Long-Run Covariability

Ulrich K. Müller and Mark W. Watson
Princeton University

September 2017

Motivation

- Study the long-run covariability/relationship between economic variables

  $\Rightarrow$ great ratios, long-run Phillips curve, nominal exchange rates and relative price levels, etc.

- Challenge: many economic time series are persistent

  $\Rightarrow$ spurious regression effects

- Cointegration framework is highly constraining

  $\Rightarrow$ Very specific model of persistence, rigid relationship between persistence and long-run covariability

This Paper

- Defines population measures about long-run covariability of general bivariate process

- Derives confidence intervals about these measures that are valid in flexible parametric model of long-run properties

  ⇒ "Descriptive statistics with confidence intervals"

- Statistical framework reflects sparsity of long-run information (cf. Müller and Watson (2008, 2016a, 2016b))

  ⇒ No consistent estimation of long-run properties

  ⇒ Inference in presence of nuisance parameters (Elliott, Müller and Watson (2015), Müller and Norets (2016))

Long-Run Correlation Matrix

		c	i	$w \times n$	TFP
y	$\hat{\rho}$	0.91*	0.53*	0.98*	0.78*
	90% CI	[0.71; 0.97]	[0.02; 0.81]	[0.95; 0.99]	[0.45; 0.95]
c	$\hat{\rho}$		0.53*	0.92*	0.70*
	90% CI		[0.03; 0.81]	[0.68; 0.97]	[0.28; 0.91]
i	$\hat{\rho}$			0.51*	0.38
	90% CI			[0.02; 0.80]	[−0.08; 0.71]
$w \times n$	$\hat{\rho}$				0.72*
	90% CI				[0.38; 0.93]

