HAC Corrections for Strongly Autocorrelated Time Series

Ulrich Müller
Princeton University

March 2013

Introduction

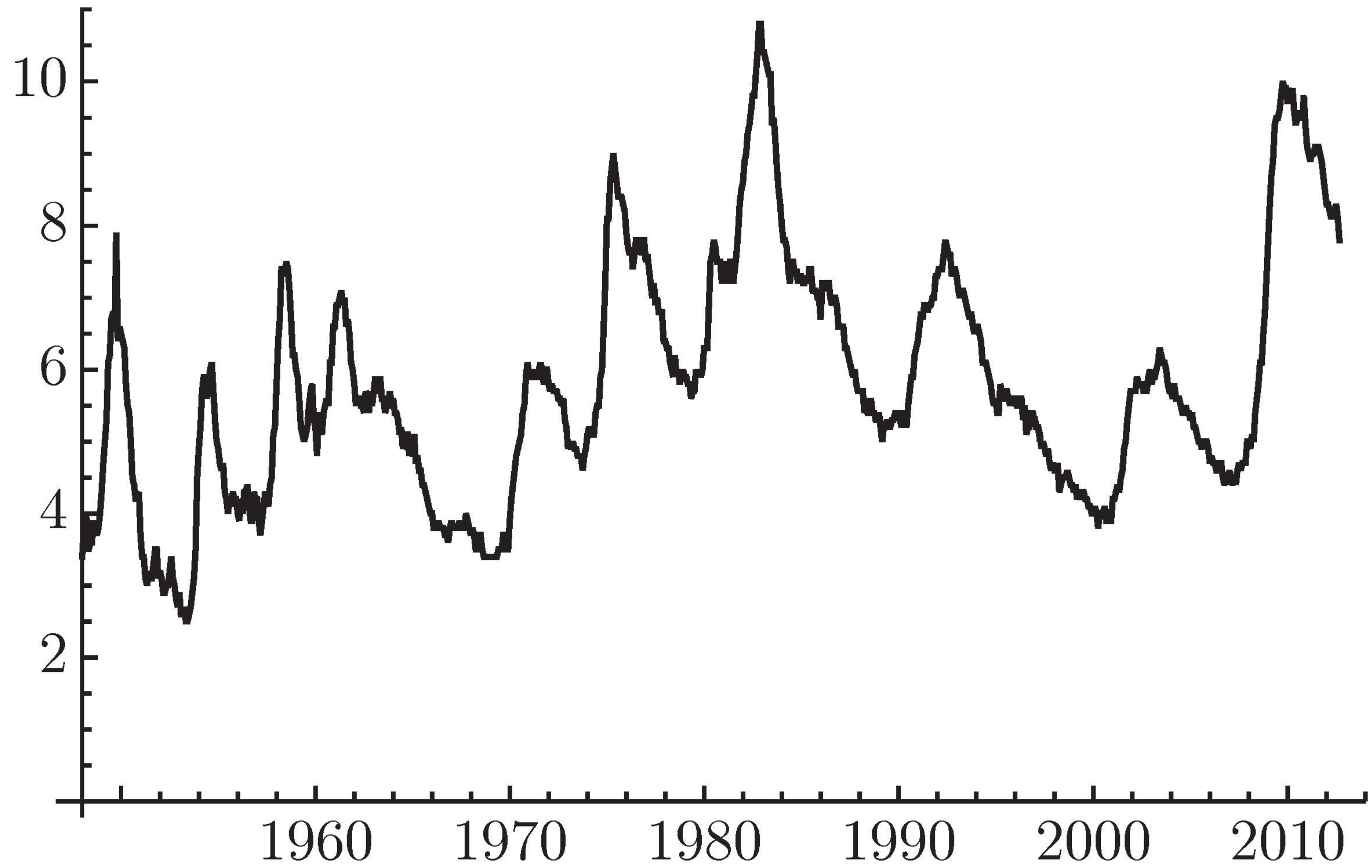
- Renewed interest in HAC corrections in time series:

  ⇒ Kiefer and Vogelsang (2002, 2005), Jansson (2004), Müller (2004, 2007), Sun, Phillips and Jin (2008), Phillips, Sun and Jin (2006, 2007), Gonçalves and Vogelsang (2011), Sun and Kaplan (2012), Achadé and Cattaneo (2012)

