AEA Continuing Education Program

Time-Series Econometrics

James Stock, Harvard University

And

Mark Watson, Princeton University

January 6-8, 2019

AEA Continuing Education Course

Time Series Econometrics

Lecture 1: Time series refresher

Mark W. Watson

January 6, 2019

4:00PM – 6:00PM

Course Topics

1. Time series refresher (MW)
2. Heteroskedasticity and autocorrelation consistent/robust (HAC, HAR) standard errors (JS)
3. Dynamic causal effects (JS)
4. Weak instruments/weak identification in IV and GMM (JS)
5. Dynamic factor models and prediction with large datasets (MW)
6. Low-frequency analysis of economic time series (MW)

(Data + Examples in (2)-(6), but not today.)

Lecture Outline: Time series refresher

26 concepts

Time Series Basics (and notation)

(References: Hayashi (2000), Hamilton (1994), Brockwell and Davis (1991)… , lots of other books)

1. $\{Y_t\}$: a sequence of random variables

2. Stochastic Process: The probability law governing $\{Y_t\}$

3. Realization: One draw from the process, $\{y_t\}$

4. Strict Stationarity: The process is strictly stationary if the probability distribution of $(Y_t, Y_{t+1}, ..., Y_{t+k})$ is identical to the probability distribution of $(Y_\tau, Y_{\tau+1}, ..., Y_{\tau+k})$ for all t, τ, and k. (Thus, all joint distributions are time invariant.)

5. Autocovariances: $\gamma_{t,k} = cov(Y_t, Y_{t+k})$

6. Autocorrelations: $\rho_{t,k} = cor(Y_t, Y_{t+k})$

7. Covariance Stationarity: The process is covariance stationary if $\mu_t = E(Y_t) = \mu$ and $\gamma_{t,k} = \gamma_k$ for all t and k.

8. White noise: A process is called white noise if it is covariance stationary and $\mu = 0$ and $\gamma_k = 0$ for $k \neq 0$.

9. Martingale: Y_t follows a martingale process if $\mathbf{E}(Y_{t+1} \mid \mathcal{F}_t) = Y_t$, where $\mathcal{F}_t \subseteq \mathcal{F}_{t+1}$ is the time t information set.

10. Martingale Difference Process: Y_t follows a martingale difference process if $\mathbf{E}(Y_{t+1} \mid \mathcal{F}_t) = 0$. $\{Y_t\}$ is called a martingale difference sequence or "mds."

11. The Lag Operator: "L" lags the elements of a sequence by one period. $\mathrm{L}y_t = y_{t-1}$, $\mathrm{L}^2 y_t = y_{t-2}$,. If b denotes a constant, then $b\mathrm{L}Y_t = \mathrm{L}(bY_t) = bY_{t-1}$.

12. Linear filter (moving averages): Let $\{c_j\}$ denote a sequence of constants and

$$c(\mathrm{L}) = c_{-r}\mathrm{L}^{-r} + c_{-r+1}\mathrm{L}^{-r+1} + \ldots + c_0 + c_1\mathrm{L} + \ldots + c_s\mathrm{L}^s$$

denote a polynomial in L. Note that $X_t = c(\mathrm{L})Y_t = \sum_{j=-r}^{s} c_j Y_{t-j}$ is a moving average of Y_t. $c(\mathrm{L})$ is sometimes called a linear filter (for reasons discussed below) and X is called a filtered version of Y.

(notational simplification: $\mathbf{E}(Y_t) = 0$)

13. AR(p) process: $\phi(\mathrm{L})Y_t = \eta_t$ where $\phi(\mathrm{L}) = (1 - \phi_1\mathrm{L} - \ldots - \phi_p\mathrm{L}^p)$ and η_t is unpredictable given lags of Y.

Alternatively: $Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \eta_t$

Jargon: η_t is the 1-period ahead forecast error in Y_t, where the forecasts are constructed using lagged values of Y_t. In the AR model η is sometimes called an "innovation" or (from concept 18 below) a "Wold" error or a "fundamental" error.

14. MA(q) process: $Y_t = \theta(\mathrm{L})\eta_t$ where $\theta(\mathrm{L}) = (1 - \theta_1\mathrm{L} - \ldots - \theta_q\mathrm{L}^q)$ and η_t is white noise.

Alternatively: $Y_t = \eta_t - \theta_1\eta_{t-1} - \theta_2\eta_{t-2} - \ldots - \theta_q\eta_{t-q}$

Note: If η_t can be recovered from current and past values of Y_t, then the MA process is said to be 'invertible'. (Algebra shows that invertibility follows when the roots of the MA polynomial, $\theta(z)$, are greater than 1 in absolute value.)

When the process is invertible, η_t is the 1-period ahead forecast error in Y_t, where the forecasts are constructed using lagged values of Y_t. In this case the η errors are 'fundamental' or 'Wold' errors like their AR counterparts.

15. ARMA(p,q): $\phi(\mathrm{L})Y_t = \theta(\mathrm{L})\eta_t$.

16. Minimum mean squared error prediction. Suppose Y is a scalar and X is a vector. You observe $X = x$; what is a good guess of the value of Y? Let $y^{prediction} = g(x)$ denote the predicted value of Y, with prediction error $e = Y - y^{prediction}$, and $m.s.p.e. = \mathbf{E}(e^2)$.

Result: the minimum $m.s.p.e.$ predictor is $\mathbf{E}(Y \mid X = x)$.

17. Linear minimum mean square error prediction.

If (X, Y) are jointly normally distributed, $\mathbf{E}(Y \mid X)$ is linear.

$$\mathbf{E}(Y \mid X) = \alpha + \beta' X \quad \text{with } \beta = \Sigma_{XX}^{-1}\Sigma_{XY} \text{ and } \alpha = \mu_Y - \beta'\mu_X.$$

with $mspe = \text{var}(Y \mid X) = \sigma_e^2 \quad \left(= \Sigma_{YY} - \Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}\right)$.

This yields the best (minimum *mspe*) **linear** predictor of Y given X (even when Y and X are not normally distributed.) This is sometimes written as the 'projection' if Y onto X (and a constant).

18. Wold decomposition theorem (e.g., Brockwell and Davis (1991))
Suppose Y_t is generated by a "linearly indeterministic" covariance stationary process. Then Y_t can be represented as

$$Y_t = \eta_t + c_1\eta_{t-1} + c_2\eta_{t-2} + \ldots ,$$

where η_t is white noise with variance σ_ε^2, $\sum_{i=1}^{\infty} c_i^2 < \infty$, and

$\eta_t = Y_t - Proj(Y_t \mid Y^{t-1})$, where $Y^{t\text{-}1} = (Y_{t-1}, Y_{t-2}, \ldots)$

Notice that η_t is a function of Y_t and lags of Y_t; it is said to be "fundamental".

19. Multiperiod prediction: Predict the value of Y_{t+h} given (Y_t, Y_{t-1}, …)
(set $\mathbf{E}(Y) = 0$ for notational convenience.)

Best linear predictor: $Y_{t+h}^{\text{Predictor}} = Proj(Y_{t+h}|Y^t) = \beta_h(\mathrm{L})Y_t$ ('Direct forecast')

(note: *need to estimate* $\beta_h(L)$ *for all h of interest*)

Iterated formulae:

$$Proj(Y_{t+1}|\ Y^t) = \beta_1(\mathrm{L})Y^t$$

$$Proj(Y_{t+2}|Y_{t+1},Y^t) = \beta_1(\mathrm{L})Y_{t+1} = \beta_{1,0}Y_{t+1} + \beta_{1,1}Y_t + \beta_{1,2}Y_{t-1} + \ldots$$

$$Proj(Y_{t+2}|\ Y^t) = \beta_{1,0}\,Proj(Y_{t+1}|Y^t) + \beta_{1,1}Y_t + \beta_{1,2}Y_{t-1} + \ldots$$

$$Proj(Y_{t+3}|\ Y^t) = \beta_{1,0}\,Proj(Y_{t+2}|Y^t) + \beta_{1,1}Proj(Y_{t+1}|Y^t) + \beta_{1,2}Y_{t-1} + \ldots$$

etc.

(note: *only need to estimate* $\beta_1(L)$, can be used for any h).

20. Vector processes:

Univariate AR(1): $Y_t = \phi Y_{t-1} + \eta_t$

Vector AR(1) (VAR(1)): $Y_t = \Phi Y_{t-1} + \eta_t$ (Y_t is $n \times 1$, Φ is $n \times n$, η_t is $n \times 1$)

VAR(p): $Y_t = \Phi Y_{t-1} + \Phi_2 Y_{t-2} + \ldots + \Phi_p Y_{t-p} + \eta_t$

or $(I - \Phi_1 \mathrm{L} - \Phi_2 \mathrm{L}^2 - \ldots \Phi_p \mathrm{L}^p) Y_t = \eta_t$

or $\Phi(\mathrm{L}) Y_t = \eta_t$

21. Invert AR to get MA

AR(1): $Y_t = \phi Y_{t-1} + \eta_t \Rightarrow Y_t = \phi^t \eta_0 + \sum_{i=0}^{t-1} \phi^i \eta_{t-i} = \sum_{i=0}^{\infty} \phi^i \eta_{t-i}$

$(1 - \phi \mathrm{L}) Y_t = \eta_t \Rightarrow Y_t = (1 - \phi \mathrm{L})^{-1} \eta_t$

$\phi(\mathrm{L}) Y_t = \eta_t \Rightarrow Y_t = \phi(\mathrm{L})^{-1} \eta_t$

AR(p): $\phi(\mathrm{L}) Y_t = \eta_t \Rightarrow Y_t = \phi(\mathrm{L})^{-1} \eta_t$

VAR(p): $\Phi(\mathrm{L}) Y_t = \eta_t \Rightarrow Y_t = \Phi(\mathrm{L})^{-1} \eta_t$

or $Y_t = \mathrm{C}(\mathrm{L}) \eta_t = \eta_t - \mathrm{C}_1 \eta_{t-1} - \mathrm{C}_2 \eta_{t-2} - \ldots$ with $\mathrm{C}(\mathrm{L}) = \Phi(\mathrm{L})^{-1}$

Jargon: $\partial Y_{i,t+h} / \partial \eta_{j,t} = C_{h,ij}$ is called an "impulse response"

22. The autocovariance generating function for a covariance stationary process is given by $\gamma(z) = \sum_{j=-\infty}^{\infty} \gamma_j z^j$, so the autocovariances are given by the coefficients on the argument z^j.

With Y represented as $Y_t = c(\mathrm{L})\eta_t$, the ACGF is $\gamma(z) = \sigma_\eta^2 c(z)c(z^{-1})'$.