Outline of Talk

1. Introduction

2. Statistical framework, running examples

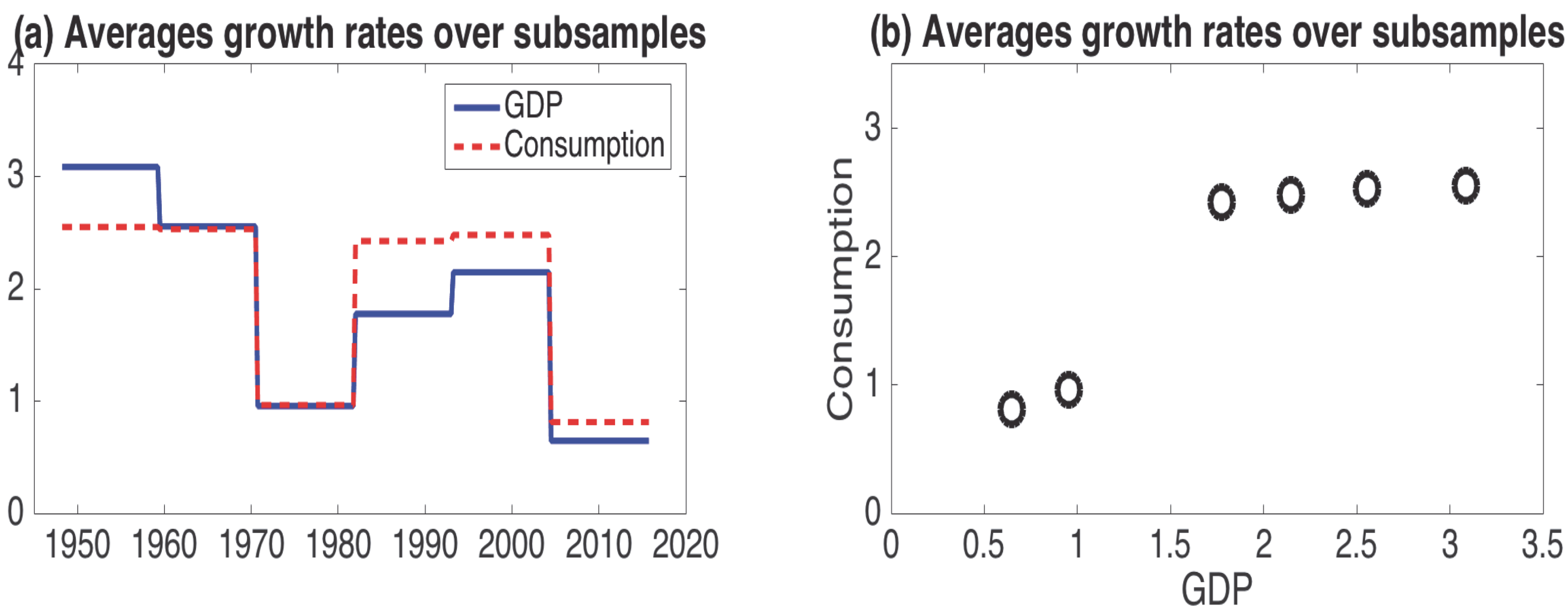
3. Measuring long-run covariability

4. A flexible parametric model of long-run properties

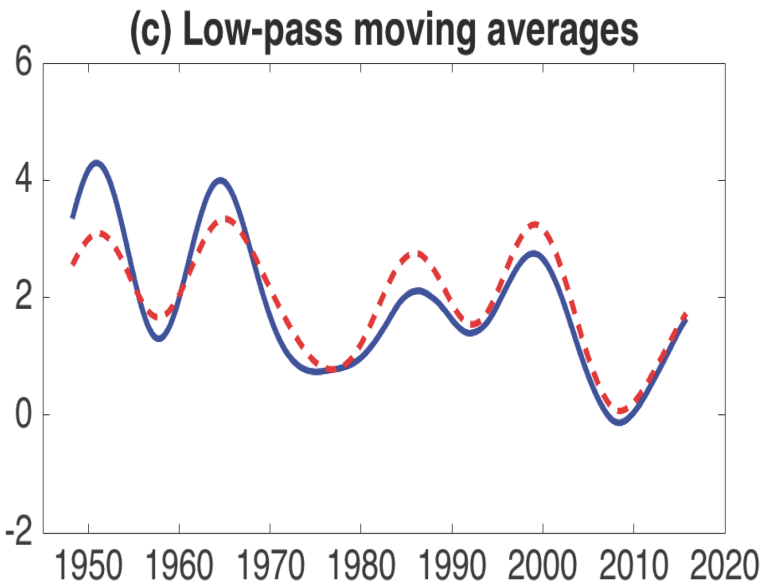
5. Construction of confidence intervals

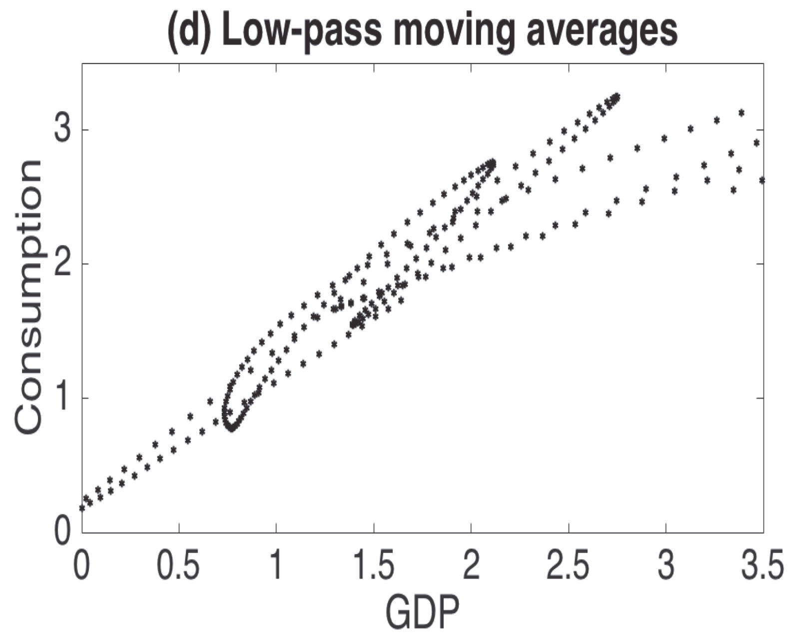
6. Additional applications

GDP and Consumption Growth

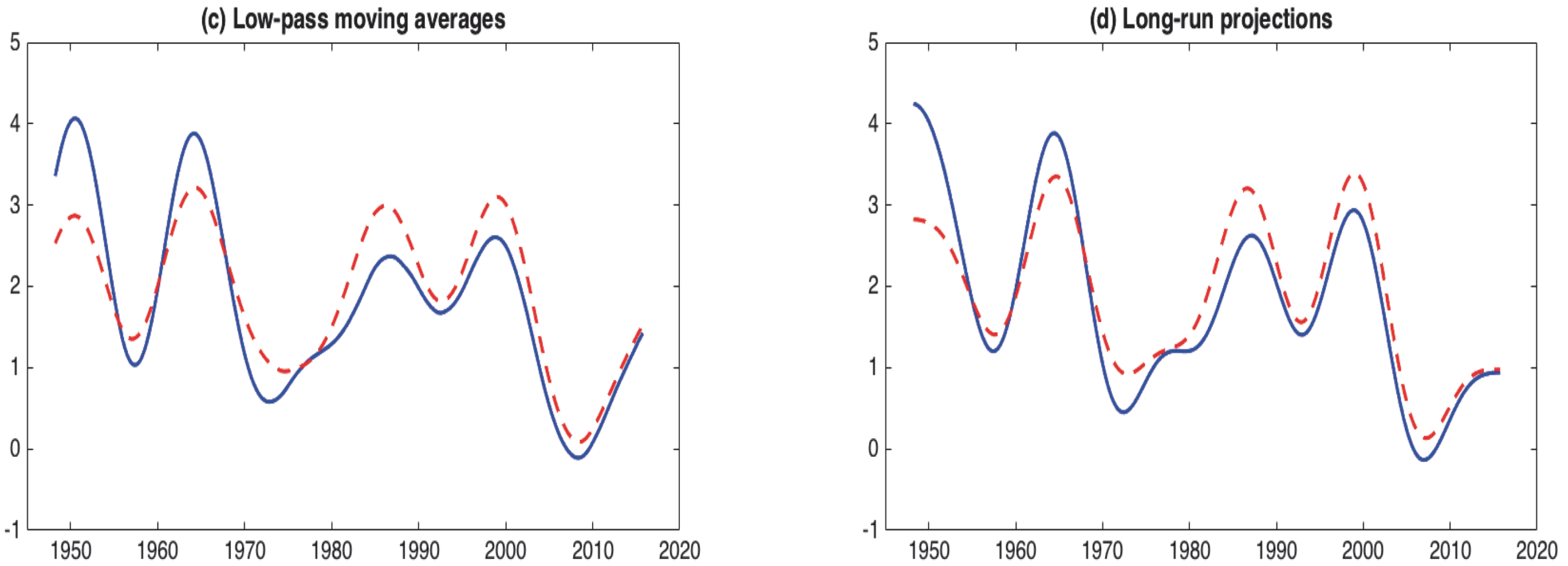


GDP and Consumption Growth





Low-Pass Moving Averages vs Long-Run Projections



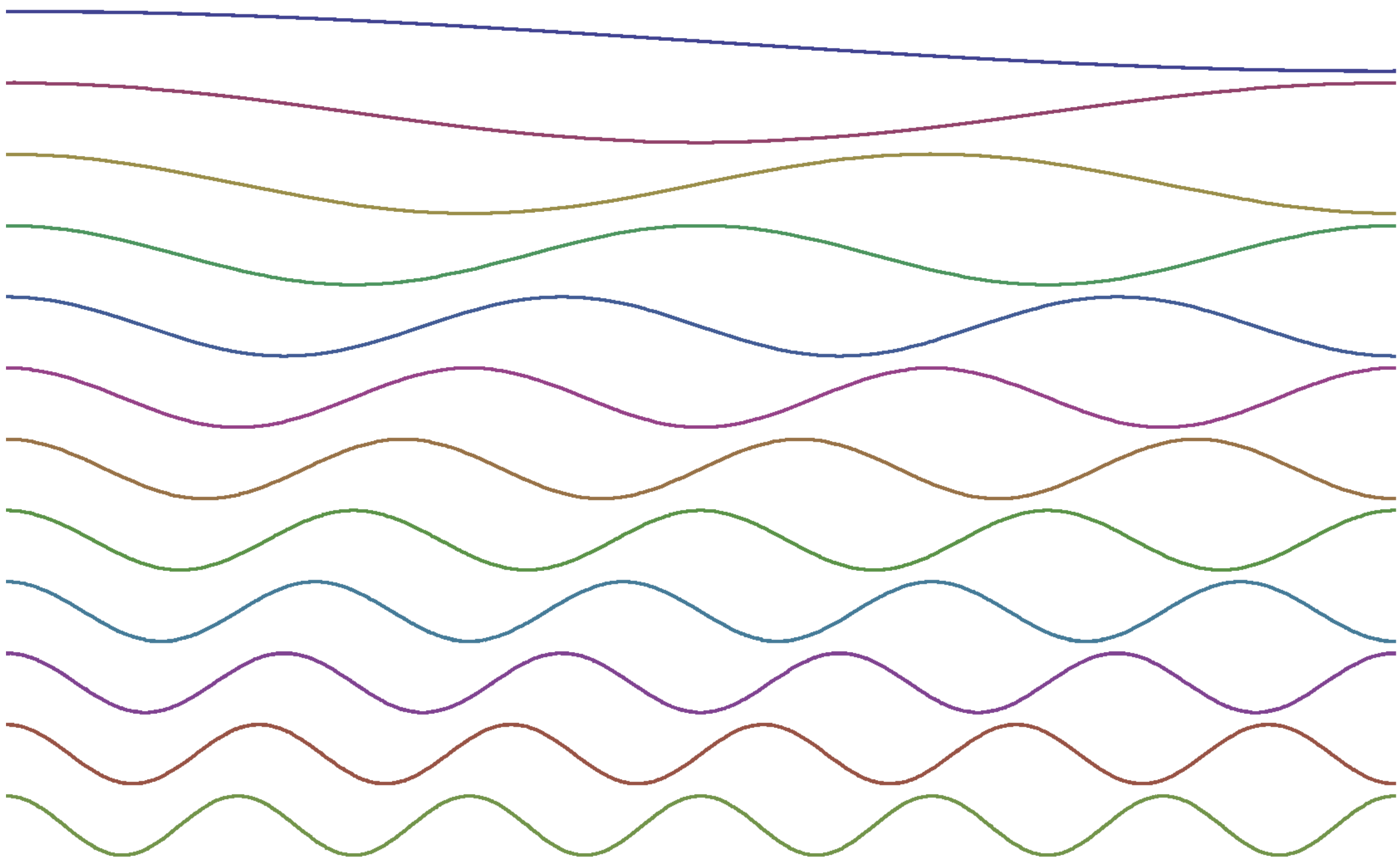
Cosine Transforms and Long-Run Projections

- Let $\Psi_j(s) = \sqrt{2}\cos(j\pi s)$

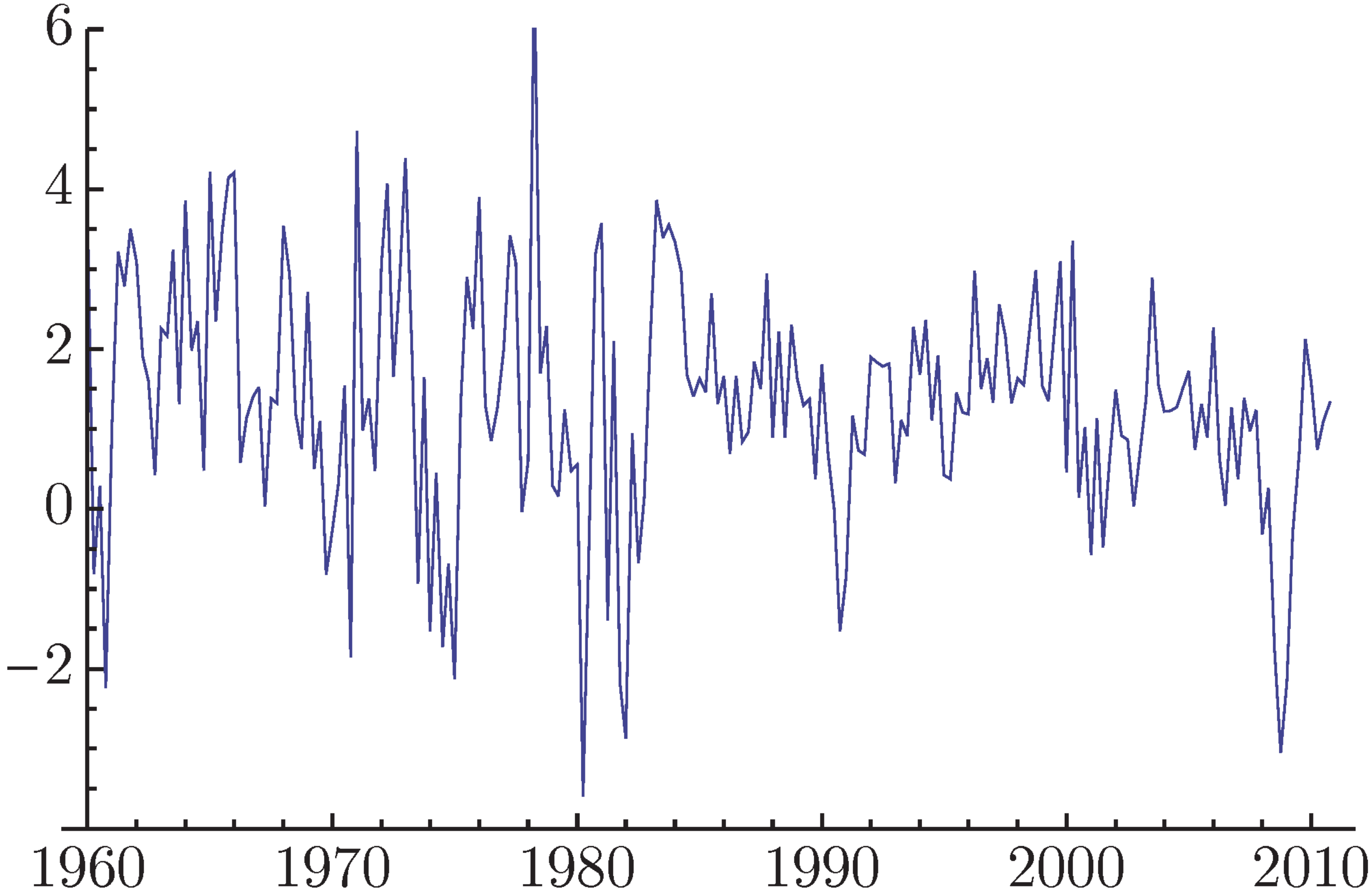
- For a scalar sequence $\{a_t\}_{t=1}^{T}$, define

$$
\begin{aligned}
A_j &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Psi_j(t/T)a_t \\
&= \frac{\sqrt{2}}{\sqrt{T}}\sum_{t=1}^{T}\cos(\pi j t/T)a_t, \quad j = 1, \cdots, q
\end{aligned}
$$