- Basic idea: account for sampling variability of HAC or "long-run variance" estimator

- Improves size control, but strong autocorrelations still induce severe over-rejections

Example: Inference about Population Mean of US Unemployment Rate



This Paper: Two Goals

1. Review of LRV estimation and inference

   - Spectral perspective
   - Both traditional "consistent" and KV-type "inconsistent" estimators
   - Illustrates causes for break-down under strong autocorrelation

This Paper

2. Derivation of valid inference under specific type of strong dependence

   - Assumption that long-run properties are approximated by stationary Gaussian AR(1)

   - AR(1) coefficient may be arbitrarily close to one

     $\Rightarrow$ allows for specific form of strong dependence

   - Numerical determination of approximately weighted average power maximizing test, following Elliott, Müller and Watson (2012)

   - Encouraging Monte Carlo results

Remainder of Talk

1. Review of time series HAC inference

   (a) based on consistent LRV estimators

   (b) based on inconsistent LRV estimators

2. New test for AR(1) persistence

3. Generalization to regression and GMM inference problems

4. Monte Carlo results

5. Conclusion

Review of LRV Estimation and Inference

- Consider inference about population mean μ of scalar time series y_t.

  Assume second-order stationarity, with absolutely summable autocovariances $\gamma(j) = E[(y_t - \mu)(y_{t-j} - \mu)]$.

- LRV is defined as

$$\omega^2 = \lim_{T\to\infty} \text{Var}[T^{1/2}\hat{\mu}] = \sum_{j=-\infty}^{\infty} \gamma(j)$$

  where $\hat{\mu} = T^{-1}\sum_{t=1}^{T} y_t$.

- By CLT, $\sqrt{T}(\hat{\mu} - \mu) \Rightarrow \mathcal{N}(0, \omega^2)$. With consistent estimator $\hat{\omega}^2 \xrightarrow{p} \omega^2$,

$$\sqrt{T}\frac{\hat{\mu} - \mu}{\hat{\omega}} \Rightarrow \mathcal{N}(0, 1).$$

Spectral Perspective and DFT

- Spectral density $f : [-\pi, \pi] \mapsto [0, \infty)$ defined as

$$f(\lambda) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} e^{\mathbf{i}\lambda j} \gamma(j) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \cos(j\lambda)\gamma(j).$$

- Thus, $\omega^2 = 2\pi f(0)$.

- Assume T odd. DFT transforms $\{y_t\}_{t=1}^{T}$ into $\hat{\mu}$ and the $T-1$ trigonometrically weighted averages $\{Z_l^{\cos}\}_{l=1}^{\lfloor T/2 \rfloor}$ and $\{Z_l^{\sin}\}_{l=1}^{\lfloor T/2 \rfloor}$, where

$$\begin{aligned} Z_l^{\cos} &= T^{-1/2}\sqrt{2} \sum_{t=1}^{T} \cos(2\pi l(t-1)/T) y_t \\ Z_l^{\sin} &= T^{-1/2}\sqrt{2} \sum_{t=1}^{T} \sin(2\pi l(t-1)/T) y_t. \end{aligned}$$

Properties of DFT

- Theorem: All pairwise correlations between the T random variables $T^{1/2}\hat{\mu}$, $\{Z_l^{\mathrm{cos}}\}_{l=1}^{\lfloor T/2 \rfloor}$ and $\{Z_l^{\mathrm{sin}}\}_{l=1}^{\lfloor T/2 \rfloor}$ converge to zero as $T \to \infty$, and

$$
\begin{aligned}
\sup_{l \le \lfloor T/2 \rfloor} \left| E[(Z_l^{\mathrm{cos}})^2] - 2\pi f(2\pi l/T) \right| &\to 0, \\
\sup_{l \le \lfloor T/2 \rfloor} \left| E[(Z_l^{\mathrm{sin}})^2] - 2\pi f(2\pi l/T) \right| &\to 0.
\end{aligned}
$$

- DFT converts autocorrelation of y_t into heteroskedasticity of $(Z_l^{\mathrm{sin}}, Z_l^{\mathrm{cos}})$, with shape governed by the spectral density f.