Example: For the scalar MA(1) model $Y_t = (1 - c_1\mathrm{L})\eta_t$

$\gamma_0 = \sigma_\eta^2(1 + c_1^2)$, $\gamma_{-1} = \gamma_1 = -\sigma_\eta^2 c_1$, and $\gamma_k = 0$ for $|k| > 1$. Thus

$$
\begin{aligned}
\gamma(z) &= \sum_{j=-\infty}^{\infty} \gamma_j z^j \\
&= \gamma_{-1} z^{-1} + \gamma_0 z^0 + \gamma_1 z^1 \\
&= \sigma_\eta^2\left(-c_1 z^{-1} + (1 + c_1^2) - c_1 z\right) \\
&= \sigma_\eta^2 (1 - c_1 z)(1 - c_1 z^{-1})
\end{aligned}
$$

23. Spectral Representation Theorem (e.g, Brockwell and Davis (1991)): Suppose Y_t is a scalar discrete time covariance stationary zero mean process, then there exists an orthogonal-increment process $Z(\omega)$ such that

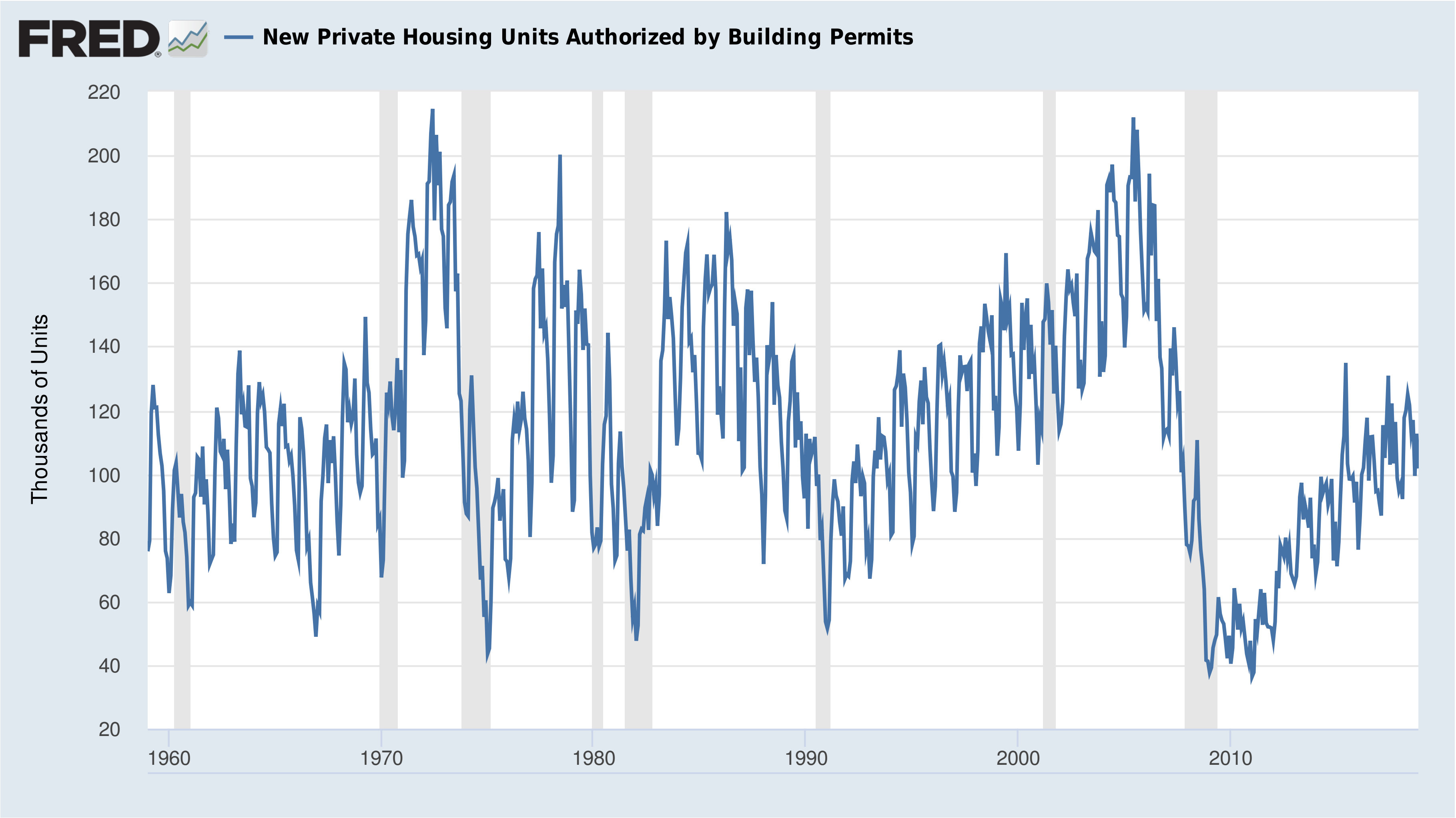
(i) $\text{Var}(Z(\omega)) = F(\omega)$

and

(ii) $Y_t = \int_{-\pi}^{\pi} e^{it\omega}\, dZ(\omega)$

where F is the spectral distribution function of the process. (The spectral density, $S(\omega)$, is the density associated with F.)

This is a useful decomposition, and we'll spend some time discussing it.



Shaded areas indicate U.S. recessions Source: U.S. Bureau of the Census myf.red/g/muwf

Some questions

1. How important are the “seasonal” or “business cycle” components in Y_t?

2. Can we measure the variability at a particular frequency? Frequency 0 (long-run) will be particularly important as that is what HAC/HAR Covariance matrices are all about.

3. Can we isolate/eliminate the “seasonal” (“business-cycle”) component? (Ex-Post vs. Real Time).

Spectral representation of a covariance stationary stochastic process

<u>Deterministic</u> processes:

(a) $Y_t = \cos(\omega t)$, strictly periodic with $period = \dfrac{2\pi}{\omega}$,

$Y_0 = 1$

amplitude = 1.

(b) $Y_t = a \times \cos(\omega t) + b \times \sin(\omega t)$, strictly period with $period = \dfrac{2\pi}{\omega}$,

$Y_0 = a$

amplitude = $\sqrt{a^2 + b^2}$

Stochastic process:

$Y_t = a \times \cos(\omega t) + b \times \sin(\omega t)$, a and b are random variables, 0-mean, mutually uncorrelated, with common variance σ^2.

2nd - moments :

$E(Y_t) = 0$

$Var(Y_t) = \sigma^2 \times \{\cos^2(\omega t) + \sin^2(\omega t)\} = \sigma^2$

$$
\begin{aligned}
Cov(Y_t, Y_{t-k}) &= \sigma^2 \{\cos(\omega t)\cos(\omega(t-k)) + \sin(\omega t)\sin(\omega(t-k))\} \\
&= \sigma^2 \cos(\omega k)
\end{aligned}
$$

Stochastic process with more components:

$Y_t = \sum_{j=1}^{n} \{a_j \cos(\omega_j t) + b_j \sin(\omega_j t)\}$, $\{a_j, b_j\}$ are uncorrelated 0-mean random variables, with $\mathrm{Var}(a_j) = \mathrm{Var}(b_j) = \sigma_j^2$

2nd - moments :

$\mathrm{E}(Y_t) = 0$

$\mathrm{Var}(Y_t) = \sum_{j=1}^{n} \sigma_j^2$ (Decomposition of variance)

$\mathrm{Cov}(Y_t Y_{t-k}) = \sum_{j=1}^{n} \sigma_j^2 \cos(\omega_j k)$ (Decomposition of auto-covariances)

Stochastic Process with even more components:

$$Y_t = \int_0^{\pi} \cos(\omega t)\, da(\omega) + \int_0^{\pi} \sin(\omega t)\, db(\omega)$$

$da(\omega)$ and $db(\omega)$: random variables, 0-mean, mutually uncorrelated, uncorrelated across frequency, with common variance that depends on frequency. This variance function, say $S(\omega)$, is called the spectrum.

.. **Digression:** A convenient change of notation:

$$Y_t = a\times\cos(\omega t) + b\times\sin(\omega t)$$

$$= \frac{1}{2}e^{i\omega}(a - ib) + \frac{1}{2}e^{-i\omega}(a + ib)$$

$$= e^{i\omega}Z + e^{-i\omega}\bar{Z}$$

where $i = \sqrt{-1}$ and $e^{i\omega} = cos(\omega) + i\times sin(\omega)$, $Z = \frac{1}{2}(a - ib)$ and $\bar{Z}$ is the complex conjugate of Z.

Similarly

$$Y_t = \int_0^{\pi} \cos(\omega t)\, da(\omega) + \int_0^{\pi} \sin(\omega t)\, db(\omega)$$

$$= \frac{1}{2}\int_0^{\pi} e^{i\omega t}(da(\omega) - i\, db(\omega)) + \frac{1}{2}\int_0^{\pi} e^{-i\omega t}(da(\omega) + i\, db(\omega))$$

$$= \int_{-\pi}^{\pi} e^{i\omega t}\, dZ(\omega)$$

where $dZ(\omega) = \frac{1}{2}(da(\omega) - idb(\omega))$ for $\omega \geq 0$ and $dZ(-\omega) = \overline{dZ(\omega)}$ for $\omega >$ 0.

Because da and db have mean zero, so does dZ. Denote the variance of $dZ(\omega)$ as $\mathrm{Var}(dZ(\omega)) = \mathrm{E}(dZ(\omega)\overline{dZ(\omega)}) = S(\omega)d\omega$, and using the assumption that da and db are uncorrelated across frequency $\mathrm{E}(dZ(\omega)\overline{dZ(\omega)'}) = 0$ for $\omega \neq \omega'$.

Second moments of Y:

$$E(Y_t) = E\left\{\int_{-\pi}^{\pi} e^{i\omega t}\, dZ(\omega)\right\} = \int_{-\pi}^{\pi} e^{i\omega t} E(dZ(\omega)) = 0$$

$$\gamma_k = E(Y_t Y_{t-k}) = E(Y_t \bar{Y}_{t-k}) = E\left\{\int_{-\pi}^{\pi} e^{i\omega t}\, dZ(\omega) \int_{-\pi}^{\pi} e^{-i\omega(t-k)}\, \overline{dZ(\omega)}\right\}$$

$$= \int_{-\pi}^{\pi} e^{i\omega t} e^{-i\omega(t-k)} E(dZ(\omega)\overline{dZ(\omega)})$$

$$= \int_{-\pi}^{\pi} e^{i\omega k} S(\omega)\, d\omega = 2\int_{0}^{\pi} \cos(\omega k) S(\omega)\, d\omega$$

where the last equality follows from $S(\omega) = S(-\omega)$.