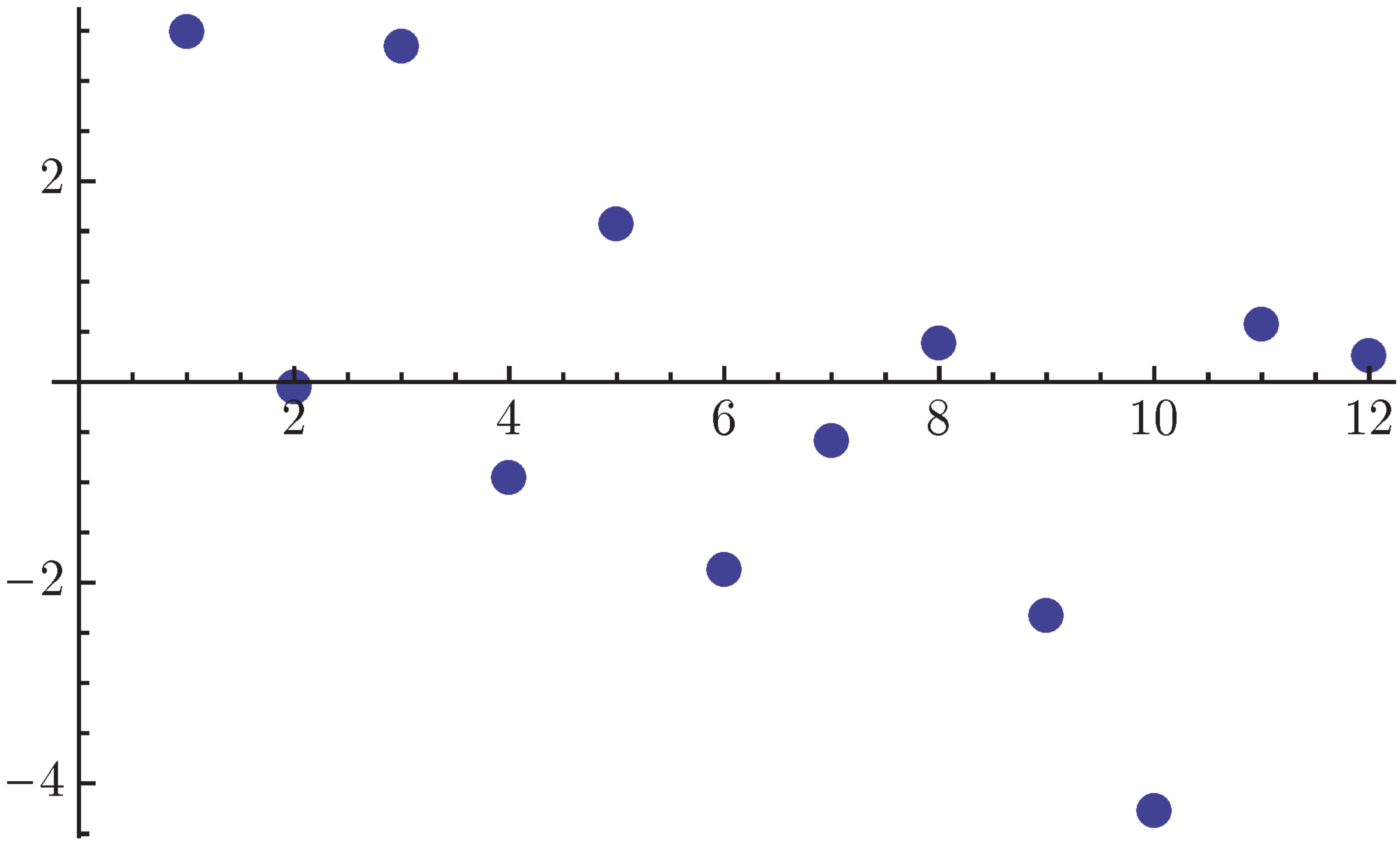
$q = 12$ Cosine Weights



GDP Growth



$q = 12$ Cosine Transforms for GDP



Long-Run Projections

- Define $\hat{a}_t$ as the predicted values from a regression of a_t on $\{\Psi_j(t/T)\}_{j=1}^{q}$

- Equivalently

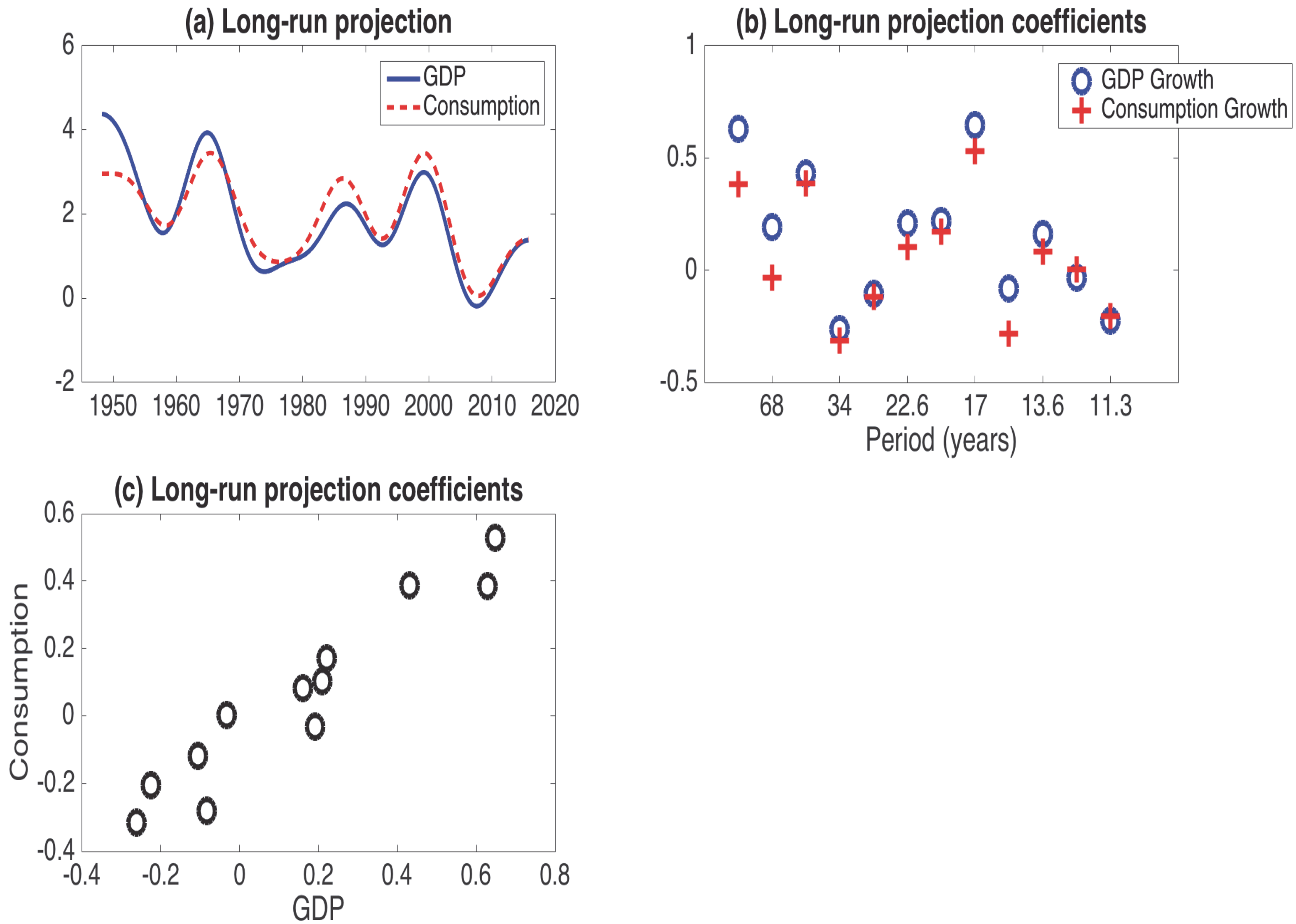
$$\hat{a}_t = \frac{1}{\sqrt{T}} \sum_{j=1}^{q} A_j \Psi_j(t/T) = \frac{\sqrt{2}}{\sqrt{T}} \sum_{j=1}^{q} A_j \cos(\pi j t/T)$$

because the weights $\Psi_j(t/T)$ are orthonormal

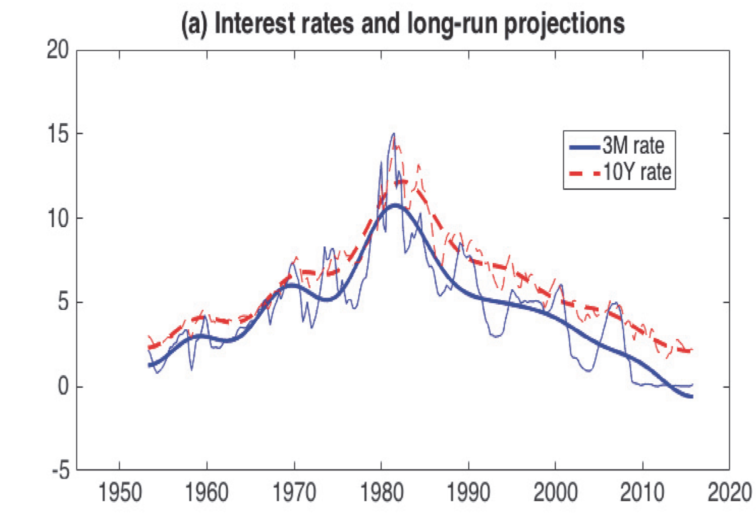
- $\hat{a}_t$ usefully thought of as low-pass filter for frequencies lower than $2T/q$ periods (66 years of data, $q = 12 \Rightarrow$ lower than 11 year cycles)

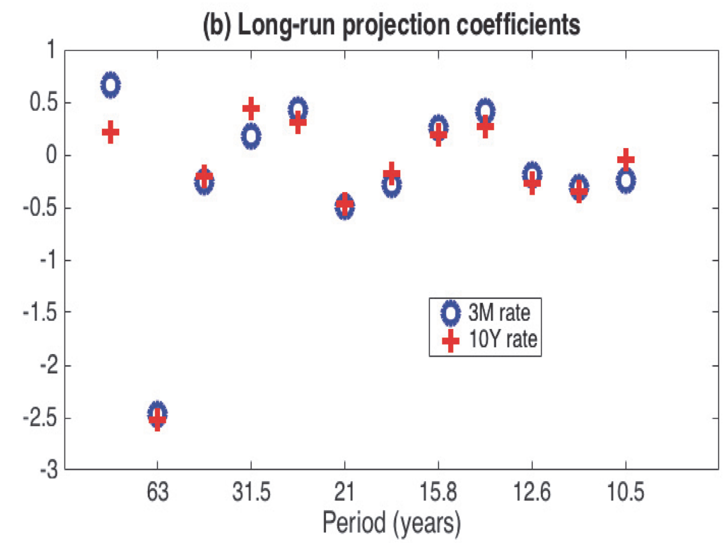
- $\hat{a}_t$ fully characterized by q regression coefficients A_j