- Truly remarkable result. Possible, because covariance matrix of y_t is Toeplitz, and by absolute summability of $\gamma(j)$, almost all of the 'action' is close to the diagonal.

Periodogram

- Under approximation $E[(Z_l^{\text{cos}})^2] \approx E[(Z_l^{\text{sin}})^2] \approx 2\pi f(2\pi l/T)$, the lth periodogram ordinate
$$p_l = \frac{(Z_l^{\text{cos}})^2 + (Z_l^{\text{sin}})^2}{2}$$
is an estimator of the spectral density at $2\pi l/T$, scaled by 2π.

- Under the assumption that f is flat over $[0, 2\pi n/T]$, suggests estimating LRV $\omega^2 = 2\pi f(0)$ via
$$\hat{\omega}^2_{p,n} = n^{-1} \sum_{l=1}^{n} p_l$$

- If f is continuous at 0, can choose sequence $n = n_T \to \infty$ with $n_T/T \to 0$ to obtain consistent estimator $\hat{\omega}^2_{p,n_T}$.

Kernel Estimators

- Many popular estimators are of the form

$$\hat{\omega}^2_{k,S_T} = \sum_{j=-T+1}^{T-1} k(j/S_T)\hat{\gamma}(j)$$

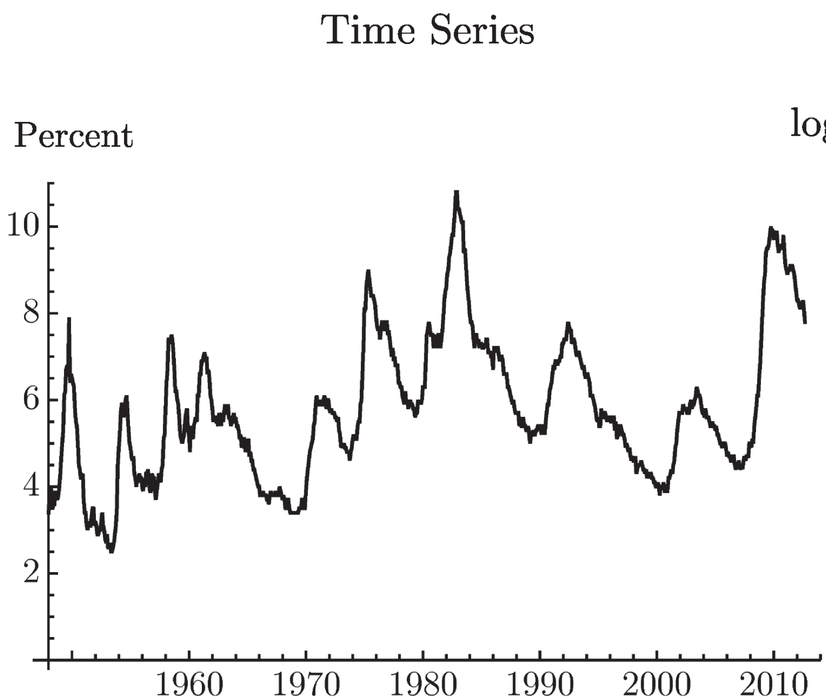
where k is even kernel with $k(0) = 1$, the bandwidth S_T satisfies $S_T \to \infty$, and $\hat{\gamma}(j) = T^{-1} \sum_{t=|j|+1}^{T} (y_t - \hat{\mu})(y_{t-|j|} - \hat{\mu})$.