Setting $k = 0$, $\gamma_0 = \mathrm{Var}(Y_t) = \int_{-\pi}^{\pi} S(\omega)\, d\omega$

… End of Digression

Summarizing

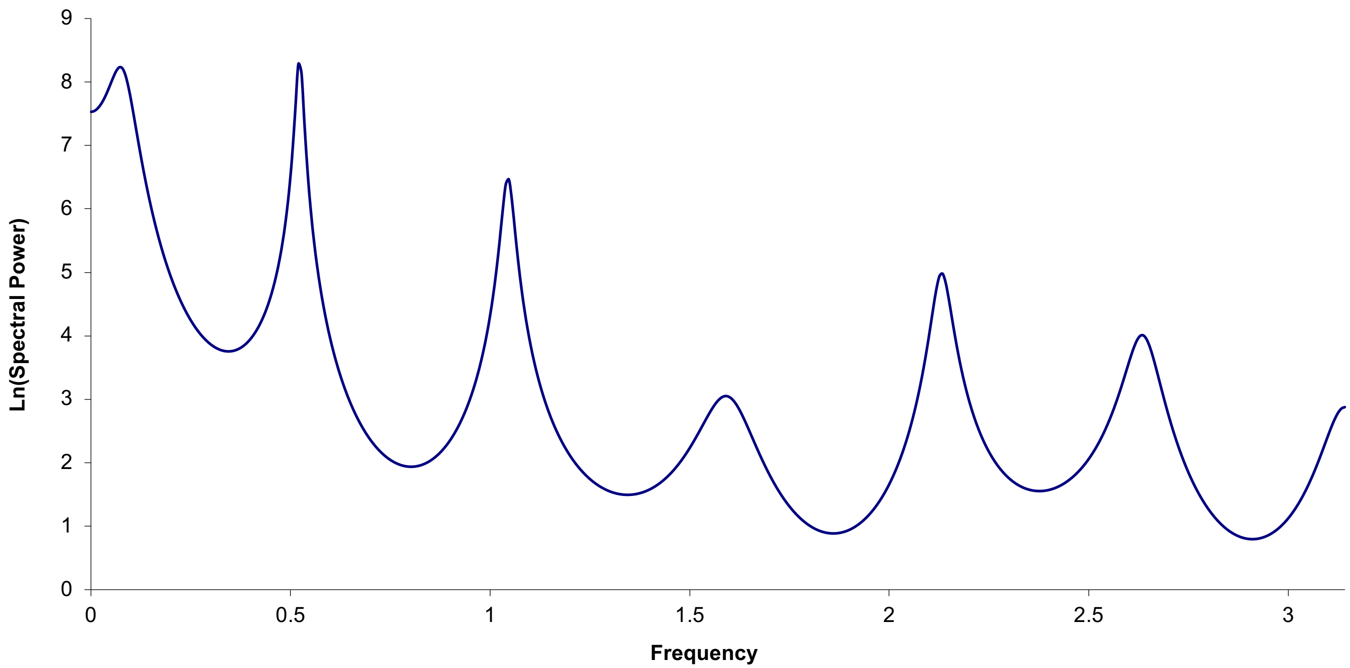
1. $S(\omega)d\omega$ can be interpreted as the variance of the cyclical component of Y corresponding to the frequency ω. The period of this component is $period = 2\pi/\omega$.

2. $S(\omega) \geq 0$ (it is a variance)

3. $S(\omega) = S(-\omega)$. Because of this symmetry, plots of the spectrum are presented for frequencies $0 \leq \omega \leq \pi$.

Example: The Spectrum of Building Permits

Figure 2
Spectrum of Building Permits



Most of the mass in the spectrum is concentrated around the seven peaks evident in the plot. (These peaks are sufficiently large that spectrum is plotted on a log scale.) The first peak occurs at frequency $\omega = 0.07$ corresponding to a period of 90 months. The other peaks occur at frequencies $2\pi/12$, $4\pi/12$, $6\pi/12$, $8\pi/12$, $10\pi/12$, and π. These are peaks for the seasonal frequencies: the first corresponds to a period of 12 months, and the others are the seasonal "harmonics" 6, 4, 3, 2.4, 2 months. (These harmonics are necessary to reproduce an arbitrary − not necessary sinusoidal − seasonal pattern.)

4. $\gamma_k = \int_{-\pi}^{\pi} e^{i\omega k} S(\omega)\, d\omega = 2\int_0^{\pi} \cos(\omega k) S(\omega)\, d\omega$ can be inverted to yield

$$S(\omega) = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty} e^{-i\omega k}\gamma_k = \frac{1}{2\pi}\left\{\gamma_0 + 2\sum_{k=1}^{\infty}\gamma_k \cos(\omega k)\right\}$$

5. From (4): $S(\omega) = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty} e^{-i\omega k}\gamma_k = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty} z^k\gamma_k = (2\pi)^{-1}\gamma(z)$

with $z = e^{-i\omega}$. Thus, $S(\omega)$ is easily computed from ACGF.

24. "Long-Run Variance" and sampling variability in the sample mean.

The long-run variance is $S(0)$, the variance of the 0-frequency (or ∞-period component).

Since $S(\omega)=\frac{1}{2\pi}\sum_{k=-\infty}^{\infty}e^{-i\omega k}\gamma_k$, then $S(0)=\frac{1}{2\pi}\sum_{k=-\infty}^{\infty}\gamma_k e^{-ik0}=\frac{1}{2\pi}\sum_{k=-\infty}^{\infty}\gamma_k$.

This plays an important role in statistical inference because (except for the factor 2π) it is the large-sample variance of the sample mean.

Suppose Y_t is stationary with autocovariances λ_i. Then

$$
\begin{aligned}
&\mathrm{var}(T^{-1/2}\sum_{t=1}^{T} Y_t) \\
&= \frac{1}{T}\{T\gamma_0 + (T-1)(\gamma_1+\gamma_{-1}) + (T-2)(\gamma_2+\gamma_{-2}) + \ldots 1(\gamma_{T-1}+\gamma_{1-T})\} \\
&= \sum_{j=-T+1}^{T-1} \gamma_j - \frac{1}{T}\sum_{j=1}^{T-1} j(\gamma_j + \gamma_{-j})
\end{aligned}
$$

If the autocovariances satisfy $\sum_{j=1}^{\infty} j\,|\,\gamma_j\,|$ (jargon: they are "1-summable") then

$$
\mathrm{var}(T^{-1/2}\sum_{t=1}^{T} Y_t) \to \sum_{j=-\infty}^{\infty} \gamma_j = 2\pi S(0)
$$

3 Estimators for $\sum_{j=-\infty}^{\infty} \gamma_j$:

(1) $\sum_{j=-k}^{k} \hat{\gamma}_j$

(2) if $Y_t = \mu + c(\mathrm{L})\eta_t$ then $\lambda(z) = \sigma^2 c(z)c(z^{-1})$ and $\sum_{j=-\infty}^{\infty} \gamma_j = \gamma(1) = \sigma^2 c(1)^2$.

with $c(\mathrm{L}) = \theta(\mathrm{L})/\phi(\mathrm{L})$ ($Y_t \sim$ ARMA), then

$$\sum_{j=-\infty}^{\infty} \gamma_j = \sigma^2 \frac{\theta(1)^2}{\phi(1)^2} = \sigma^2 \frac{(1-\theta_1 - \ldots - \theta_q)^2}{(1-\phi_1 - \ldots - \phi_p)^2}$$

(3) 'spectral estimators' based on low-frequency weighted averages of Y. (JS)

25. Recursive prediction, signal extraction and the Kalman filter.

Linear Gaussian Model

$$y_t = Hs_t + \varepsilon_t$$

$$s_t = Fs_{t-1} + \eta_t$$

$$\begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} \sim iidN\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_\varepsilon & 0 \\ 0 & \Sigma_\eta \end{pmatrix} \right)$$

Applications:

- Unobserved component models (*s* is serially correlated part of *y*)
- Factor Models (many *y's*, few *s's*)
- TVP Regression models ($H = H_t = x_t$, $s_t = \beta_t$)
- Extensions: (many)

Recall that if

$$\begin{pmatrix} a \\ b \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_a \\ \mu_b \end{pmatrix}, \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix}\right),$$

then $(a|b) \sim N(\mu_{a|b}, \Sigma_{a|b})$

where $\mu_{a|b} = \mu_a + \Sigma_{ab}\Sigma_{bb}^{-1}(b - \mu_b)$ and $\Sigma_{a|b} = \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba}$.

Interpreting *a* and *b* appropriately yields the Kalman Filter and Kalman Smoother.

Model: $y_t = Hs_t + \varepsilon_t$, $s_t = Fs_{t-1} + \eta_t$, $\begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} \sim iidN\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_\varepsilon & 0 \\ 0 & \Sigma_\eta \end{pmatrix}\right)$

Let $s_{t/k} = E(s_t \mid y_{1:k})$, $P_{t/k} = \mathrm{Var}(s_t \mid y_{1:k})$,

$\mu_{t/t-1} = E(y_t \mid y_{1:t-1})$, $\Sigma_{t/t-1} = \mathrm{Var}(y_t \mid y_{1:t-1})$.

Deriving Kalman Filter:

Starting point: $s_{t-1} \mid y_{1:t-1} \sim N(s_{t-1/t-1}, P_{t-1/t-1})$. Then

$$\begin{pmatrix} s_t \\ y_t \end{pmatrix} \mid y_{1:t-1} \sim N\left(\begin{pmatrix} s_{t/t-1} \\ y_{t/t-1} \end{pmatrix}, \begin{pmatrix} P_{t/t-1} & P_{t/t-1}H' \\ HP_{t/t-1} & HP_{t/t-1}H' + \Sigma_\varepsilon \end{pmatrix}\right)$$

interpreting s_t as "a" and y_t as "b" yields the Kalman Filter.

Model: $y_t = Hs_t + \varepsilon_t,\ \ s_t = Fs_{t-1} + \eta_t, \begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} \sim iidN\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_\varepsilon & 0 \\ 0 & \Sigma_\eta \end{pmatrix}\right)$ and

$$\begin{pmatrix} s_t \\ y_t \end{pmatrix} | \, y_{1:t-1} \sim N\left(\begin{pmatrix} s_{t/t-1} \\ y_{t/t-1} \end{pmatrix}, \begin{pmatrix} P_{t/t-1} & P_{t/t-1}H' \\ HP_{t/t-1} & HP_{t/t-1}H' + \Sigma_\varepsilon \end{pmatrix}\right)$$

Details of KF :

(i) $s_{t/t-1} = Fs_{t-1/t-1}$

(ii) $P_{t/t-1} = FP_{t-1/t-1}F' + \Sigma_\eta$,

(iii) $\mu_{t/t-1} = Hs_{t/t-1}$,

(iv) $\Sigma_{t/t-1} = HP_{t/t-1}H' + \Sigma_\varepsilon$

(v) $K_t = P_{t/t-1}H'\Sigma_{t/t-1}^{-1}$

(vi) $s_{t/t} = s_{t/t-1} + K_t(y_t - \mu_{t/t-1})$

(vii) $P_{t/t} = (I - K_t)P_{t/t-1}$.

The density of $Y_{1:T}$ is $f(Y_{1:T}) = f(Y_T|Y_{1:T-1})f(Y_{1:T-1}) = \prod_{t=2}^{T} f(y_t \mid y_{1:t-1}) f(y_1)$

so the log-likelihood is

$$L(Y_{1:T}) = \text{constant } -0.5\sum_{t=1}^{T}\left\{\ln|\Sigma_{t|t-1}| + (y_t - \mu_{t|t-1})'\Sigma_{t|t-1}^{-1}(y_t - \mu_{t/t-1})\right\}$$

The Kalman Smoother (for $s_{t|T}$ and $P_{t|T}$) is derived in analogous fashion (see Anderson and Moore (2012), or Hamilton (1990).)