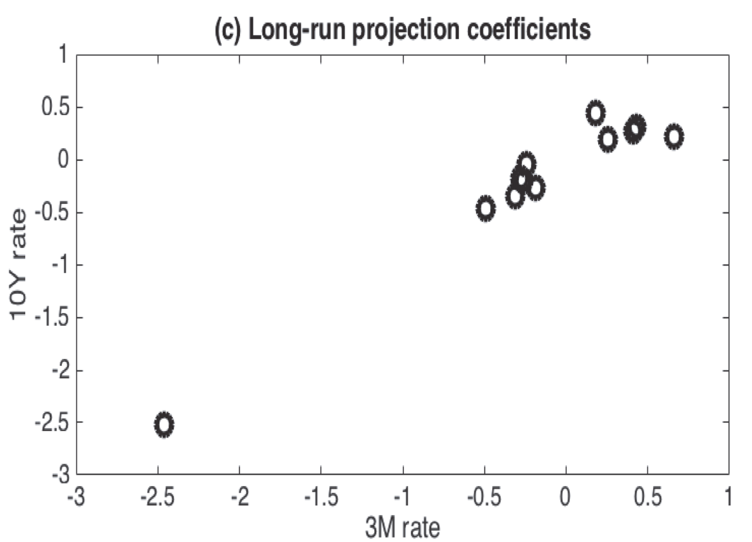
GDP and Consumption Growth



Short and Long-Term Interest Rates







Asymptotic Behavior of Cosine Transforms

- Let $(X_T', Y_T')'$ be $2q \times 1$ cosine transforms of $\{(x_t, y_t)\}_{t=1}^T$

- Parameters of interest and suggested confidence intervals are defined in terms of $(X_T', Y_T')'$

- Consider asymptotics where q is fixed, as in Müller (2004, 2007, 2014), Phillips (2006), Müller and Watson (2008, 2013, 2016a, 2016b), Sun (2013, 2016)

  $\Rightarrow$ Defines notion of "long-run": Periods of interest are $2T/q$ and longer

  $\Rightarrow$ Reflects that in samples of interest, reasonable q are small

  $\Rightarrow$ Ignoring data information beyond $(X_T', Y_T')'$ avoids modelling higher frequency properties

Asymptotic Behavior of Cosine Transforms

- Assume $(\Delta x_t, \Delta y_t)$ stationary with a spectral density f_T. Let $F_T(\lambda) = f_T(\lambda)/|1 - e^{\mathbf{i}\lambda}|^2$ be the pseudo-spectrum of (x_t, y_t), and suppose

$$F_T(\omega/T) \to S(\omega)$$

in a suitable sense.

- Under linear process assumption and additional regularity conditions, by CLT in Müller and Watson (2016)

$$\begin{pmatrix} X_T \\ Y_T \end{pmatrix} \Rightarrow \begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}(0, \boldsymbol{\Sigma}), \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{XX} & \boldsymbol{\Sigma}_{XY} \\ \boldsymbol{\Sigma}_{YX} & \boldsymbol{\Sigma}_{YY} \end{pmatrix}$$

and

$$\boldsymbol{\Sigma}_T = \mathsf{Var}\begin{pmatrix} X_T \\ Y_T \end{pmatrix} \to \boldsymbol{\Sigma}$$

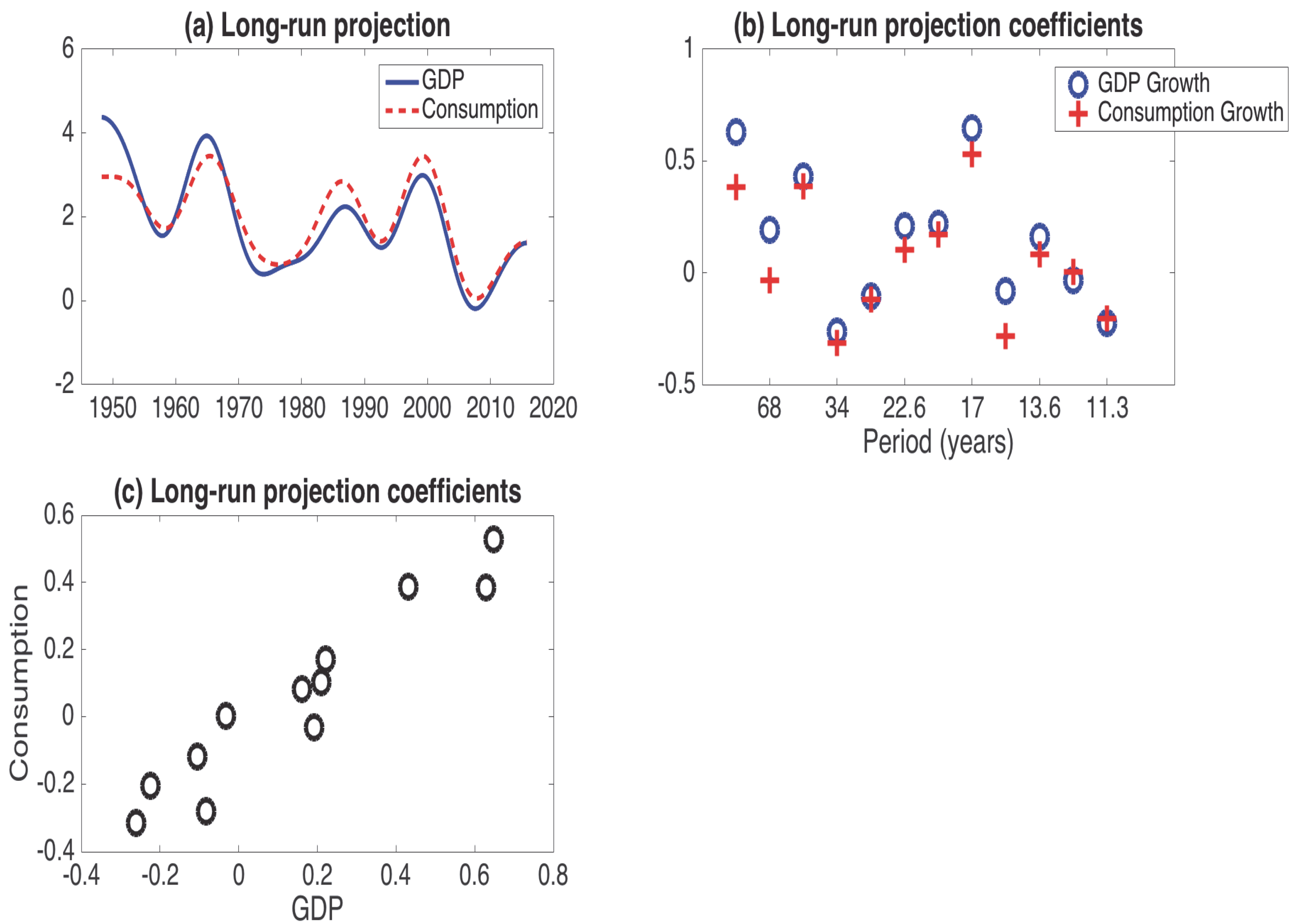
where $\boldsymbol{\Sigma}$ is a function of S.

Measuring Long-Run Covariability

$$\begin{aligned}\Omega_T &= T^{-1}\sum_{t=1}^{T} E\left[\begin{pmatrix}\hat{x}_t\\ \hat{y}_t\end{pmatrix}\begin{pmatrix}\hat{x}_t\\ \hat{y}_t\end{pmatrix}'\right] = \sum_{j=1}^{q} E\left[\begin{pmatrix}X_{T,j}\\ Y_{T,j}\end{pmatrix}\begin{pmatrix}X_{T,j}\\ Y_{T,j}\end{pmatrix}'\right]\\ &= \begin{pmatrix}\mathrm{tr}\,\Sigma_{XX,T} & \mathrm{tr}\,\Sigma_{XY,T}\\ \mathrm{tr}\,\Sigma_{YX,T} & \mathrm{tr}\,\Sigma_{YY,T}\end{pmatrix} \to \begin{pmatrix}\mathrm{tr}\,\Sigma_{XX} & \mathrm{tr}\,\Sigma_{XY}\\ \mathrm{tr}\,\Sigma_{YX} & \mathrm{tr}\,\Sigma_{YY}\end{pmatrix}\\ &= \begin{pmatrix}\Omega_{XX} & \Omega_{XY}\\ \Omega_{YX} & \Omega_{YY}\end{pmatrix} = \Omega\end{aligned}$$

- Scalar parameters derived from Ω
  - $\rho = \Omega_{XY}/\sqrt{\Omega_{XX}\Omega_{YY}}$
  - $\beta = \Omega_{XY}/\Omega_{XX}$
  - $\sigma = \sqrt{\Omega_{YY} - \Omega_{XY}^2/\Omega_{XX}}$

GDP and Consumption Growth



Interpretation of β and σ

- Define

$$\beta_T = \operatorname*{argmin}_b E\left[T^{-1}\sum_{t=1}^{T}(\hat{y}_t - b\hat{x}_t)^2\right] = \operatorname*{argmin}_b E\left[\sum_{j=1}^{q}(Y_{T,j} - bX_{T,j})^2\right]$$

$$\sigma_T = \sqrt{\min_b E\left[T^{-1}\sum_{t=1}^{T}(\hat{y}_t - b\hat{x}_t)^2\right]} = \sqrt{\min_b E\left[\sum_{j=1}^{q}(Y_{T,j} - bX_{T,j})^2\right]}$$

- Then $\beta_T \to \beta$ and $\sigma_T \to \sigma$, since

$$\begin{aligned} E\left[T^{-1}\sum_{t=1}^{T}(\hat{y}_t - b\hat{x}_t)^2\right] &= T^{-1}E\left[\sum_{t=1}^{T}\hat{y}_t^2 - 2b\sum_{t=1}^{T}\hat{x}_t\hat{y}_t + b^2\sum_{t=1}^{T}\hat{x}_t^2\right] \\ &= \Omega_{YY,T} - 2b\Omega_{XY,T} + b^2\Omega_{XX,T} \end{aligned}$$

- Population R^2 is $\Omega_{XY,T}^2/(\Omega_{XX,T}\Omega_{YY,T}) \to \rho^2$

Comparison with Standard Spectral Analysis

- Recall that F_T is pseudo-spectrum of (x_t, y_t), and $F_T(\omega/T) \to S(\omega)$.