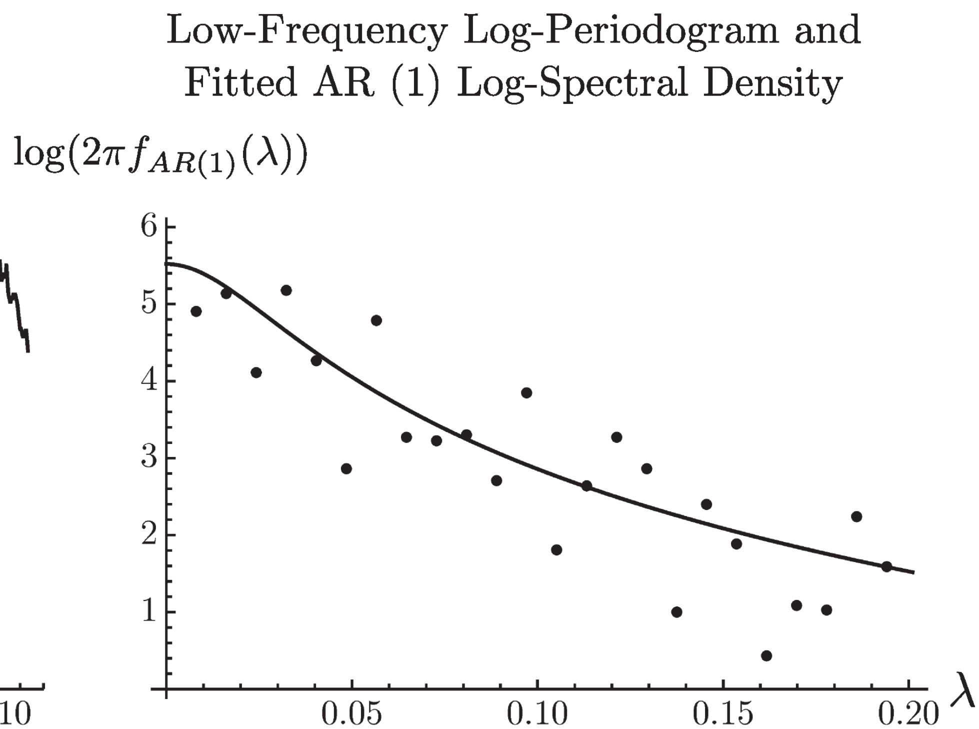
- For instance, Bartlett kernel $k(x) = \max(1 - |x|, 0)$ yields popular Newey and West (1987) estimator.

- Approximately

$$\hat{\omega}^2_{k,S_T} \approx \sum_{l=1}^{\lfloor T/2 \rfloor} K_{T,l} p_l, \text{ where } K_{T,l} = \frac{2}{T} \sum_{j=-T+1}^{T-1} \cos(2\pi j l/T) k(j/S_T).$$