26. Recursive prediction, signal extraction, a more general formulation.

Models and objects of interest

General Model (Nonlinear, non-Gaussian state-space model)

$$y_t = H(s_t,\ \varepsilon_t)$$

$$s_t = F(s_{t-1},\ \eta_t)$$

$$\varepsilon_t \text{ and } \eta_t \sim i.i.d.$$

Jargon: This is sometimes called a "hidden Markov model" because s_t is "hidden" by the measurement error ε_t.

Example 1: Linear Gaussian Model

$$y_t = Hs_t + \varepsilon_t$$

$$s_t = Fs_{t-1} + \eta_t$$

$$\begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} \sim iidN\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_\varepsilon & 0 \\ 0 & \Sigma_\eta \end{pmatrix} \right)$$

Example 2: Hamilton Regime-Switching Model

$$y_t = \mu(s_t) + \sigma(s_t)\varepsilon_t$$

$s_t = 0$ or 1 with $P(s_t = i \mid s_{t-1} = j) = p_{ij}$

(using $s_t = F(s_{t-1}, \eta_t)$ notation:

$s_t = \mathbf{1}(\eta_t \leq p_{10} + (p_{11} - p_{10})s_{t-1})$, where $\eta \sim \mathrm{U}[0,1]$)

Example 3: Stochastic volatility model

$$y_t = e^{s_t}\varepsilon_t$$

$$s_t = \mu + \phi(s_{t-1} - \mu) + \eta_t$$

with, say, $\varepsilon_t \sim$ iid(0,1) and $e^{s_t} = \sigma_t$

the model for y is

$$y_t \mid s_t \sim N(0, \sigma_t^2)$$

Some things you might want to calculate

Notation: $y_{1:t} = (y_1, y_2, \ldots, y_t)$, $s_{1:t} = (s_1, s_2, \ldots, s_t)$,

$f(\,.\,|\,.\,)$ a generic density function.

A. Prediction and Likelihood

(i) $f(s_t \mid y_{1:t-1})$

(ii) $f(y_t \mid y_{1:t-1})$ … Note $f(y_{1:T}) = \prod_{t=1}^{T} f(y_t \mid y_{1:t-1})$ is the likelihood

B. Filtering: $f(s_t \mid y_{1:t})$

C. Smoothing: $f(s_t \mid y_{1:T})$.

2. General Formulae (Kitagawa (1987))

Model: $y_t = H(s_t, \varepsilon_t)$, $s_t = F(s_{t-1}, \eta_t)$, ε and η ~ iid

A. Prediction of s_t and y_t given Y_{t-1}.

$$
\begin{aligned}
f(s_t \mid y_{1:t-1}) &= \int f(s_t, s_{t-1} \mid y_{1:t-1})\, ds_{t-1} \\
\text{(i)} \qquad &= \int f(s_t \mid s_{t-1}, y_{1:t-1})\, f(s_{t-1} \mid y_{1:t-1})\, ds_{t-1} \\
&= \int f(s_t \mid s_{t-1})\, f(s_{t-1} \mid y_{1:t-1})\, ds_{t-1}
\end{aligned}
$$

$$f(y_t \mid y_{1;t-1}) = \int f(y_t, s_t \mid y_{1:t-1})\, ds_t$$

(ii) $$= \int f(y_t \mid s_t, y_{1:t-1}) f(s_t \mid y_{1:t-1})\, ds_t$$ ("t" component of likelihood)

$$= \int f(y_t \mid s_t) f(s_t \mid y_{1:t-1})\, ds_t$$

B. Filtering

$$f(s_t \mid y_{1:t}) = f(s_t \mid y_t, y_{1:t-1})$$

$$= \frac{f(y_t \mid s_t, y_{1:t-1}) f(s_t \mid y_{1:t-1})}{f(y_t \mid y_{1:t-1})}$$

$$= \frac{f(y_t \mid s_t) f(s_t \mid y_{1:t-1})}{f(y_t \mid y_{1:t-1})}$$

C. Smoothing

$$f(s_t \mid y_{1:T}) = \int f(s_t, s_{t+1} \mid y_{1:T})\, ds_{t+1}$$

$$= \int f(s_t \mid s_{t+1}, y_{1:T}) f(s_{t+1} \mid y_{1:T})\, ds_{t+1}$$

$$= \int f(s_t \mid s_{t+1}, y_{1:t}) f(s_{t+1} \mid y_{1:T})\, ds_{t+1}$$

$$= \int \left[\frac{f(s_{t+1} \mid s_t) f(s_t \mid y_{1:t})}{f(s_{t+1} \mid y_{1:t})} \right] f(s_{t+1} \mid y_{1:T}) ds_{t+1}$$

$$= f(s_t \mid y_{1:t}) \int f(s_{t+1} \mid s_t) \frac{f(s_{t+1} \mid y_{1:T})}{f(s_{t+1} \mid y_{1:t})} ds_{t+1}$$

Solving these integral equations depends on the structure of the problem.

Easy: Linear and normal (Kalman filter)

Pretty Easy: Hamilton model

$$y_t = \mu(s_t) + \sigma(s_t)\varepsilon_t$$

$s_t = 0$ or 1 with $P(s_t = i \mid s_{t-1} = j) = p_{ij}$

(simple recursive formulae for likelihood and filter – exercise: work this out).

Harder: Stochastic volatility

$$y_t = e^{s_t}\varepsilon_t$$

$$s_t = \mu + \phi(s_{t-1} - \mu) + \eta_t$$

…

Numerical methods are used to evaluate the required integrals: Importance sampling, MCMC and Particle Filtering.

References for Lecture 1

Anderson, B.D.O and J.B. Moore (2012), *Optimal Filtering*, Dover

Brockwell, P.J., and R.A. Davis (1991), *Time Series*: *Theory and Methods*, 2nd Edition, New York: Springer Verlag.

Hamilton, J.D. (1994), *Time Series Analysis*, Princeton: Princeton University Press

Hayashi, F. (2000), *Econometrics*. Princeton: Princeton University Press.

Kitagawa, G. (1987), "Non-Gaussian State-Space Modeling of Nonstationary Time Series," *Journal of the American Statistical Association*, 82(4): 1032 -1041.

Priestly, M.B. (1981), *Spectral Analysis and Time Series*, London: Academic Press.

Stock, James H., and Mark W. Watson (2018), *Introduction to Econometrics*, 4th Edition, Prentice Hall.

AEA Continuing Education Course
Time Series Econometrics

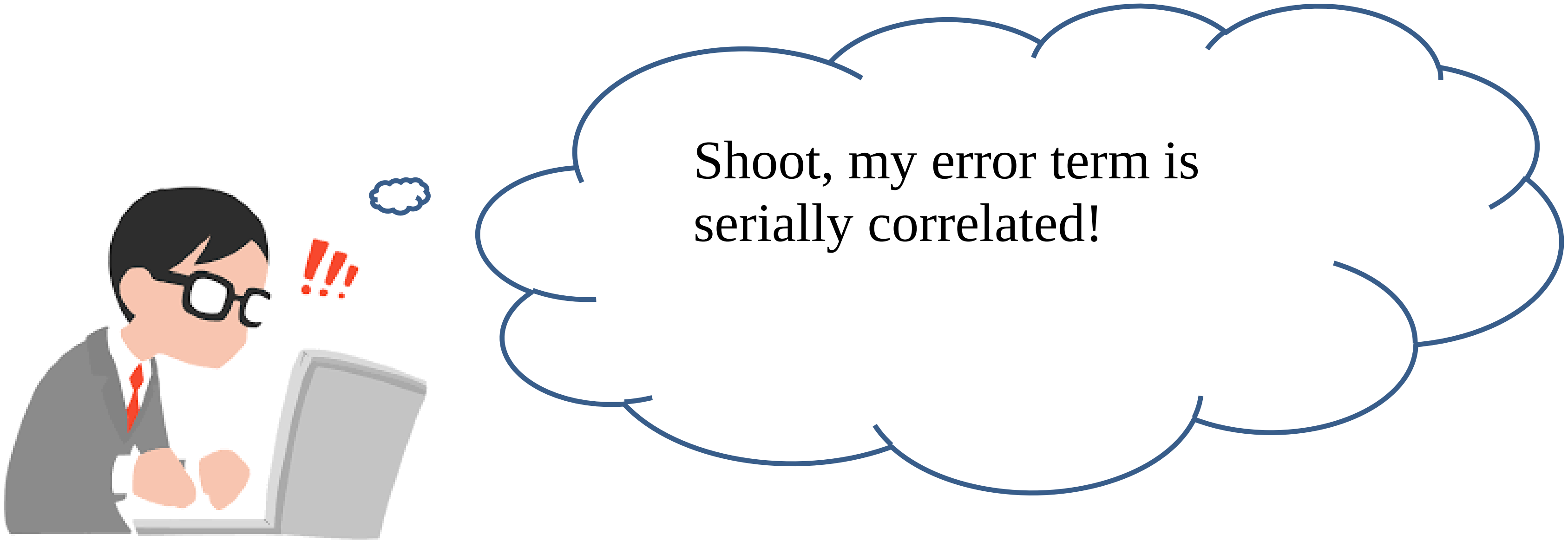
Lecture 2

Heteroskedasticity- and Autocorrelation-Robust Inference

James H. Stock
Harvard University

January 7, 2019, 8:15-9:45am

The Heteroskedasticity- and Autocorrelation Robust Inference (HAR) problem



The HAR problem

Guess I need to use Newey-West SEs.

The HAR problem

Guess I need to use Newey-West SEs.

- what truncation parameter S should I use?

The HAR problem

Guess I need to use Newey-West SEs.

- what truncation parameter S should I use?
- what critical value?

The HAR problem

Guess I need to use Newey-West SEs.

- what truncation parameter *S* should I use?
- what critical value?
- what about all those hard papers by Vogelsang, Müller, Sun, and others*?

*Kiefer, Vogelsang, Bunzel (2000), Velasco and Robinson (2001), Kiefer and Vogelsang (2002, 2005), Jansson (2004), Phillips (2005), Müller (2007, 2014), Sun, Phillips, & Jin (2008), Ibragimov and Müller (2010), Sun (2011, 2013, 2014a, 2014b), Gonçalves & Vogelsang (2011), Zhang and Shao (2013), Pötscher and Preinerstorfer (2016, 2017),…

The HAR problem

Guess I need to use Newey-West SEs.

- what truncation parameter *S* should I use?
- what critical value?
- what about all those hard papers by Vogelsang, Müller, Sun, and others*?
- nobody uses them anyway…

The HAR problem

Guess I need to use Newey-West SEs.

- what truncation parameter *S* should I use?
- what critical value?
- what about all those hard papers by Vogelsang, Müller, Sun, and others*?
- nobody uses them anyway…

  and the referees never complain…

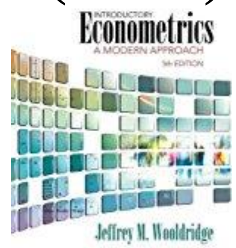
The HAR problem

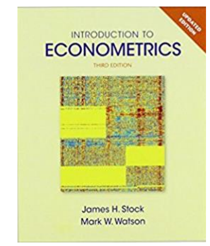
Guess I'll just use NW with ±1.96 and $S =$

The HAR problem

Guess I'll just use NW with ± 1.96 and $S =$

$4(T/100)^{2/9}$ or $0.75T^{1/3}$

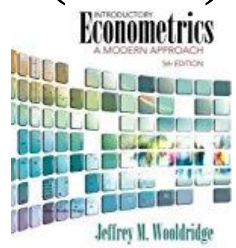


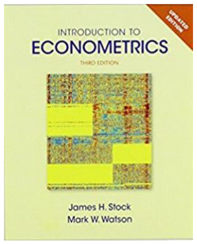


The HAR problem

Guess I'll just use NW with ±1.96 and S =

$4(T/100)^{2/9}$ or $0.75T^{1/3}$





hmm… they give the same answer…
must be OK…

20 years of research says: Bad idea.