- Up to reasonable approximation, Ω_T is to average over (pseudo) spectral density $F_T(\lambda)$ over frequencies $\lambda \in [\pi/T, q\pi/T]$, corresponding to average of $S(\omega)$ over $\omega \in [\pi, q\pi]$.

- Two departures from classic spectral analysis:

  1. Derive inference for q fixed, no LLN applicable

  2. Allow for curvature in F_T that is non-negligble in $1/T$ neighborhood (=don't assume S is flat).

Parametrizing S, and Implied Ω

Two important special cases:

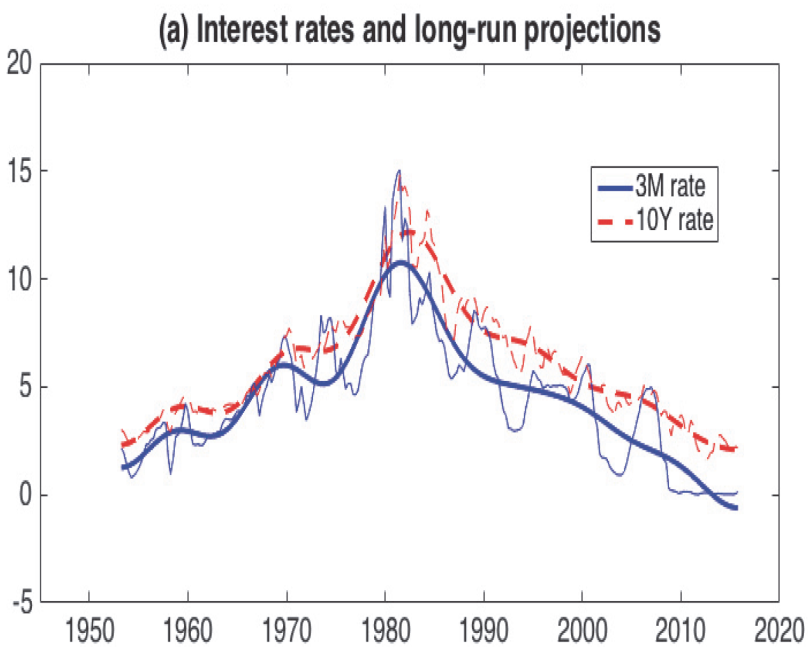
1. In I(0) model, $S(\omega) \propto \Lambda$, $(X_j, Y_j)' \sim iid\mathcal{N}(0, \Lambda)$

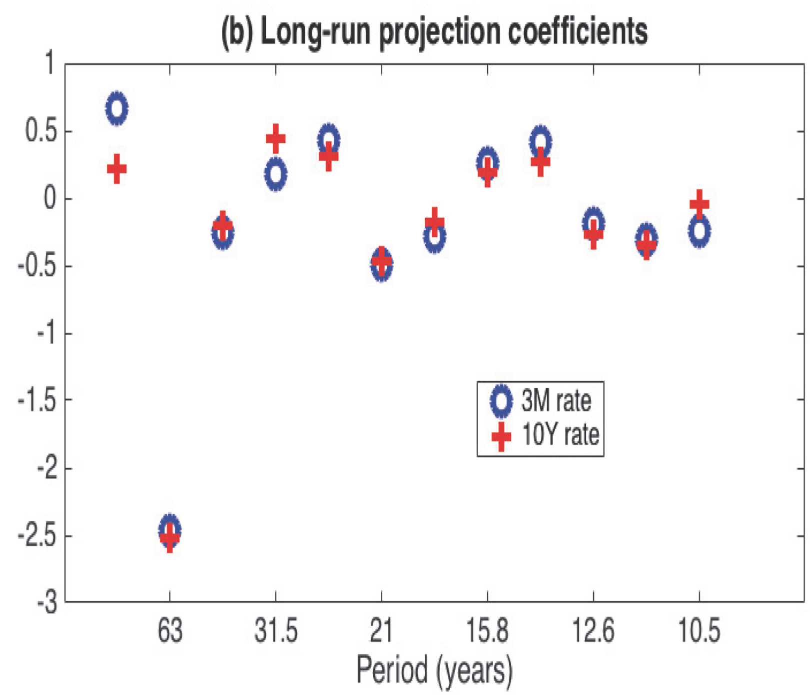
   $\Rightarrow \Omega \propto \Lambda$

2. In I(1) model, $S(\omega) \propto \Lambda/\omega^2$, $(X_j, Y_j)' \sim id\mathcal{N}(0, \Lambda/j^2)$

   $\Rightarrow \Omega \propto \Lambda$

Short and Long-Term Interest Rates





I(0) and I(1) Inference

⇒ Follows from well-known small sample results (cf. Anderson (1984))

		c and y		I-Rates	
		I(0)	I(1)	I(0)	I(1)
ρ	Estimates	0.93*	0.93*	0.97*	0.94*
	90% CI	[0.81;0.97]	[0.82;0.97]	[0.93;0.99]	[0.82;0.97]
β	Estimates	0.76*	0.84*	0.96*	0.85*
	90% CI	[0.60;0.92]	[0.67;1.01]	[0.84;1.08]	[0.68;1.03]
σ	Estimates	0.35	0.35	0.63	0.48
	90% CI	[0.26;0.55]	[0.26;0.54]	[0.47;0.97]	[0.36;0.74]

A Flexible Parametric Model for S

- (A, B, c, d) model:

$$S(\omega) = A \begin{pmatrix} (\omega^2 + c_1^2)^{-d_1} & 0 \\ 0 & (\omega^2 + c_2^2)^{-d_2} \end{pmatrix} A' + BB'$$

  with $d_i \in [-0.4, 1]$, $c_i \in [0, \infty]$ and B lower triangular

- 11 parameter model that embeds bivariate fractional, local-to-unity and local-level models as special cases

- Allows for various long-run phenomena such as (stochastic) breaks in means, slow mean reversion, overdifferencing, etc.

- Ω can be computed from implied $\Sigma = \Sigma(\theta)$, $\theta = (A, B, c, d)$

Construction of Confidence Intervals

- Under asymptotic approximation, a parametric small sample problem:
  - Observe $Z = (X', Y')' \sim \mathcal{N}(0, \mathbf{\Sigma})$, where $\mathbf{\Sigma} = \mathbf{\Sigma}(\theta)$, $\theta = (A, B, c, d) \in \mathbf{\Theta}$
  - Seek confidence interval $H(Z)$ for parameter of interest $\gamma = g(\theta)$, for known g
  - Impose appropriate invariance on H

- W-weighted average expected length minimizing program

$$\min_H \int E_\theta[\text{lgth}(H(Z))]dW(\theta) \text{ s.t. } P_\theta(g(\theta) \in H(Z)) \geq 1 - \alpha \; \forall \theta \in \mathbf{\Theta}$$

$\Rightarrow$ Optimal H depends on unknown Lagrange multipliers

Form of Optimal Confidence Set

- For simplicity, ignore invariance. Then (cf. Pratt 1961), optimal H inverts tests φ_{γ_0} of $H_0 : g(\theta) = \gamma_0$ of the form

$$\varphi_{\gamma_0}(z) = \mathbf{1}[\int f_\theta(z)dW(\theta) > \mathsf{cv} \int f_\theta(z)d\Lambda_{\gamma_0}(\theta)]$$

  for some probability distribution Λ_{γ_0} on Θ with support on a subset of $\{\theta : g(\theta) = \gamma_0\}$, $\Lambda_{\gamma_0}(\{\theta : P_\theta(g(\theta) \in H(Z)) > 1 - \alpha\}) = 0$ and $E_\theta[\varphi_{\gamma_0}(Z)] \leq \alpha$ for all $g(\theta) = \gamma_0$.

- Similar form under invariance, but not a family of distributions Λ if parameter of interest is affected by invariance

  $\Rightarrow$ see Müller and Norets (2016) for details. Effective dimension of parameter space after invariance is 8 for all three problems.

Computation of Λ

- Arbitrary Λ induces lower bound on expected length criterion that holds for all valid CIs (cf. Elliott, Müller and Watson (2015))

- Basic algorithm

  1. Let $\Theta_c = \{\theta_1, \ldots, \theta_m\}$ be candidate set for support of Λ.

  2. Compute Λ_c weights and cv such that size of test is $\alpha - \epsilon$ on Θ_c.

  3. Check size control on Θ:

     (a) If size is controlled, we are done (compare length to bound generated from cv′ that induces Λ-weighted size equal to α).

     (b) If size violated, add violating θ to Θ_c and go to Step 2.