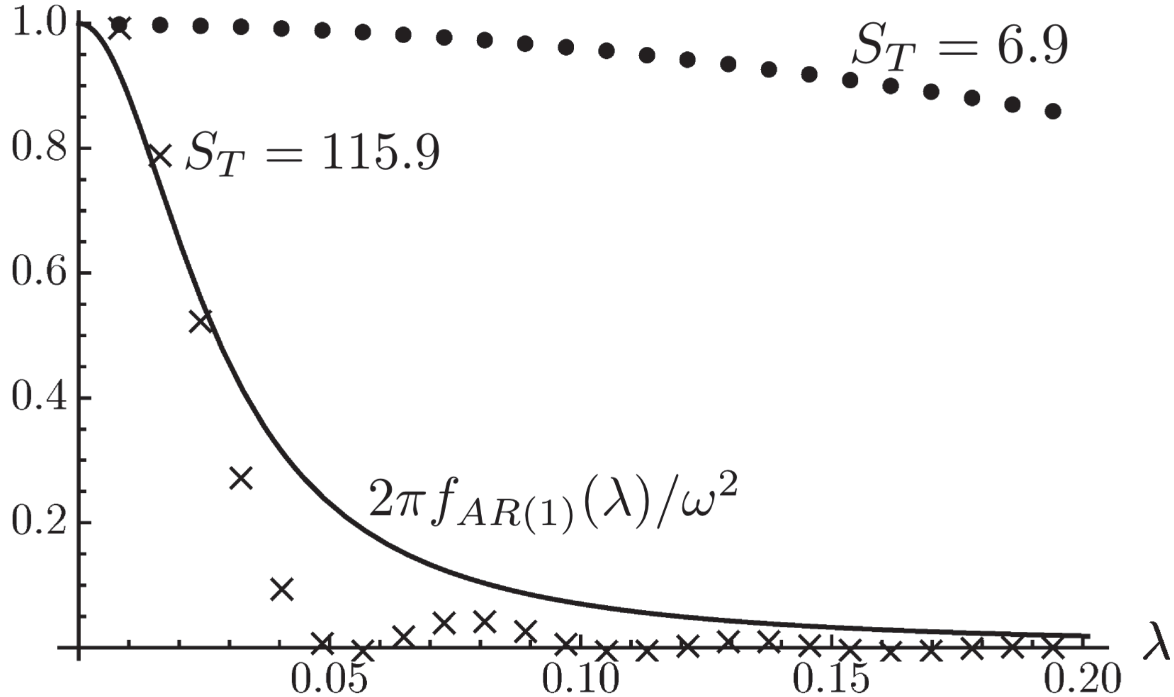
US Unemployment Example





fitted AR(1) has coefficient $\hat{\rho} = 0.973$ for $T = 777$ monthly observations

Implied Weights $K_{T,l}$ of Newey-West Estimator



$S_T = 6.9$: Default suggested in Stock and Watson's textbook for moderate autocorrelation

$S_T = 115.9$: Andrews' (1991) suggestion for AR(1) with coefficient of 0.973

$\Rightarrow$ no good choice for S_T: LRV estimator is very biased, or very variable

Parametric LRV Estimators

- Model y_t as ARMA with parameter θ and spectrum $f_{ARMA}(\lambda;\theta)$ and estimate ω^2 from implied spectrum $\hat{\omega}^2_{ARMA} = 2\pi f_{ARMA}(0;\hat{\theta})$.

- Approximate uncorrelatedness of DFT implies Whittle approximation to Gaussian (quasi) log-likelihood

$$-\sum_{l=1}^{\lfloor T/2 \rfloor} \log(2\pi f_{ARMA}(2\pi l/T;\theta)) - \sum_{l=1}^{\lfloor T/2 \rfloor} \frac{p_l}{2\pi f_{ARMA}(2\pi l/T;\theta)}$$

$\Rightarrow \hat{\theta}$ (and thus $\hat{\omega}^2_{ARMA}$) is determined by all frequencies, as encoded by p_l, $l = 1, \cdots, \lfloor T/2 \rfloor$

- For AR(1), $\omega^2 = \sigma^2/(1-\rho)^2$. Very sensitive to estimation error in ρ. For instance, with $\rho = 0.973$ and $T = 777$, one standard deviation estimation error in ρ leads to an estimation error of ω^2 by a factor of 0.6 and 2.0, respectively. Same problem for prewhitening.

Inconsistent LRV Estimators

- For strongly autocorrelated series, reasonable bandwidth choices lead to highly variable $\hat{\omega}^2$

  $\Rightarrow$ needs to be accounted for in distributional approximation for $\sqrt{T}(\hat{\mu} - \mu)/\hat{\omega}$

- By appropriate CLT and previous DFT result,

$$\begin{pmatrix} Z_l^{\text{sin}} \\ Z_l^{\text{cos}} \end{pmatrix} \overset{a}{\sim} i.n.i.d.\mathcal{N}\left(0, 2\pi f(2\pi l/T) I_2\right).$$

- Under the assumption that f is flat over $[0, 2\pi n/T]$, continue to rely on $\hat{\omega}_{p,n}^2 = n^{-1}\sum_{l=1}^{n} p_l = (2n)^{-1}\sum_{l=1}^{n}((Z_l^{\text{cos}})^2 + (Z_l^{\text{sin}})^2)$, but now

$$\sqrt{T}\frac{\hat{\mu} - \mu}{\hat{\omega}_{p,n}} \overset{a}{\sim} t_{2n}$$

Alternative: Discrete Cosine Transform

- Müller (2004, 2007) and Müller and Watson (2008) work with

$$Y_l = T^{-1/2}\sqrt{2}\sum_{t=1}^{T} \cos(\pi l(t-1/2)/T) y_t, \quad l \geq 1$$

- DCT behaves very similarly to DFT, so that under weak dependence, Y_l^2 are approximately independent estimators of $2\pi f(\pi l/T)$.

- If f is assumed flat over $[0, 2\pi q/T]$, use $\hat{\omega}_{Y,q}^2 = \frac{1}{q}\sum_{l=1}^{q} Y_l^2$ and

$$\sqrt{T}\frac{\hat{\mu}-\mu}{\hat{\omega}_{Y,q}} \overset{a}{\sim} t_q.$$

- Similar ideas pursued in Phillips (2005) and Sun (2012).