Rejection rates of HAR tests with nominal level 5% ($b = S/T$)

$$y_t = \beta_0 + \beta_1 x_t + u_t,\ x_t \ \&\ u_t \text{ Gaussian AR(1)},\ \rho_x = \rho_u = 0.7^{1/2},\ T = 200$$

Estimator	Truncation rule for b	Critical values	Null imposed?	$\rho = 0.3$	$\rho = 0.5$	$\rho = 0.7$
NW	$0.75T^{-2/3}$	N(0,1)	No	**0.079**	**0.105**	**0.164**
NW	$1.3T^{-1/2}$	fixed-b (nonstandard)	No	0.067	0.080	0.107
EWP	$1.95T^{-2/3}$	fixed-b (t_v)	No	0.063	0.074	0.100
NW	$1.3T^{-1/2}$	fixed-b (nonstandard)	Yes	**0.057**	**0.062**	**0.073**
EWP	$1.95T^{-2/3}$	fixed-b (t_v)	Yes	0.052	0.056	0.066
Theoretical bound based on Edgeworth expansions for the Gaussian location model						
NW	$1.3T^{-1/2}$	fixed-b (nonstandard)	No	0.054	0.058	0.067
EWP	$1.95T^{-2/3}$	fixed-b (t_v)	No	0.052	0.056	0.073

Outline

HAC = Heteroskedasticity- and Autocorrelation-Consistent

HAR = Heteroskedasticity- and Autocorrelation-Robust

1) The HAR Problem and the long-run variance matrix Ω
2) PSD estimators of Ω
3) From MSE to size and power
4) Fixed-*b* critical values
5) Size-power tradeoff
6) Choice of kernel and bandwidth
7) Monte Carlo results
8) Summary

1) The HAR Problem

The task: valid inference on β when X_t and u_t are possibly serially correlated:

$$Y_t = X_t'\beta + u_t,\ E(u_t|X_t) = 0,\ t = 1,\ldots, T$$

Asymptotic distribution of OLS estimator:

$$\sqrt{T}(\hat{\beta}-\beta) = \left(\frac{1}{T}\sum_{t=1}^{T} X_t X_t'\right)^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} X_t u_t\right)$$

Assume throughout that WLLN and CLT hold:

$$\frac{1}{T}\sum_{t=1}^{T} X_t X_t' \xrightarrow{p} \Sigma_{XX} \text{ and } \frac{1}{\sqrt{T}}\sum_{t=1}^{T} X_t u_t \xrightarrow{d} \mathrm{N}(0, \Omega),$$

so $$\sqrt{T}(\hat{\beta}-\beta) \xrightarrow{d} N\left(0, \Sigma_{XX}^{-1}\Omega\Sigma_{XX}^{-1}\right).$$

Σ_{XX} is easy to estimate, but what is Ω and how should it be estimated?

Some OLS situations in which HAR SEs are needed

Distributed lag regressions: $Y_t = \alpha + \beta(\mathrm{L})X_t + u_t$ where $E\left(u_t \mid X_t, X_{t-1},...\right) = 0$

Multiperiod asset returns: $\Delta\ln(P_{t+h}/P_t) = \alpha + \beta X_t + u_{t+h}^{(h)}$, e.g. X_t = dividend yield$_t$

Multiperiod-ahead forecasts: $y_{t+h} = \alpha + \beta X_t + \gamma(\mathrm{L})Y_t + u_{t+h}^{(h)}$

Local projections: $y_{t+h} = \alpha + \beta X_t + \gamma' W_t + u_{t+h}^{(h)}$, W_t = control variables

- In all these cases, u_t and X_t are serially correlated and the regression exogeneity condition holds or is assumed to hold, i.e.
  $E\left(u_t \mid X_t, X_{t-1},...\right) = 0$ (weak exogeneity: past, or past and present)

- GLS can't be used in any of these settings because X_t is weakly exogeneous but not strictly exogenous, i.e.
  $E\left(u_t \mid ...X_{t+1}, X_t, X_{t-1},...\right) \neq 0$ (past, present, and future)

Ω: The Long-Run Variance of $X_t u_t$

Let $Z_t = X_t u_t$. Note that $EZ_t = 0$ (because $E(u_t|X_t) = 0$). Suppose Z_t is second order stationary. Then

$$\begin{aligned}
\Omega_T &= \text{var}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} Z_t\right) = E\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} Z_t\right)^2 \\
&= \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T} E\left(Z_t Z_s'\right) \\
&= \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T} \Gamma_{t-s} \quad (Z_t \text{ is second order stationary}) \\
&= \frac{1}{T}\sum_{j=-(T-1)}^{T-1} \left(T - |j|\right)\Gamma_{t-s} \quad (\text{adding along the diagonals}) \\
&= \sum_{j=-(T-1)}^{T-1} \left(1 - \left|\frac{j}{T}\right|\right)\Gamma_j \to \sum_{j=-\infty}^{\infty} \Gamma_j
\end{aligned}$$

so

$$\Omega = \sum_{j=-\infty}^{\infty} \Gamma_j = 2\pi S_Z(0) \quad \left(\text{recall that } S_Z(\omega) = \frac{1}{2\pi}\sum_{j=-\infty}^{\infty} \Gamma_j e^{-i\omega j}\right)$$

Standard approach (Newey-West/Andrews): 3 elements

1) Bartlett kernel (triangle weight function)

Newey-West estimator: declining average of sample autocovariances

$$\hat{\Omega}^{NW} = \sum_{j=-S}^{S} k\left(\left|\frac{j}{S}\right|\right)\hat{\Gamma}_j, \text{ where } k(u) = (1 - |u|) \text{ (Bartlett kernel)}$$

where $\hat{\Gamma}_j = \frac{1}{T}\sum_{t=1}^{T}\hat{Z}_t\hat{Z}_{t-j}{}'$, where $\hat{Z}_t = X_t\hat{u}_t$.

2) Truncation parameter: Andrews (1991) minimum MSE

- Andrews truncation parameter: $S = S_T = .75T^{1/3}$ (e.g. Stock and Watson, *Introduction to Econometrics*, 4th edition, equation (16.17)).
- Expressed as a fraction of the sample size: $\boldsymbol{b} = S/T = 0.75T^{-2/3}$

3) Critical values: Normal/chi-squared

First order asymptotics imply that $\hat{\Omega}^{NW} \xrightarrow{p} \Omega$ so that $t \xrightarrow{d} N(0,1)$.

What's new? A lot.

1) **Bartlett kernel, or maybe equal-weighted periodogram**

Equal-weighted periodogram is the simple average of the first $B/2$ periodogram ordinates (details later)

2) **Truncation parameter: Balance size-power tradeoff for HAR test**

To reduce size distortions, use a larger truncation parameter:

$b = 1.3T^{1/2}$ (Proposed rule for NW kernel)

3) **Critical values: Fixed *b***

- The larger bandwidth induces sampling variability in $\hat{\Omega}$
- This sampling variability is resolved (to higher order asymptotically) by using fixed-*b* asymptotics (Kiefer-Vogelsang (2005)) – fixed *b* supposes that *S* increases proportionally to *T*, i.e. explicitly treats *S* as large.
- For the EWP estimator, fixed-*b* critical values are t_B (scalar case)

2) PSD Estimators of Ω

Estimation of $\Omega = \sum_{j=-\infty}^{\infty} \Gamma_j$ is hard: the sum needs truncation.

- Sum-of-covariances kernel estimator: $\hat{\Omega}^{sc} = \sum_{j=-S}^{S} k\left(\left|\frac{j}{S}\right|\right)\hat{\Gamma}_j$
- Weighted periodogram estimator: $\hat{\Omega}^{wp} = 2\pi \sum_{j=-(T-1)}^{T-1} K(j/B) I_{\hat{Z}\hat{Z}}(2\pi j/T)$

where $I_{ZZ}(\omega) = \frac{1}{2\pi} d_Z(\omega)\overline{d_Z(\omega)}'$ and $d_Z(\omega) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} Z_t e^{-i\omega t}$

Time domain/frequency domain duality:

- All sum-of-covariances estimators have a kernel representation and vice versa (frequency domain/time domain duality).
- Small $b \leftrightarrow$ large B (for EWP, $b = B^{-1}$)
- PSD theorem: kernel estimators are psd w.p. 1 if & only if their frequency domain weights are nonnegative ($K_T(l) \geq 0$ for all l)

3) From MSE to size and power

$$\hat{\Omega}^{SC} = \sum_{j=-S}^{S} k\left(\left|\frac{j}{S}\right|\right)\hat{\Gamma}_j$$

Overarching questions: What kernel k? What value of S, given k?

Historical approach: minimize MSE($\hat{\Omega}^{SC}$) (delivers $S = 0.75T^{1/3}$)

- Early history of spectral estimation, applied to the HAR problem:
  Grenander (1951), Parzen (1969), Epanechnikov (1969); Brillinger (1975), Priestley (1981); Andrews (1991)

Problems with the NW/Andrews paradigm: Early MCs

- Den Haan & Levin 1994

From MSE to size and power, ctd.

New view of HAR inference: focus on size and power

- Size control through fixed-*b* critical values.

  Kiefer-Vogelsang-Bunzel (2000), **Kiefer-Vogelsang (2005)**, Sun (2014)

- Study size and power using Edgeworth expansions in the Gaussian location model.

  Velasco and Robinson (2001), Jansson (2004), Sun, Phillips, Jin (2008), Sun (2013, 2014a, b), Lazarus, Lewis, & Stock (2018), Lazarus, Lewis, Stock, & Watson (2018) (with discussion, *JBES*)

4) Fixed-*b* critical values

It is easiest to understand fixed-*b* critical values by looking at the Equal Weighted Periodogram (EWP) estimator in the scalar case.

$$\hat{\Omega}^{EWP} = 2\pi \frac{1}{B/2}\sum_{j=1}^{B/2} I_{\hat{Z}\hat{Z}}(2\pi j/T) = \frac{1}{B/2}\sum_{j=1}^{B/2}\left\| \frac{1}{\sqrt{T}}\sum_{t=1}^{T} Z_t e^{-i2\pi jt/T}\right\|^2$$

$$= \frac{1}{B/2}\sum_{j=1}^{B/2}\left\| \frac{1}{\sqrt{T}}\sum_{t=1}^{T} Z_t \left[\cos(2\pi jt/T) + i\sin(2\pi jt/T)\right]\right\|^2$$

$$\approx \frac{1}{B}\left\{\sum_{j=1}^{B/2}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} Z_t\sqrt{2}\cos(2\pi jt/T)\right]^2 + \sum_{j=1}^{B/2}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} Z_t\sqrt{2}\sin(2\pi jt/T)\right]^2\right\}$$

$$\xrightarrow{d} \left(\chi_B^2 / B\right)\Omega$$

so $$t^{EWP} = \frac{\bar{Z}}{\sqrt{\hat{\Omega}^{EWP}}} \xrightarrow{d} \frac{N(0,1)}{\sqrt{\chi_B^2 / B}} \sim t_B \text{ for } B \text{ fixed.}$$

Recall that for EWP, $b = B^{-1}$; so fixed B is fixed b. So, (*) is the fixed B distribution of the EWP test statistic.