Computational Details

- Λs have about 30-100 points of support
- CS are within 5% of lower bound on length criterion for $\alpha = 0.1$
- Size control: 500 consecutive BFGS searches with random starting values don't find violation
- Single problem takes about 30 minutes on fast PC using Fortran

Choice of Weighting Function

- Recall

$$S(\omega) = A \begin{pmatrix} (\omega^2 + c_1^2)^{-d_1} & 0 \\ 0 & (\omega^2 + c_2^2)^{-d_2} \end{pmatrix} A' + BB'$$

- Set $B = 0$, $c_1 = c_2 = 0$, $d_1, d_2 \sim iidU[-0.4, 1.4]$,

$$A \sim R(\pi U_1)\,\text{diag}(15^{U_0}, 1) R\,(\pi U_2)$$

where $U_j \sim iidU[0, 1]$ and $R(\phi)$ is 2×2 rotation matrix of angle ϕ.

Credibility of Resulting CS

- Optimal $H(z)$ could be empty for some z, or unreasonably short

  $\Rightarrow$ Perversely, this is optimal for very uninformative draws Z, as coverage then costly in terms of length

- Generic problem in nonstandard problems: Frequentist (optimality) properties don't rule out unreasonable descriptions of uncertainty for some realizations

  $\Rightarrow$ Analyzed in detail in Müller and Norets (2016), building on old literature (Fisher (1956), Buehler (1959), Wallace (1959), Cornfield (1969), Pierce (1973), Robinson (1977), etc.)

- Suggested solution: H constrained to be superset of equal-tailed credible set relative to prior W

Spurious Regression?

- Minimal coverage of weighted average length minimizing nominal 90% CIs for ρ under $\rho = 0$

DGP\Assumed Model	I(0)	I(1)	$ABcd$
I(0)	0.90	0.01	0.91
I(1)	0.01	0.90	0.90
$ABcd$	0.01	0.00	0.90

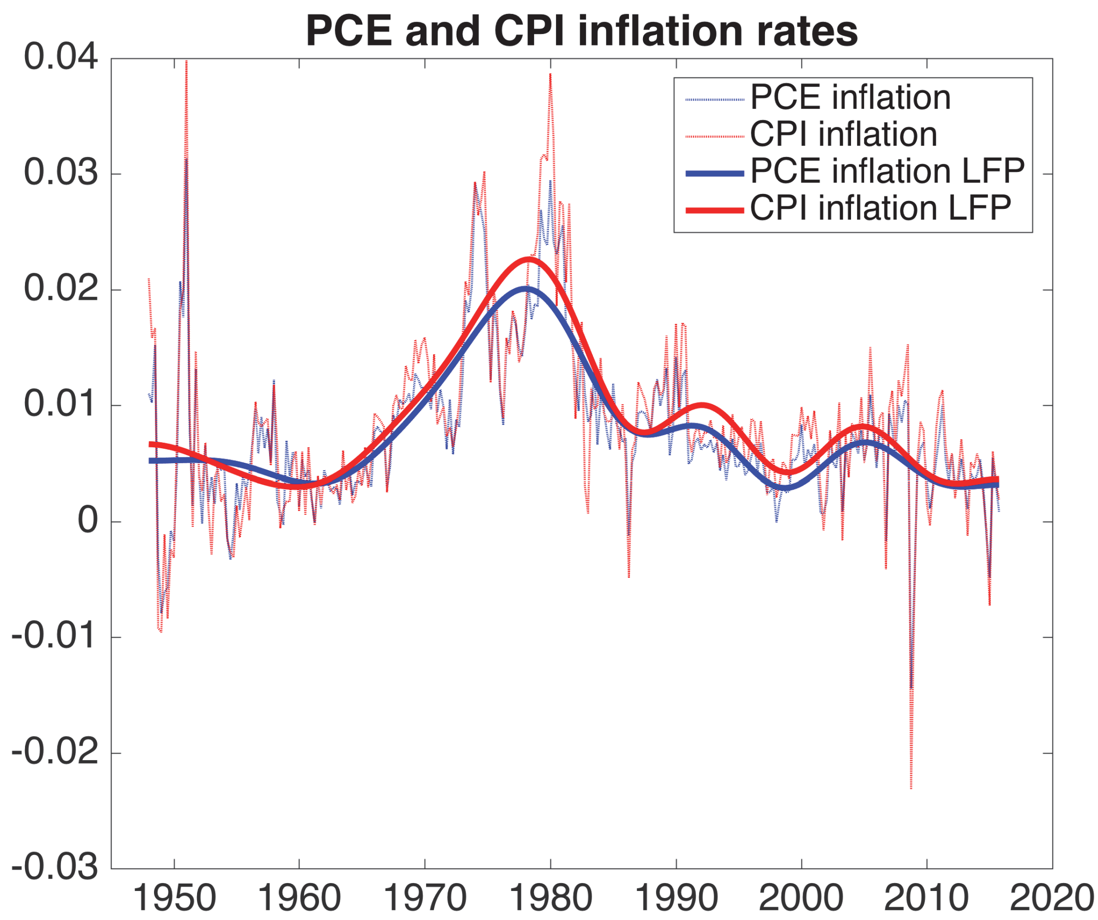
Results in Running Examples

		I(0)	I(1)	$ABcd$
			c and y	
ρ	Estimates	0.93*	0.93*	0.91*
	90% CI	[0.81;0.97]	[0.82;0.97]	[0.71;0.97]
β	Estimates	0.76*	0.84*	0.76*
	90% CI	[0.60;0.92]	[0.67;1.01]	[0.48;0.95]
σ	Estimates	0.35	0.35	0.40
	90% CI	[0.26;0.55]	[0.26;0.54]	[0.27;0.66]
			I-Rates	
ρ	Estimates	0.97*	0.94*	0.96*
	90% CI	[0.93;0.99]	[0.82;0.97]	[0.89;0.99]
β	Estimates	0.96*	0.85*	0.92*
	90% CI	[0.84;1.08]	[0.68;1.03]	[0.75;1.14]
σ	Estimates	0.63	0.48	0.70
	90% CI	[0.47;0.97]	[0.36;0.74]	[0.53;0.92]

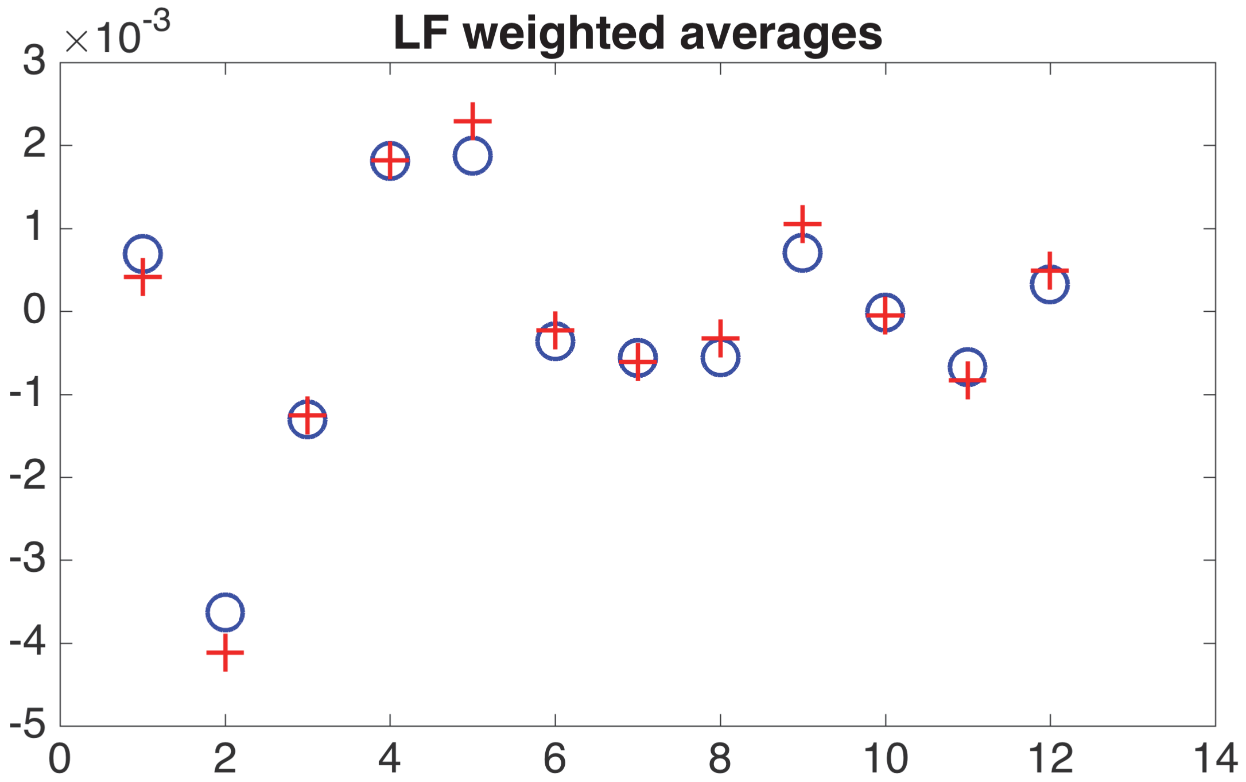
Real Variables: Correlation

		c	i	$w \times n$	TFP
y	$\hat{\rho}$	0.91*	0.53*	0.98*	0.78*
	90% CI	[0.71; 0.97]	[0.02; 0.81]	[0.95; 0.99]	[0.45; 0.95]
c	$\hat{\rho}$		0.53*	0.92*	0.70*
	90% CI		[0.03; 0.81]	[0.68; 0.97]	[0.28; 0.91]
i	$\hat{\rho}$			0.51*	0.38
	90% CI			[0.02; 0.80]	[−0.08; 0.71]
$w \times n$	$\hat{\rho}$				0.72*
	90% CI				[0.38; 0.93]