Choice of q

- Might make sense in macroeconomic applications to assume that spectrum is flat below business cycle frequencies

  $\Rightarrow$ roughly says that business cycles are i.i.d.

- Define business cycles to be at most 8 year cycles

  $\Rightarrow$ choose q equal to the largest integer smaller than the span of the data in years divided by 4

  $\Rightarrow$ in unemployment example, with $T = 777$ monthly observations, $(777/12)/4 \approx 16.2$, so $q = 16$.

Data-Dependent Choice of q?

- If $y_t \sim iid(\mu, \sigma^2)$, then asymptotically more powerful inference is obtained by relying on OLS variance estimator (or any other consistent HAC estimator).

- Suppose we know for sure that

$$\sqrt{T}(\hat{\mu} - \mu) \Rightarrow \mathcal{N}(0, \sigma^2) \quad \text{and} \quad Y_j \Rightarrow iid\mathcal{N}(0, \sigma^2), \ \ j = 1, \cdots, q \quad (1)$$

  for some given q, but not whether $y_t \sim iid(\mu, \sigma^2)$.

  Could try to pretest whether $y_t \sim iid(\mu, \sigma^2)$. If it looks that way, rely on OLS standard errors. Otherwise, revert to $\sqrt{T}(\hat{\mu} - \mu)/\hat{\omega}_{Y,q} \overset{a}{\sim} t_q$.

- Is it possible to do this pretesting in a way such that inference remains asymptotically valid whenever (1) holds?

  Müller (2011): No. Asymptotically efficient inference is obtained by treating (1) as the only relevant knowledge.

Fixed-b Asymptotics

- Set $S_T = bT$ in kernel estimator

$$\hat{\omega}^2_{k,S_T} = \sum_{j=-T+1}^{T-1} k(j/S_T)\hat{\gamma}(j)$$

  $\Rightarrow$ under weak dependence, leads to nonstandard limiting distribution of $\sqrt{T}(\hat{\mu} - \mu)/\hat{\omega}_{k,bT}$ derived in Kiefer and Vogelsang (2005) (which only depends on b and k)

- In general, no straightforward spectral interpretation.

- Special case of Newey-West estimator with $S_T = T$:

$$\hat{\omega}^2_{KVB} = \sum_{l=1}^{T-1} \frac{1}{\pi^2 l^2} Y_l^2 + o_p(1).$$

Effect of Second-Moment Instabilities

- Distributional approximations for inconsistent LRV estimators (essentially) require second-order stationarity: They break down if, say,

$$y_t = \mu + (1 + \mathbf{1}[t > T/2])\varepsilon_t \quad \text{with} \quad \varepsilon_t \sim iid(0, \sigma^2).$$

- In contrast, the OLS variance estimator (and any consistent HAC estimator) consistent for the right object,

$$\hat{\sigma}^2 \xrightarrow{p} \lim_{T \to \infty} \mathrm{Var}[T^{1/2}\hat{\mu}].$$

- In that sense, inference with inconsistent LRV estimators is not uniformly more robust.

Ibragimov and Müller (2011) Approach

- Partition data in q blocks of contiguous data, and estimate μ from each of the blocks.

- Under weak dependence, resulting estimators $\hat{\mu}_l$, $l = 1, \cdots, q$, are approximately independent and Gaussian with mean μ.

- Consider usual t-statistic for the mean using these q estimators only,

$$t = \frac{\sqrt{q}(\overline{\hat{\mu}} - \mu)}{\sqrt{q^{-1}\sum_{l=1}^{q}(\hat{\mu}_l - \overline{\hat{\mu}})^2}}, \qquad \overline{\hat{\mu}} = q^{-1}\sum_{l=1}^{q}\hat{\mu}_l.$$

- By remarkable result of Bakirov and Szekely (2005), using student-t critical values at conventional significance levels yields correct inference, even under pronounced variance heterogeneity in $\hat{\mu}_l$.

Inference under AR(1) Persistence

- Suppose initially that y_t is exactly a Gaussian AR(1) with mean μ, coefficient $\rho < 1$ and error variance σ^2.