- This result dates to Brillinger (1975, exercise 5.13.25).

General fixed-*b* distribution

Recall that time-domain kernels can be represented in the frequency domain. This motivates the general fixed-*b* distribution (e.g. Sun (2014)),

$$t^{SC} = \frac{\bar{Z}}{\sqrt{\hat{\Omega}^{SC}}} \xrightarrow{d} \frac{z}{\sqrt{\sum_{j=1}^{\infty} w_j \xi_j}}, \qquad (**)$$

where z ~ N(0,1) and ξ_j i.i.d. χ_1^2, and $z \perp \{\xi_j\}$ [(**) is exact for Z Gaussian]

- There is a time-domain representation in terms of weighted Brownian Bridges – more frequently used in papers, but less intuitive than (**).
- The distribution (**) is *t*-like and can be approximated by t_v, where

  $v = \left(b\int_{-\infty}^{\infty} k^2(x)dx\right)^{-1}$ = Tukey's (1949) approximate degrees-of-freedom
- For EWP, $\int_{-\infty}^{\infty} k^2(x)dx = 1$ and $v = B = b^{-1}$.

5) Size-power tradeoff

Setup & theory: Notation

- **Spectral curvature**:

$$\omega^{(q)} = \sum_{j=-\infty}^{\infty} |j|^q \, \Gamma_j \Omega^{-1} \text{ (scalar case)}$$

$$\omega^{(2)} = -S_z''(0) / S_z(0)$$

AR(1) case: $\omega^{(1)} = 2\rho/(1-\rho^2)$ $\omega^{(2)} = 2\rho/(1-\rho)^2$.

- **Parzen characteristic exponent (q)** and generalized derivative:

$k^{(q)}(0) = \lim_{x\to 0} \dfrac{1-k(x)}{|x|^q}$, where k = kernel

q = Parzen characteristic exponent = max q: $k^{(q)}(0) < \infty$

- The Edgeworth expressions are derived for the Gaussian location model – but they are a guide (we hope) to non-Gaussian location and regression.

Sketch of asymptotic expansion of size distortion

For details see Velasco and Robinson (2001), Sun, Phillips, and Jin (2008)

Consider the Gaussian location model, $y_t = \beta + u_t$, u_t stationary, Gaussian

Then $Z_t = X_t u_t = u_t$ so the test statistic is, $W_T = \dfrac{\left(T^{-1/2}\sum_1^T Z_t\right)^2}{\hat{\Omega}}$.

The probability of 2-sided rejection under the null thus is,

$$\Pr[|t| < c] = \Pr\left[\frac{\left(T^{-1/2}\sum_1^T Z_t\right)^2}{\hat{\Omega}} < c\right]$$

where c is the asymptotic critical value (3.84 for a 5% test). The size distortion is obtained by expanding this probability…

Under Gaussianity, $T^{-1/2}\sum_1^T Z_t$ and $\hat{\Omega}$ are asymptotically independent. Now

$$\Pr[W_T < c] = \Pr\left[\frac{\left(T^{-1/2}\sum_1^T Z_t\right)^2}{\hat{\Omega}} < c\right] = \Pr\left[\frac{\left(T^{-1/2}\sum_1^T Z_t\right)^2}{\Omega} < c\frac{\hat{\Omega}}{\Omega}\right]$$

$$= E\left\{\Pr\left[\frac{\left(T^{-1/2}\sum_1^T Z_t\right)^2}{\Omega} < c\frac{\hat{\Omega}}{\Omega}\right] \middle| \hat{\Omega}\right\}$$

$$\approx E\left[F\left(c\frac{\hat{\Omega}}{\Omega}\right)\right],\text{ where } F = \text{chi-squared c.d.f}$$

$$= E\left[F(c) + cF'(c)\left(\frac{\hat{\Omega}-\Omega}{\Omega}\right) + \frac{1}{2}cF''(c)\left(\frac{\hat{\Omega}-\Omega}{\Omega}\right)^2 + \ldots\right]$$

so the size distortion approximation is,

$$\Pr[W_T < c] - F(c) \approx cF'(c)\frac{bias(\hat{\Omega})}{\Omega} + \frac{1}{2}cF''(c)\frac{\mathrm{var}(\hat{\Omega})}{\Omega^2} + \text{smaller terms}$$

Expressions for bias and variance for small *b*

Bias. Use frequency domain representation, scalar case – for time-domain kernel with two derivatives at origin (QS, EWP, not Bartlett):

$$E\hat{\Omega} = \sum_{j=1}^{M} K\left(\frac{j}{M}\right) E\left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \hat{Z}_t e^{-i2\pi jt/T} \right\|^2$$

$$= 2\pi \sum_{j=1}^{M} K\left(\frac{j}{M}\right) s_z(2\pi j/T)$$

$$\approx 2\pi \sum_{j=1}^{M} K\left(\frac{j}{M}\right)\left[s_z(0) + (2\pi j/T) s_z{}'(0) + \frac{1}{2}(2\pi j/T)^2 s_z{}''(0) \right]$$

$$= \Omega + \frac{1}{2}\sum_{j=1}^{M} K\left(\frac{j}{M}\right)(2\pi j/T)^2 \frac{s_z{}''(0)}{s_z(0)}[2\pi s_z(0)]$$

$$= \Omega\left\{ 1 + 2\pi^2 \omega^{(2)} \left(\frac{M}{T}\right)^2 \sum_{j=1}^{M} \left(\frac{j}{M}\right)^2 K\left(\frac{j}{M}\right) \right\}$$

$$\to \Omega\left[1 + \left(2\pi^2 \omega^{(2)} \int_0^1 u^2 K(u)\,du \right) M^2 T^{-2} \right]$$

Expressions for bias and variance for small *b*, ctd.

Variance.

$$\operatorname{var}(\hat{\Omega}) = \operatorname{var}\left[\sum_{j=1}^{M} K\left(\frac{j}{M}\right)\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\hat{Z}_t e^{-i2\pi jt/T}\right\|^2\right]$$

$$\approx \sum_{j=1}^{M} K^2\left(\frac{j}{M}\right)\operatorname{var}\left(\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\hat{Z}_t e^{-i2\pi jt/T}\right\|^2\right)$$

$$\approx \sum_{j=1}^{M} K^2\left(\frac{j}{M}\right)\operatorname{var}\left(\chi_2^2 / 2\right)\Omega^2$$

$$= M^{-1}\Omega^2\left[M\sum_{j=1}^{M} K^2\left(\frac{j}{M}\right)\right] \to M^{-1}\Omega^2\int_0^1 K^2(u)\,du$$

Rejection rate expansion, bias, and variance: summary

$$\Pr[W_T < c] - F(c) \approx cF'(c)\frac{bias(\hat{\Omega})}{\Omega} + \frac{1}{2}cF''(c)\frac{\mathrm{var}(\hat{\Omega})}{\Omega^2} + \text{smaller terms}$$

$$\frac{bias(\hat{\Omega})}{\Omega^2} \approx 1 + \left(2\pi^2\omega^{(2)}\int_0^1 u^2 K(u)du\right)M^2 T^{-2}$$

$$\frac{\mathrm{var}(\hat{\Omega})}{\Omega^2} \approx \int_0^1 K^2(u)du M^{-1}$$

<u>Comments</u>

1) Controlling size (using normal critical values) places more emphasis on bias reduction than minimizing MSE → larger S (larger b/ smaller B)

2) The second term, $\frac{1}{2}cF''(c)\frac{\mathrm{var}(\hat{\Omega})}{\Omega^2}$, depends on b and the kernel, but not on the time series properties of z. This term thus can be used to provide a higher-order correction to the critical values.

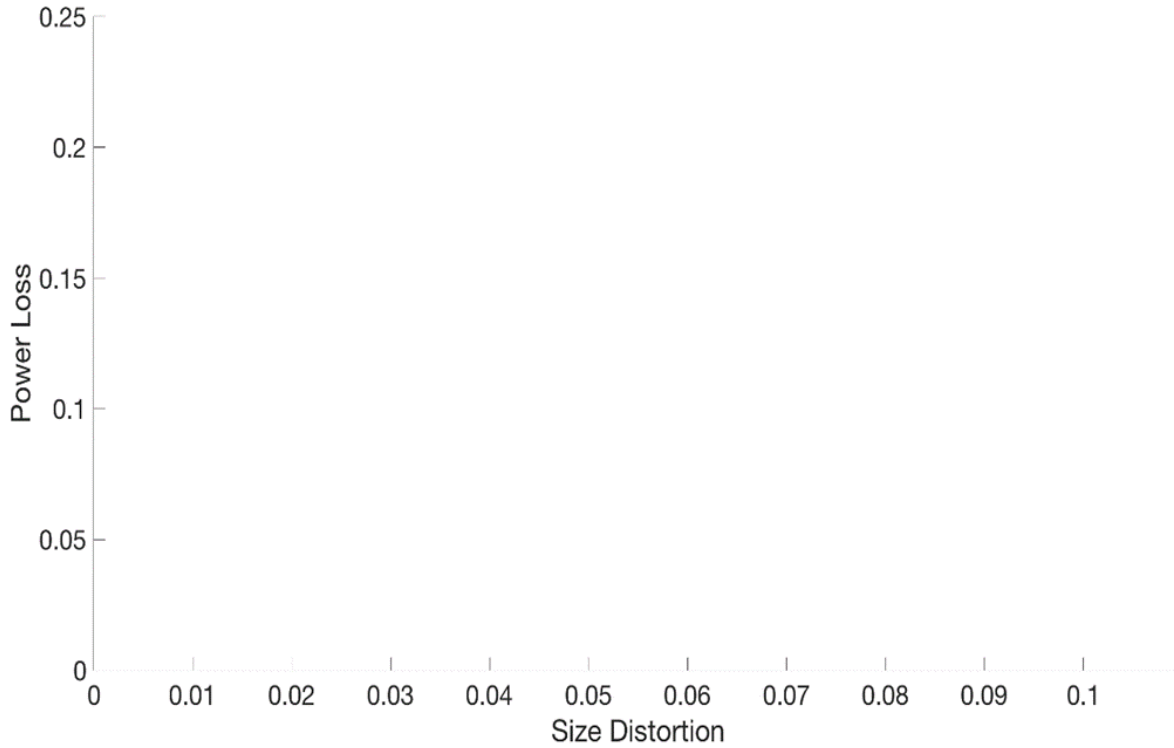
- In i.i.d. normal means, this term approximates t critical values
- In the HAR problem, Jansson (2004) showed that using fixed-b critical values eliminate this term.