PCE and CPI Inflation

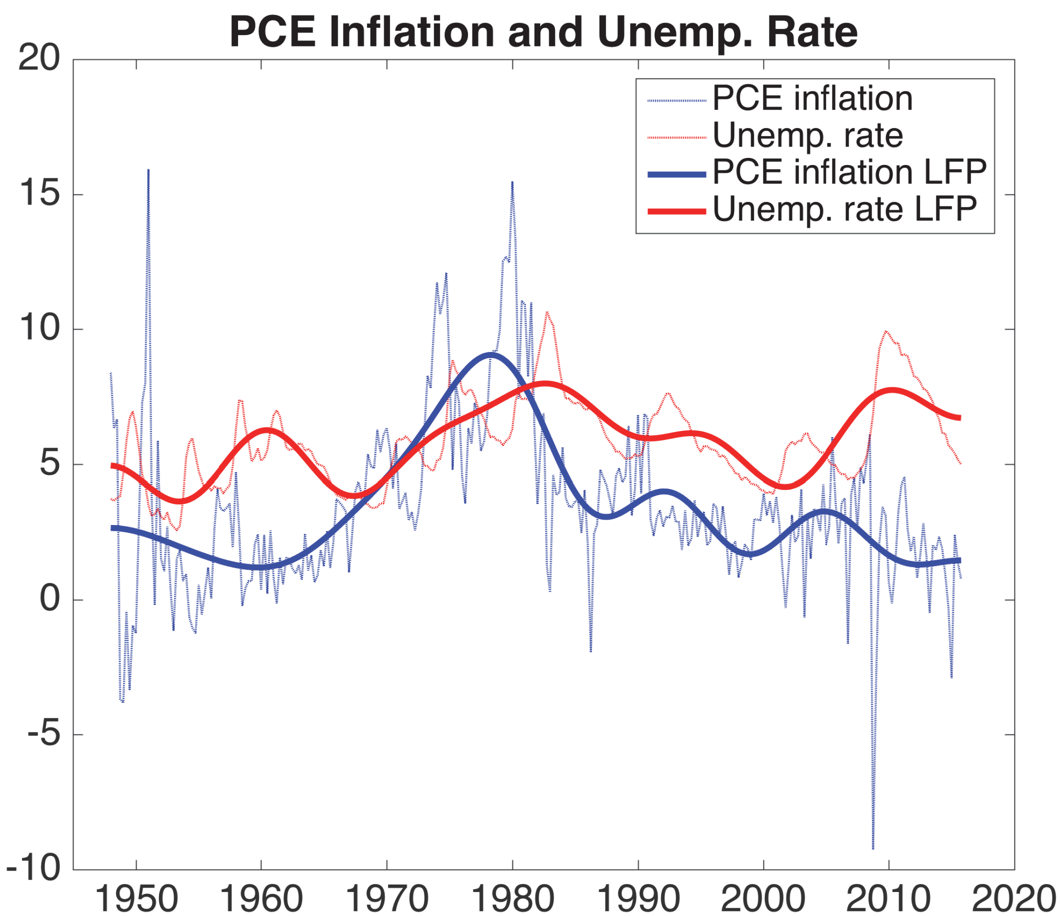


PCE and CPI Inflation

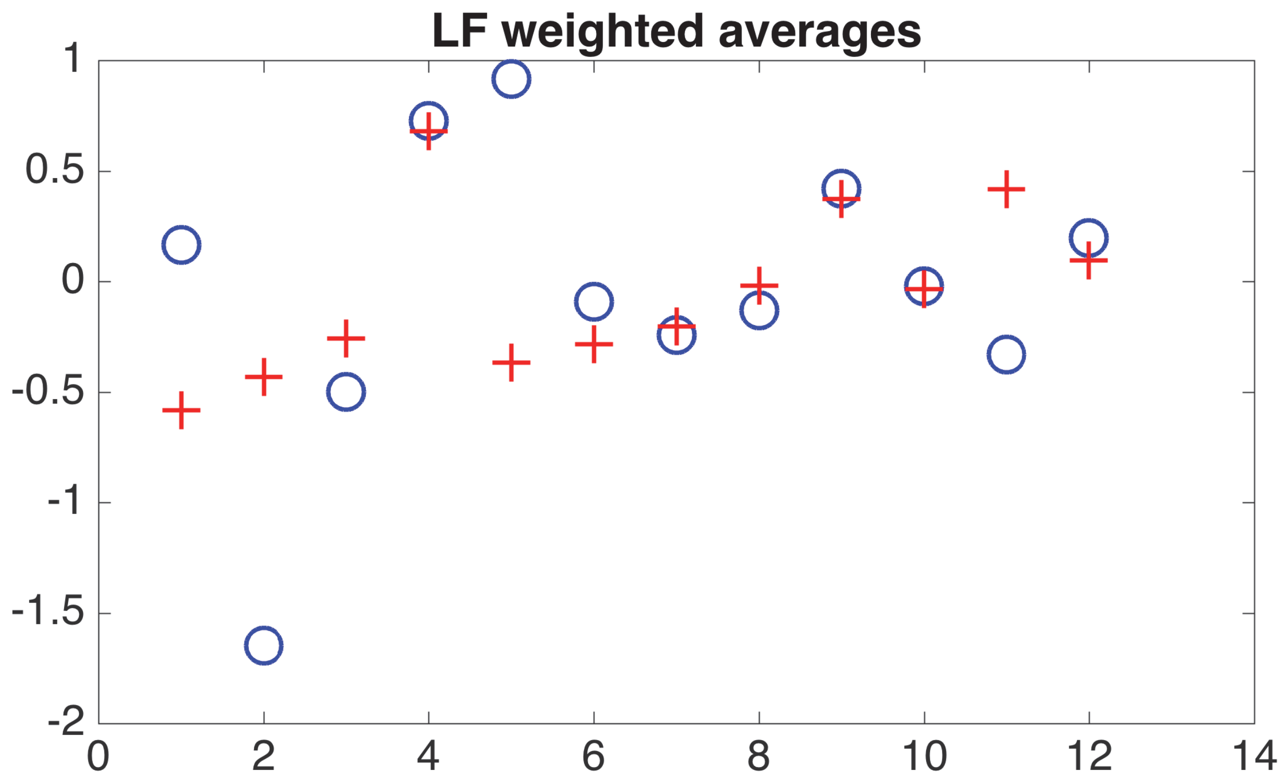


	ρ	β
Estimate	0.98*	1.13*
90% CI	[0.95;0.99]	[0.98;1.24]