- Let $Y_0 = \sqrt{T}\hat{\mu}$, and Y_l the lth DCT. Then

$$Y = (Y_0, \cdots, Y_q)' \sim \mathcal{N}(T^{1/2}\mu\iota_1, \sigma^2\Omega(\rho))$$

  where ι_1 is first column of I_{q+1}.

- For fixed ρ, as $T \to \infty$, $\sigma^2\Omega(\rho)$ becomes proportional to I_{q+1} (with factor of proportionality $\omega^2 = \sigma^2/(1-\rho)^2$).

- But for any T, there exists ρ sufficiently close to 1 for which $\Omega(\rho)$ is far from being proportional to I_{q+1}.

Local-To-Unity Asymptotics

- Under local-to-unity embedding $\rho = \rho_T = 1 - c/T$ with some fixed $c > 0$,

  $$T^{-2}\Omega(\rho_T) \to \Omega_0(c)$$

  so that

  $$T^{-1}Y \overset{a}{\sim} \mathcal{N}(T^{-1/2}\mu\iota_1, \sigma^2\Omega_0(c)).$$

- Same distributional approximation holds more generally under local-to-unity asymptotics.

- As $c \to \infty$, $\Omega(c)$ becomes again proportional to I_{q+1}. Thus, LTU theory contains weak dependence asymptotics as a special case.

  $c \to 0$ corresponds to very nearly unit root behavior of y_t.

- Number q determines the over which frequency band LTU approximation is imposed.

Hypothesis Testing with Nuisance Parameter

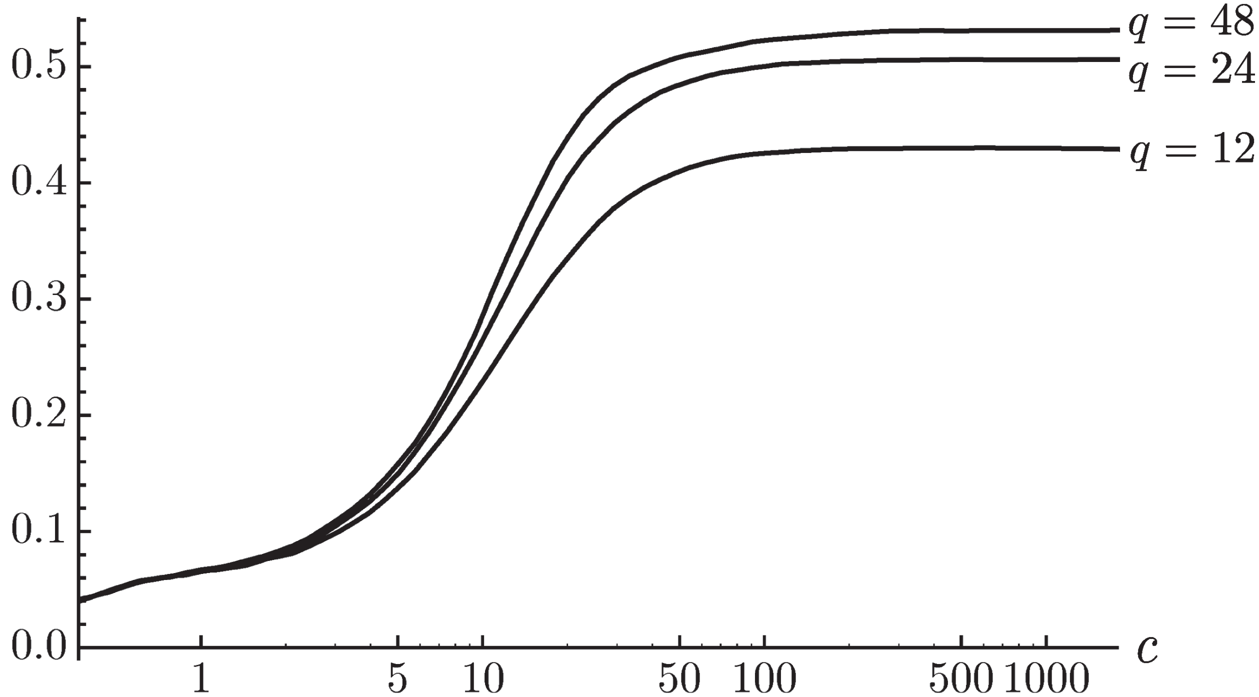
- In $T^{-1}Y \sim \mathcal{N}(T^{-1/2}\mu\iota_1, \sigma^2\Omega_0(c))$, consider tests of $H_0 : \mu = 0$ that maximize weighted average power against alternatives $\mu \neq 0$, where the weight on μ is $\mathcal{N}(0, 10T\sigma^2/c^2)$. (This variance yields approximately 50% power for 5% level tests for known and large c.)