The size-power tradeoff

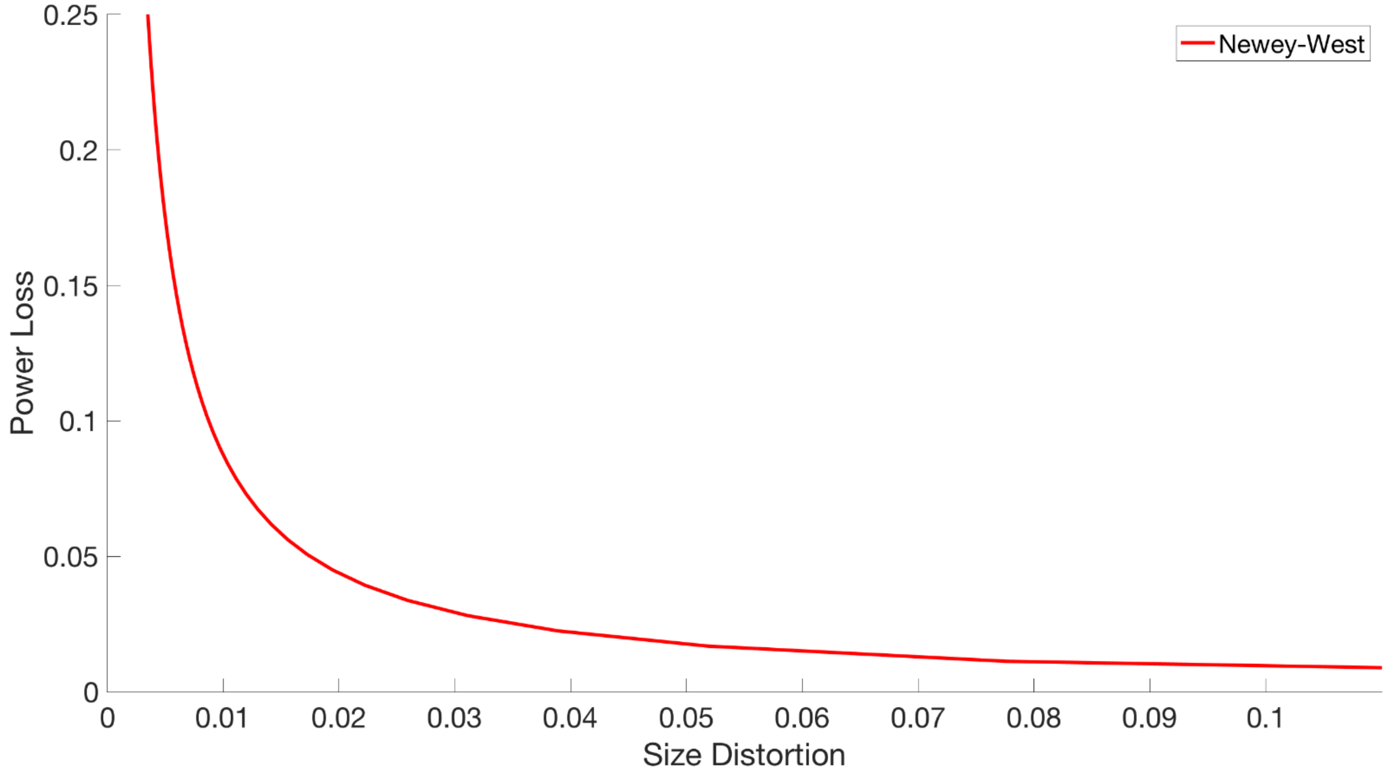
Intuition: Using larger S reduces bias, but it increases variance of $\hat{\Omega}^{SC}$.

- Bias results in size distortion: the estimator, and thus the test statistic, is centered at the wrong place
- Variance results in power loss – like using t-inference with a small d.f.

Kiefer and Vogelsang (2005) speculated that there must be a way to use the Edgeworth expansions to construct a size-power tradeoff using fixed-b critical values.



They were right!



Size-power tradeoff, NW, AR(1), ρ_u = 0.7, *T* = 200.

Key insight: using fixed-*b* critical values:

- Size depends on the bias
- Power loss depends on the variance (e.g., *t* degrees of freedom, or *v*)
    - Here, power = size-adjusted power (as usual)

The tradeoff is obtained using Edgeworth expansion …

Note: the expressions for bias and variance used here use the time-domain kernel, where $K(\omega) = (2\pi)^{-1}\int_{-\infty}^{\infty} k(u)e^{-i\omega u}du$

Under null:

$$\Pr_0\left[F_T^* > c_m^\alpha(b)\right] = \alpha + G_m'(\chi_m^\alpha)\chi_m^\alpha \omega^{(q)}k^{(q)}(0)(bT)^{-q} + o(b) + o\left((bT)^{-q}\right)$$

The use of fixed-*b* critical values has eliminated the leading term in v^{-1}

Size adjusted critical value:

$$c_{m,T}^\alpha(b) = c_m^\alpha(b) + d_{m,T} = \left[1+\omega^{(q)}k^{(q)}(0)(bT)^{-q}\right]c_m^\alpha(b)$$

where

v = $\left(b\int_{-\infty}^{\infty} k^2(x)dx\right)^{-1}$ = Tukey equivalent d.f.

F_T^* = HAR *F* statistic testing the *m* restrictions, $R\beta = r_0$, in "*F*" form

$c_m^\alpha(b)$ = fixed-*b* asymptotic critical value for level α test using F_T^*

G_{m,δ^2} = noncentral chi-squared cdf with *m* df & noncentrality parm δ^2

χ_m^α = chi-squared *m* critical value for test of level α

Under a standardized local alternative δ:

$$\Pr_{\delta}\left[F_T^* > c_m^{\alpha}(b)\right] = [1-G_{m,\delta^2}(\chi_m^{\alpha})]$$ **Power of oracle test against δ**

$$+G'_{m,\delta^2}(\chi_m^{\alpha})\chi_m^{\alpha}\omega^{(q)}k^{(q)}(0)(bT)^{-q}$$ **Bias term inherited from size distortion under null, eliminated by using size-adjusted critical values**

$$-\frac{1}{2}\delta^2 G'_{(m+2),\delta^2}(\chi_m^{\alpha})\chi_m^{\alpha}\nu^{-1}$$ **Power loss from using *t*-like inference**

$$+o(b)+o\left((bT)^{-q}\right)+O(\log T/\sqrt{T})$$

- Size adjusted critical value:
  $$c_{m,T}^{\alpha}(b) = c_m^{\alpha}(b)+d_{m,T} = \left[1+\omega^{(q)}k^{(q)}(0)(bT)^{-q}\right]c_m^{\alpha}(b)$$
- Power difference between two tests (“1” and “2”) with same second-order size depends only on $\nu_1^{-1}-\nu_2^{-1}$.

Size/power tradeoff for given kernel

Size distortion:

$$(1)\ \Delta_S = \Pr_0\left[F_T^* > c_m^\alpha(b)\right] - \alpha \approx G'_{m,\delta^2}(\chi_m^\alpha)\chi_m^\alpha \omega^{(q)} k^{(q)}(0)(bT)^{-q}$$

Maximum (size-adjusted) power loss:

$$(2)\ \Delta_p^{\max} = \max_\delta [1 - G_{m,\delta^2}(\chi_m^\alpha)] - \Pr_\delta\left[F_T^* > c_{m,T}^\alpha(b)\right] \approx \frac{1}{2}\left[\max_\delta \delta^2 G'_{(m+2),\delta^2}(\chi_m^\alpha)\chi_m^\alpha\right]\nu^{-1}$$

Because $\nu = \left(b\int_{-\infty}^{\infty} k^2(x)dx\right)^{-1}$ (1) and (2) are parametric equations in b that map out the size/size-adjusted power tradeoff for a given kernel/implied mean kernel:

$$\Delta_p \left|\Delta_S\right|^{1/q} \approx \bar{a}_{m,\alpha,q}\left[\left(k^{(q)}(0)\right)^{1/q}\int_{-\infty}^{\infty} k^2(x)dx\right]\left|\omega^{(q)}\right|^{1/q} T^{-1}.$$

6) Choice of kernel and bandwidth

For a given kernel, the tradeoff is,

$$\Delta_p \left|\Delta_S\right|^{1/q} \approx \bar{a}_{m,\alpha,q}\left[\left(k^{(q)}(0)\right)^{1/q}\int_{-\infty}^{\infty} k^2(x)dx\right]\left|\omega^{(q)}\right|^{1/q} T^{-1}$$

- $q = 2$ dominates $q = 1$: $\Delta_p, \Delta_S = \begin{cases} T^{-1/2}, q=1 \\ T^{-2/3}, q=2 \end{cases}$
- So, the frontier is given by the psd kernel that maximizes $\sqrt{k^{(2)}(0)}\int_{-\infty}^{\infty} k^2(x)dx$
- This is the classic problem solved by the QS kernel, for which

$\sqrt{k^{(2)}(0)}\int_{-\infty}^{\infty} k^2(x)dx = 3\pi\sqrt{10}/25.$

- For $m = 1$, $\bar{a}_{1,.05,2} = .2825$, so for 5% tests of a single restriction,

$$\Delta_p \left|\Delta_S\right|^{1/q} \geq 0.2825\frac{3\pi\sqrt{10}}{25}\sqrt{\omega^{(2)}}T^{-1} = 0.3368\sqrt{\omega^{(2)}}T^{-1}$$

