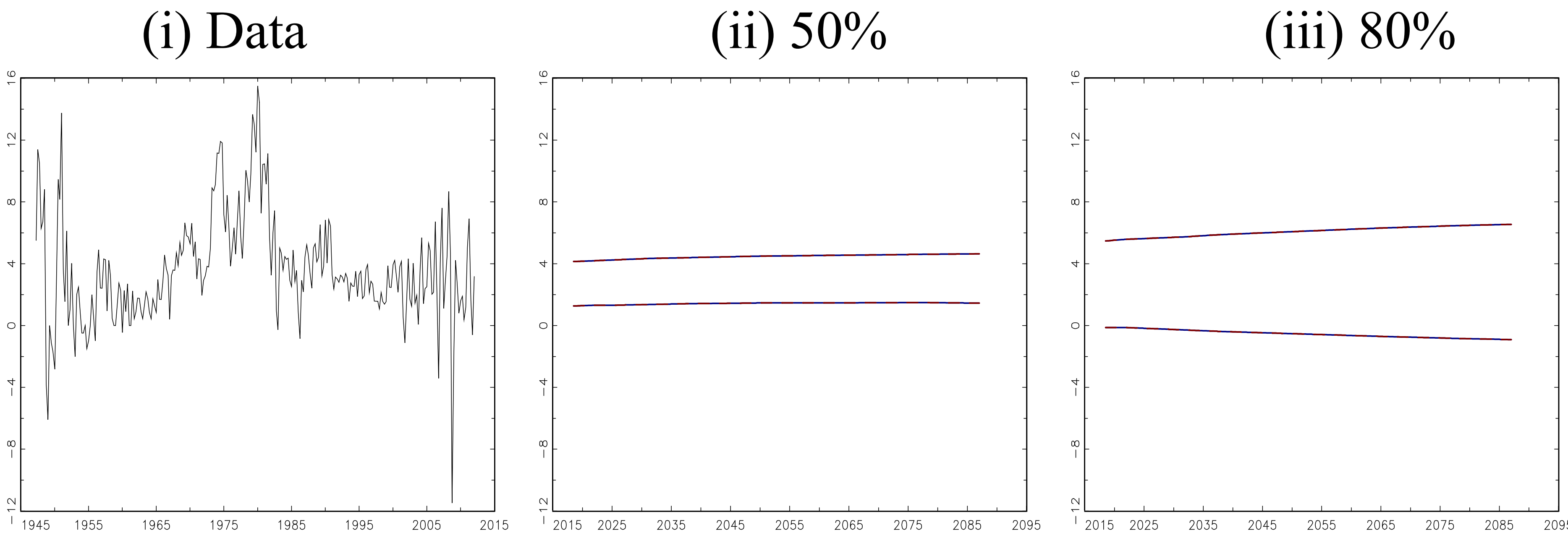
(iii) 80%



Bayes: Solid

MN-Frequentist: Dashed

Bayes and Frequentist Prediction Sets: CPI Inflation



Bayes: Solid
MN-Frequentist: Dashed

Some limitations of current work:

- No stochastic volatility
    - Perhaps not too important (low frequency)

- Univariate, with Multivariate extensions underway

Multivariate Example (Gordon BPEA (2003))

GDP = (GDP/Hrs) × (Hrs/Emp) × (Emp/LabFor) × (LabFor/Pop) × Pop

- Larger information set may lead to shorter unconditional prediction intervals – probably not a big deal

- Decomposition of forecast uncertainty

- Conditional Forecasts ("Scenarios)
  - Prediction Interval for GDP growth <u>given</u> a predetermined path (or interval) for Pop or (LabFor/Pop), etc.
    - But conditional distributions may not be "causal".

- Multivariate Extension conceptually straightforward, but there are technical challenges (numerical, parameterizing multivariate persistence).