- Transforms problem into testing $H_0 : T^{-1}Y \sim \mathcal{N}(0, \sigma^2\Omega_0(c))$ against $H_1 : T^{-1}Y \sim \mathcal{N}(0, \sigma^2\Omega_1(c))$, where $\Omega_1(c) = \Omega_0(c) + (10/c^2)\iota_1\iota_1'$.

- Impose scale invariance to eliminate σ from the testing problem.

- Generic testing problem with unknown and not consistently estimable nuisance parameter $c > 0$. Use insights of Elliott, Müller and Watson (2012) to derive scale invariant test that maximizes weighted average power over c (with weight function that is approximately uniform on $[0, 7]$ for $\log(c)$).

Weighted Average Power for $\mu \sim \mathcal{N}(0, 10T\sigma^2/c^2)$



Size control for $c > 0$ makes it impossible to generate power for very small c, as y_0 and μ are not individually identified under $y_0 - \mu \sim \mathcal{N}(0, (1-\rho^2)^{-1})$ for $\rho \to 1$.

Application to Unemployment Series

All CIs are of the form $\hat{\mu} = 5.80 \pm$m.e., where

	S_{24}	$\hat{\omega}^2_{A91}$	$\hat{\omega}^2_{AM}$	$\hat{\omega}^2_{KVB}$	$\hat{\omega}^2_{Y,12}$	IM_8
m.e.	1.31	0.75	2.21	1.46	0.85	1.02

Generalization to Regression and GMM

- For inference about regression coefficient, problem is to estimate LRV of suitably linear combination of product of regressor times regression error.

- Can be recast as problem of inference about mean of scalar time series.

- But validity and optimality of AR(1)-type inference under strong dependence requires very special assumptions.

Monte Carlo Results

AR(1) design, $T = 200$

ρ	S_{24}	$\hat{\omega}^2_{A91}$	$\hat{\omega}^2_{AM}$	$\hat{\omega}^2_{KVB}$	$\hat{\omega}^2_{Y,12}$	IM_8
			Size			
0.0	4.9	5.5	5.5	5.0	5.2	5.1
0.7	4.9	10.0	7.0	6.1	6.3	6.0
0.98	4.7	44.2	26.5	23.0	48.3	37.3
			Size-adjusted power			
0.0	42.8	50.0	49.9	37.3	44.3	40.1
0.7	41.7	46.4	47.0	36.2	45.1	41.5
0.98	12.5	40.0	38.1	35.1	51.4	47.2

Monte Carlo Results

Linear regression with AR(1) regressor and independent AR(1) error, $T = 200$

ρ	S_{24}	$\hat{\omega}^2_{A91}$	$\hat{\omega}^2_{AM}$	$\hat{\omega}^2_{KVB}$	$\hat{\omega}^2_{Y,12}$	IM_8
			Size			
0.0	4.9	5.9	6.0	5.1	5.4	5.0
0.7	5.0	10.2	8.0	6.7	7.3	5.3
0.98	2.7	36.4	26.0	22.4	31.2	5.2
			Size-adjusted power			
0.0	59.1	68.9	68.7	51.9	62.1	53.7
0.7	44.6	49.9	49.4	38.6	46.4	45.8
0.98	25.2	44.6	39.3	38.0	46.4	93.2

Conclusions

- Choice of LRV estimator strongly influence empirical conclusions

  $\Rightarrow$ necessary to understand rationale behind alternative estimators

- Valid inference always requires some regularity. In my mind, most natural and interpretable to impose regularity from a spectral point of view.

- Most popular approaches (consistent and inconsistent) exploit flatness of spectrum close to origin.

- New test instead takes particular parametric stand on spectrum close to origin, and accounts for the parameter uncertainty of this parametric form.