PCE Inflation and Unemployment

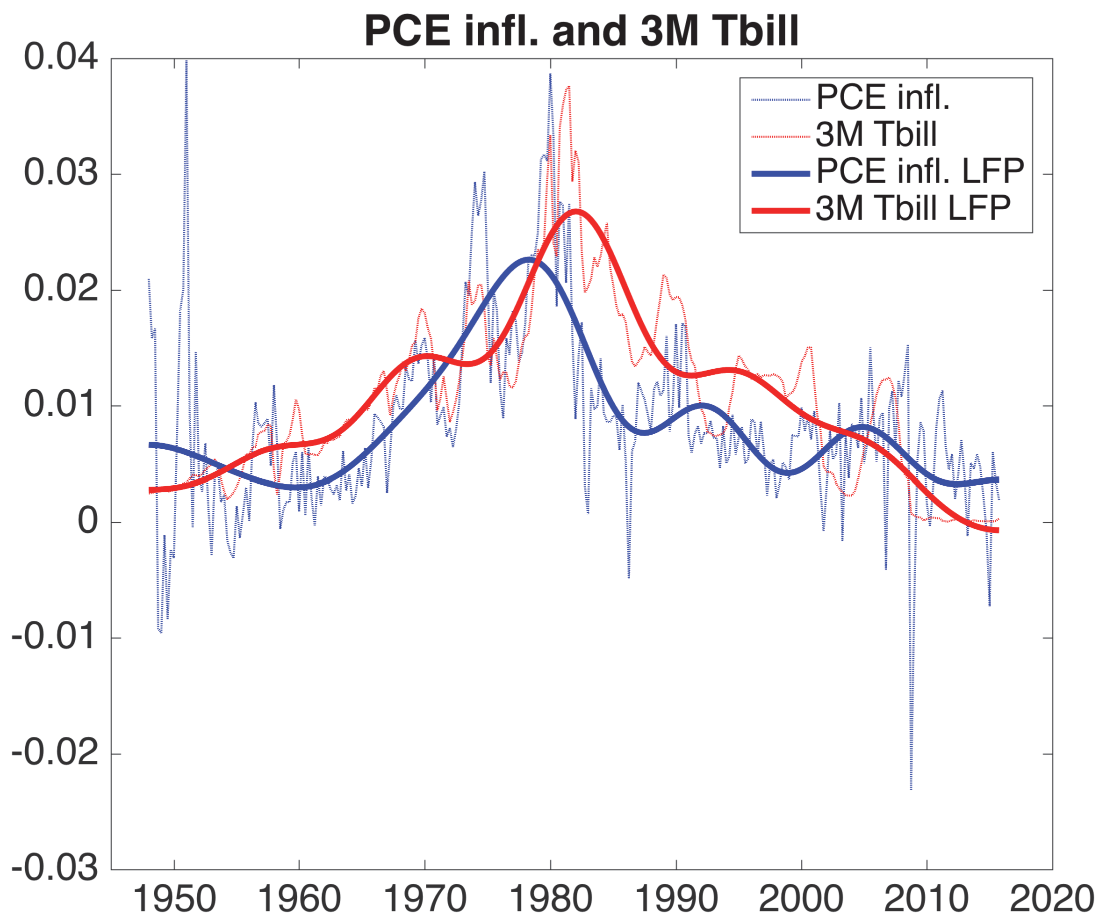


PCE Inflation and Unemployment

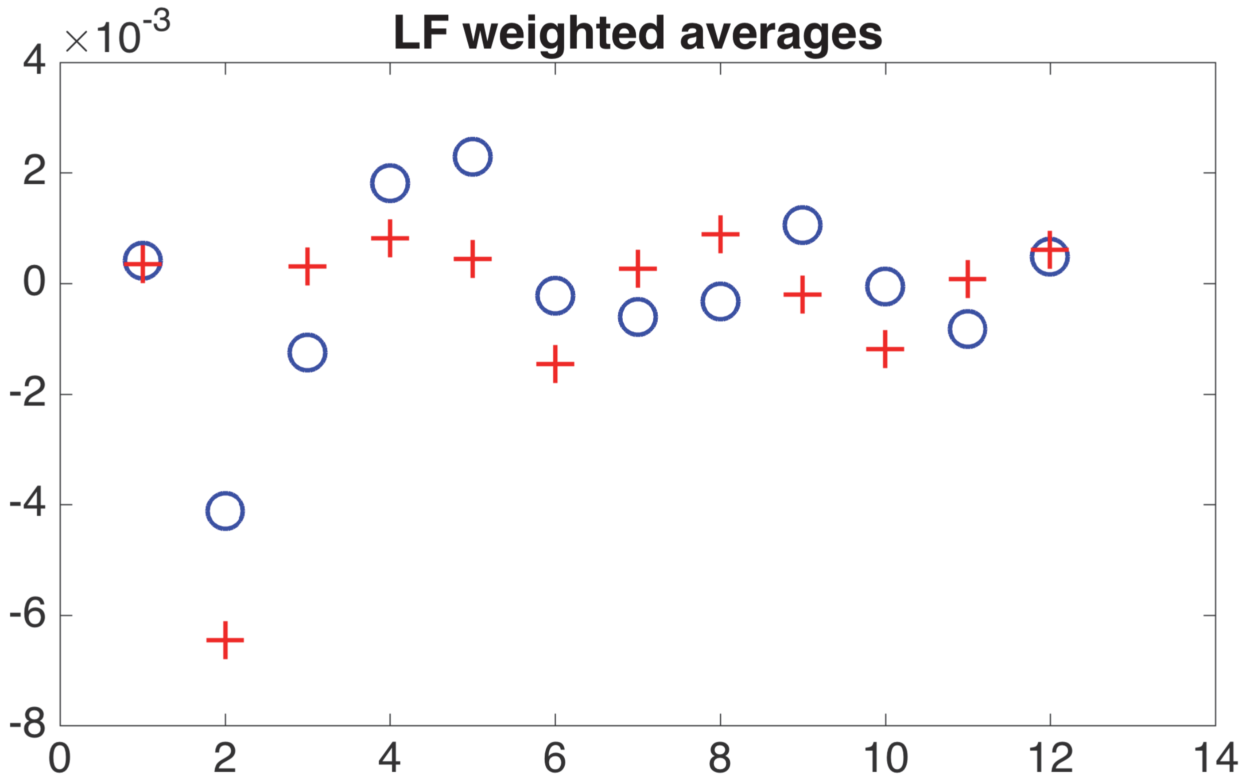


	ρ	β
Estimate	0.25	0.21
90% CI	[-0.27;0.82]	[-0.24;0.78]

PCE Inflation and Short I-Rates

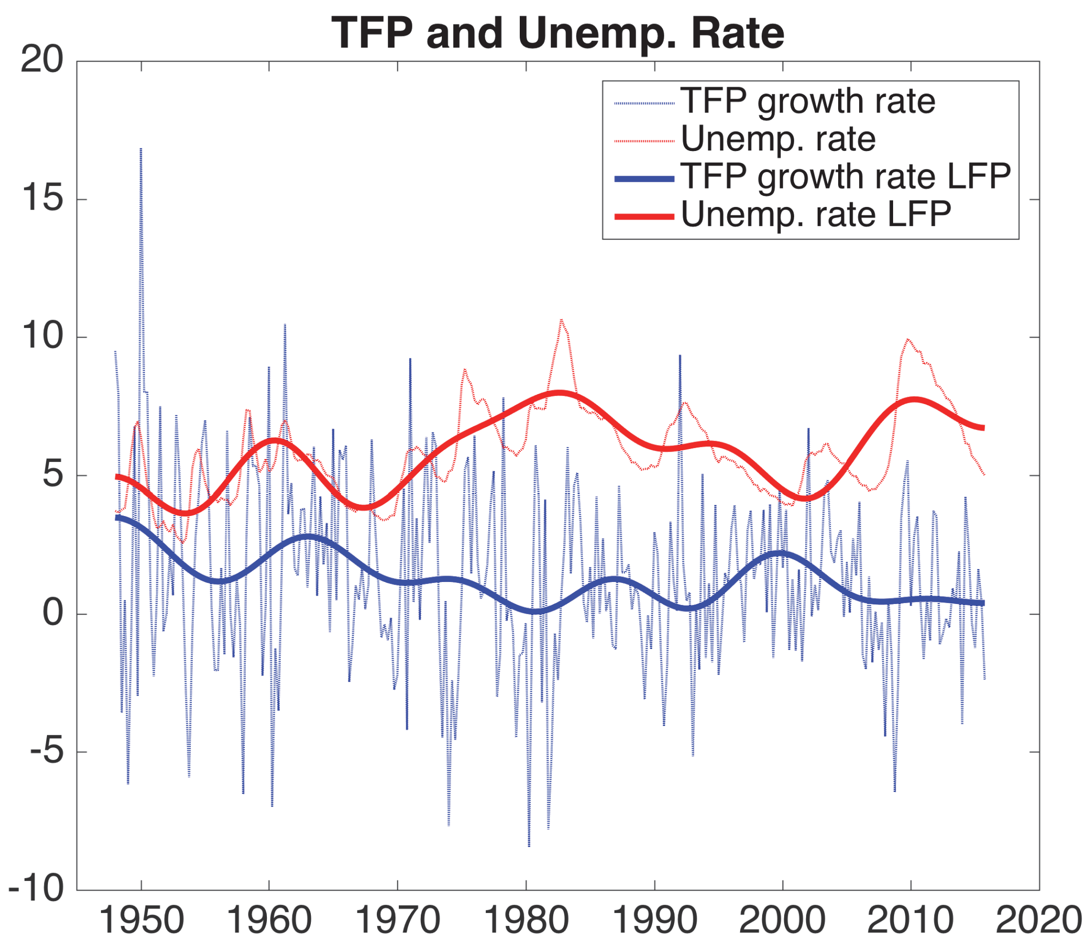


PCE Inflation and Short I-Rates

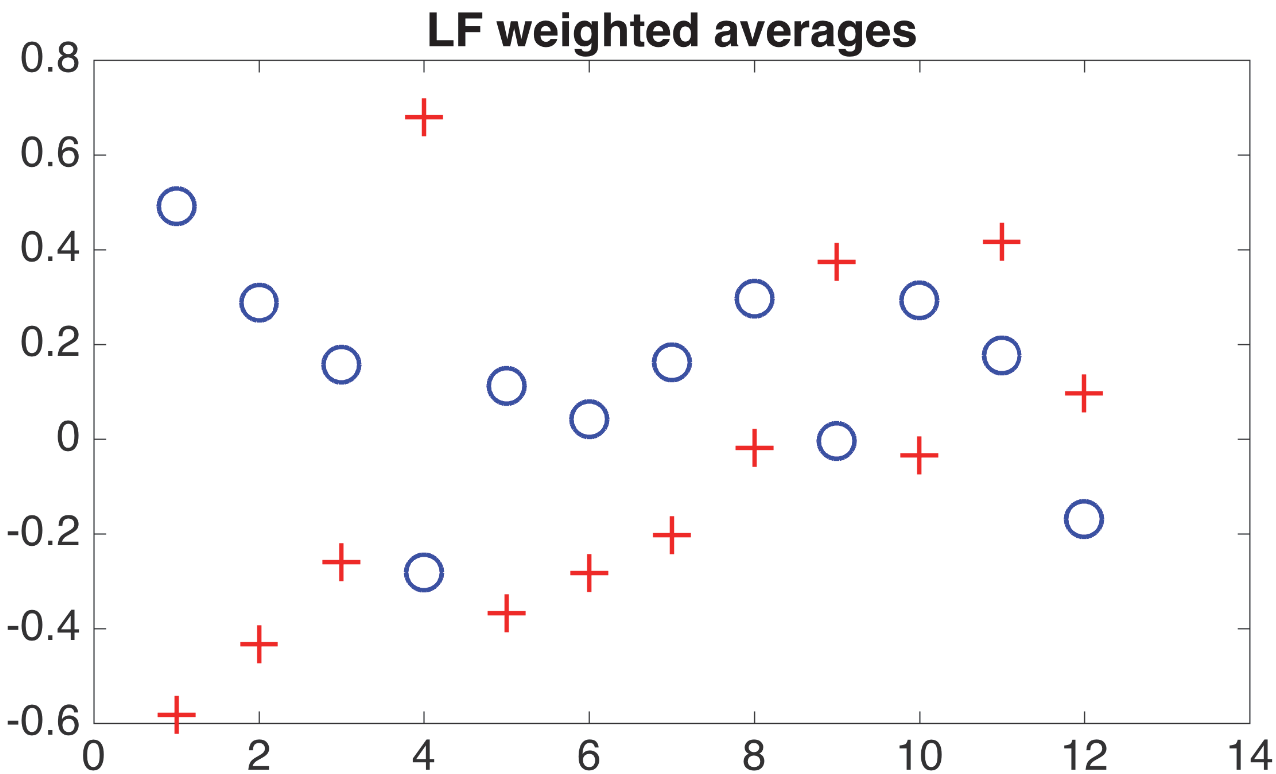


	ρ	β
Estimate	0.47	0.73
90% CI	[0.00;0.91]	[-0.09;1.91]

TFP and Unemployment



TFP and Unemployment



	ρ	β
Estimate	-0.65*	-1.00*
90% CI	[-0.75;-0.35]	[-1.64;-0.27]

Conclusions

- Study statistical significance of long-run relationships that is robust to many DGPs

- Underlying derivations are involved, but not difficult or computationally intensive to apply to data

- In paper: Sensitivity of results to alternative q and frequencies of interest