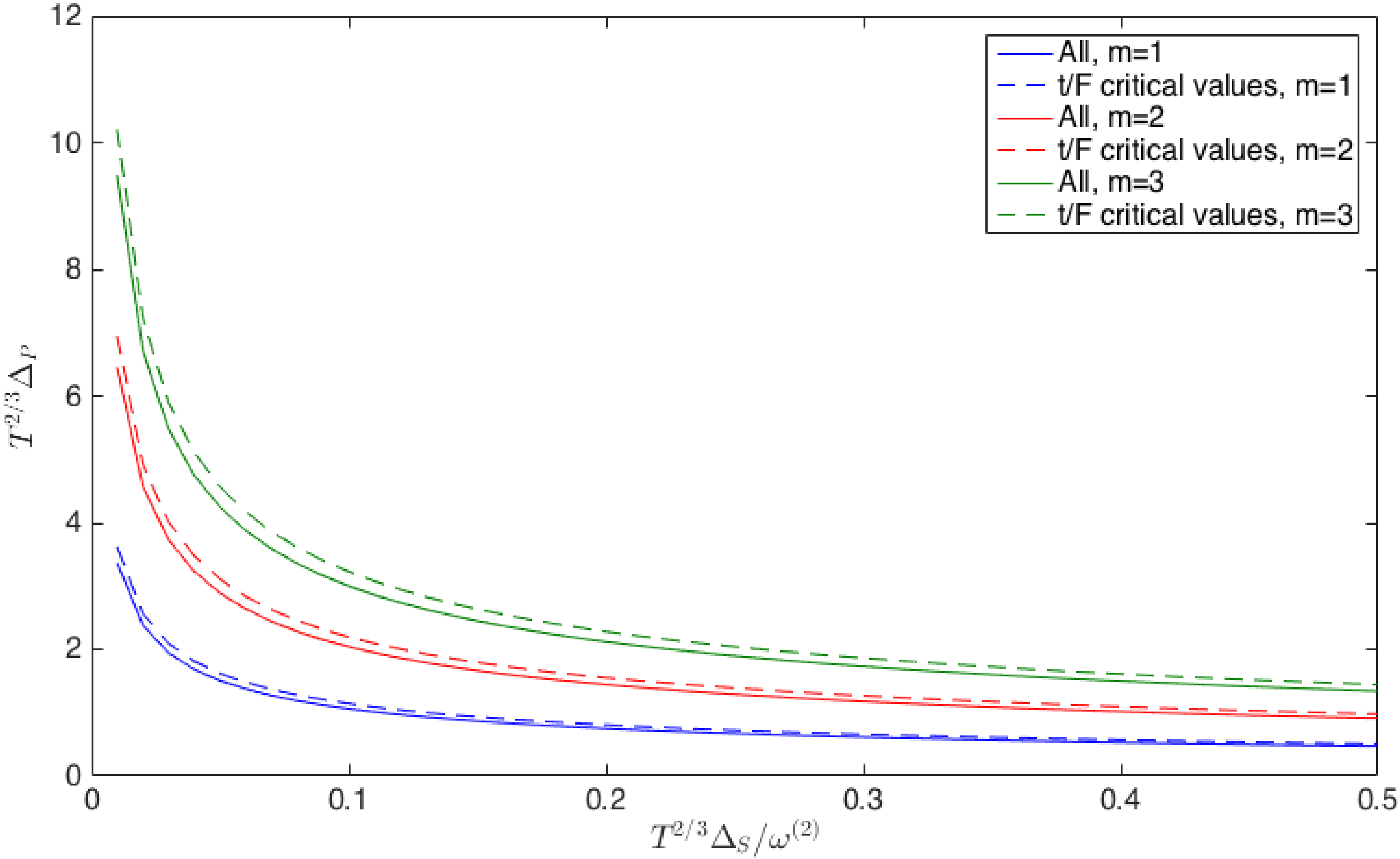
Achieving the frontier, t_ν fixed-b inference

- Lazarus, Lewis, and Stock show that the EWP test achieves the frontier for tests with exact t_ν fixed-b asymptotic critical values
- For EWP, $\int_{-\infty}^{\infty} k^2(x) = 1$ and $\sqrt{k^{(2)}(0)} = \pi / \sqrt{6}$, so

$$\Delta_p \sqrt{|\Delta_S|} \geq 0.2825 \frac{\pi}{\sqrt{6}} \sqrt{\omega^{(2)}} T^{-1} = 0.3624 \sqrt{\omega^{(2)}} T^{-1}$$

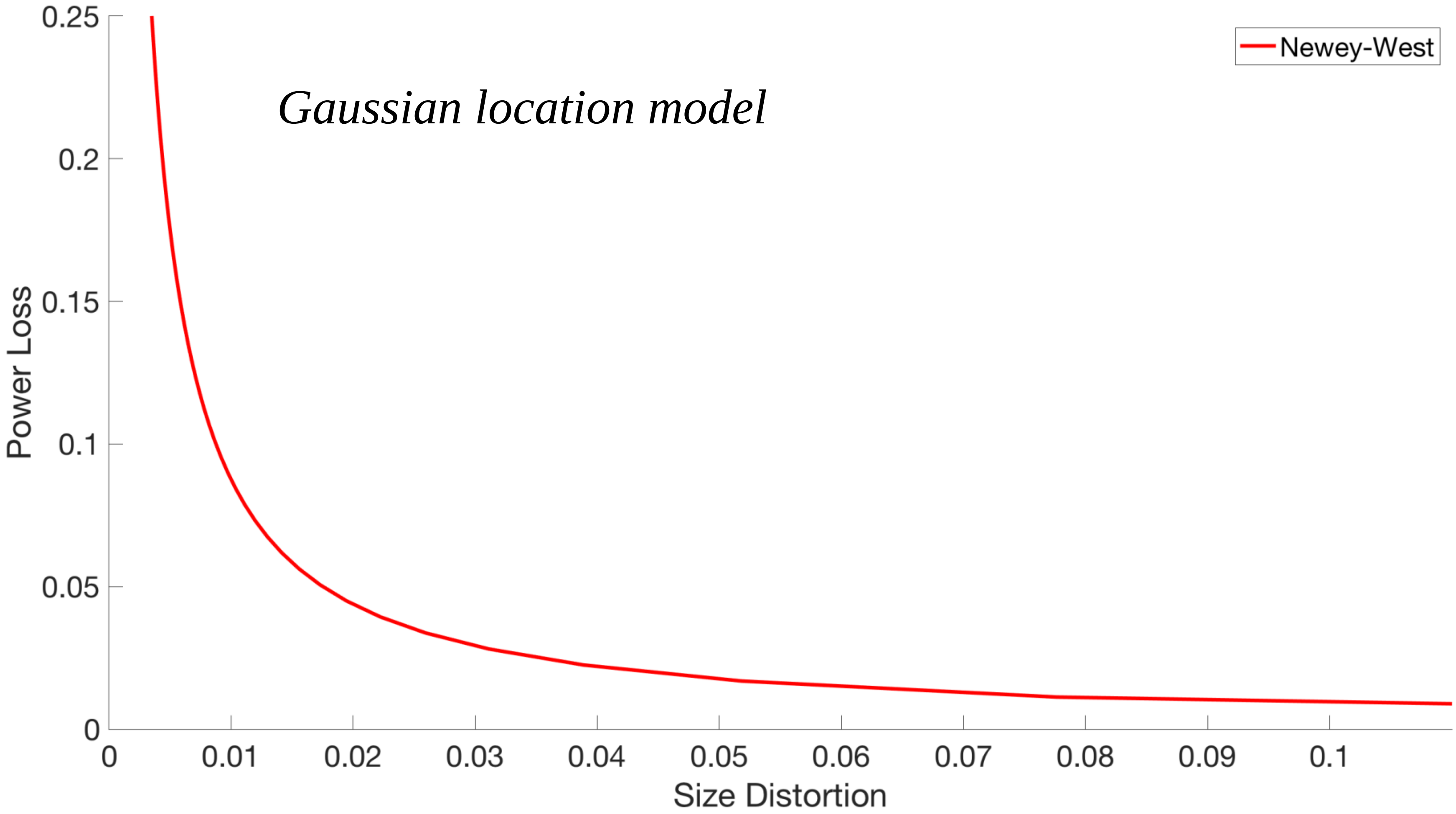
LLS summary, in pictures:

Frontier for psd kernel tests



- Vertical axis: $T^{2/3}\Delta_P$
- Horizontal axis: $T^{2/3}\Delta_S/\omega^{(2)}$
- Overall frontier solid, the fixed-b t and F frontier dashed
- These frontiers are universal: this figure covers all T & all stationary z_t

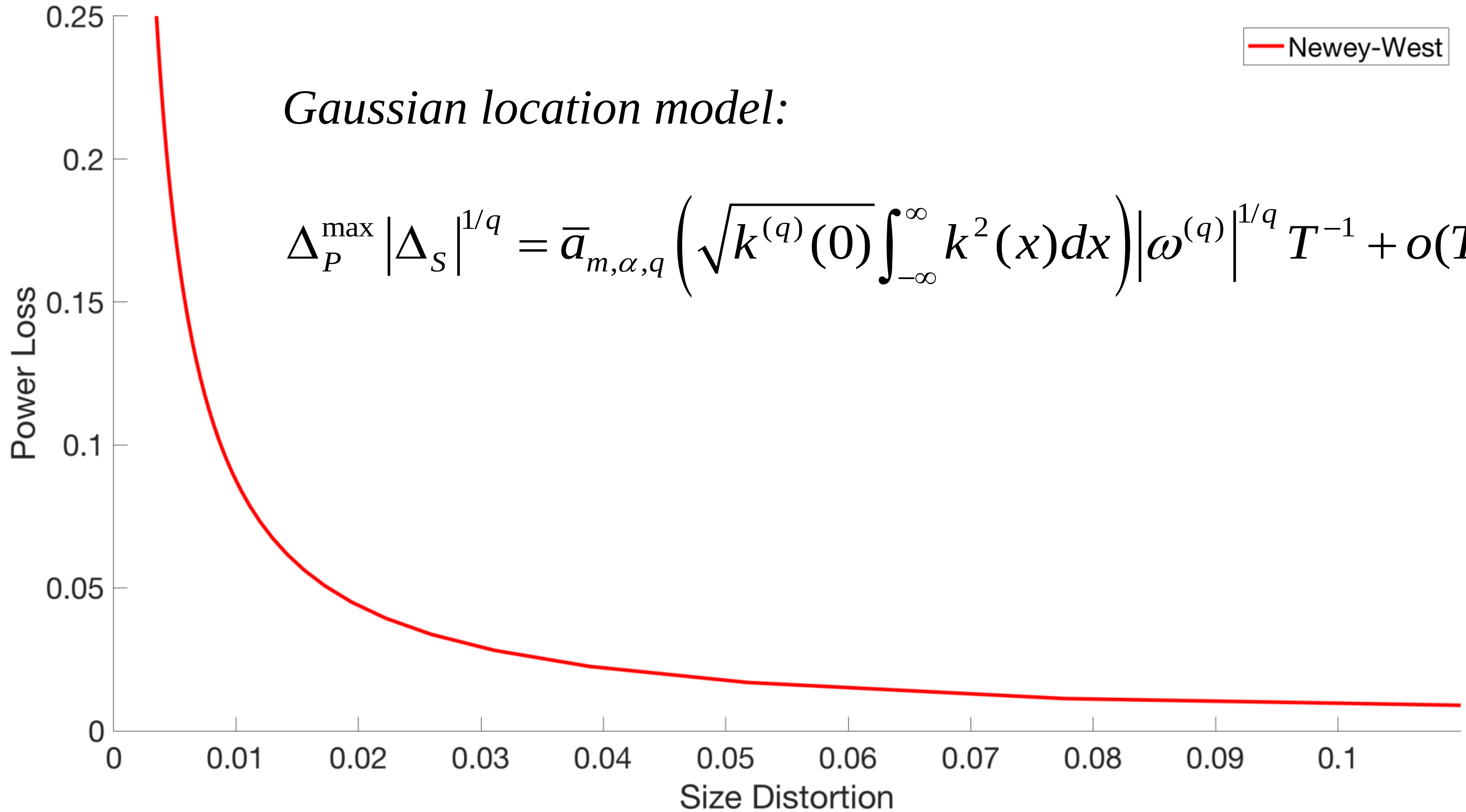
Theoretical size-power tradeoff, AR(1), $\rho_z = 0.7$, $T = 200$



x axis: Δ_S = |rejection rate| – 0.05 = size distortion

y axis: Δ_P^{max} = maximum power loss, compared to oracle (Ω known) test with same second-order size

Theoretical size-power tradeoff, AR(1), $\rho_z = 0.7$, $T = 200$



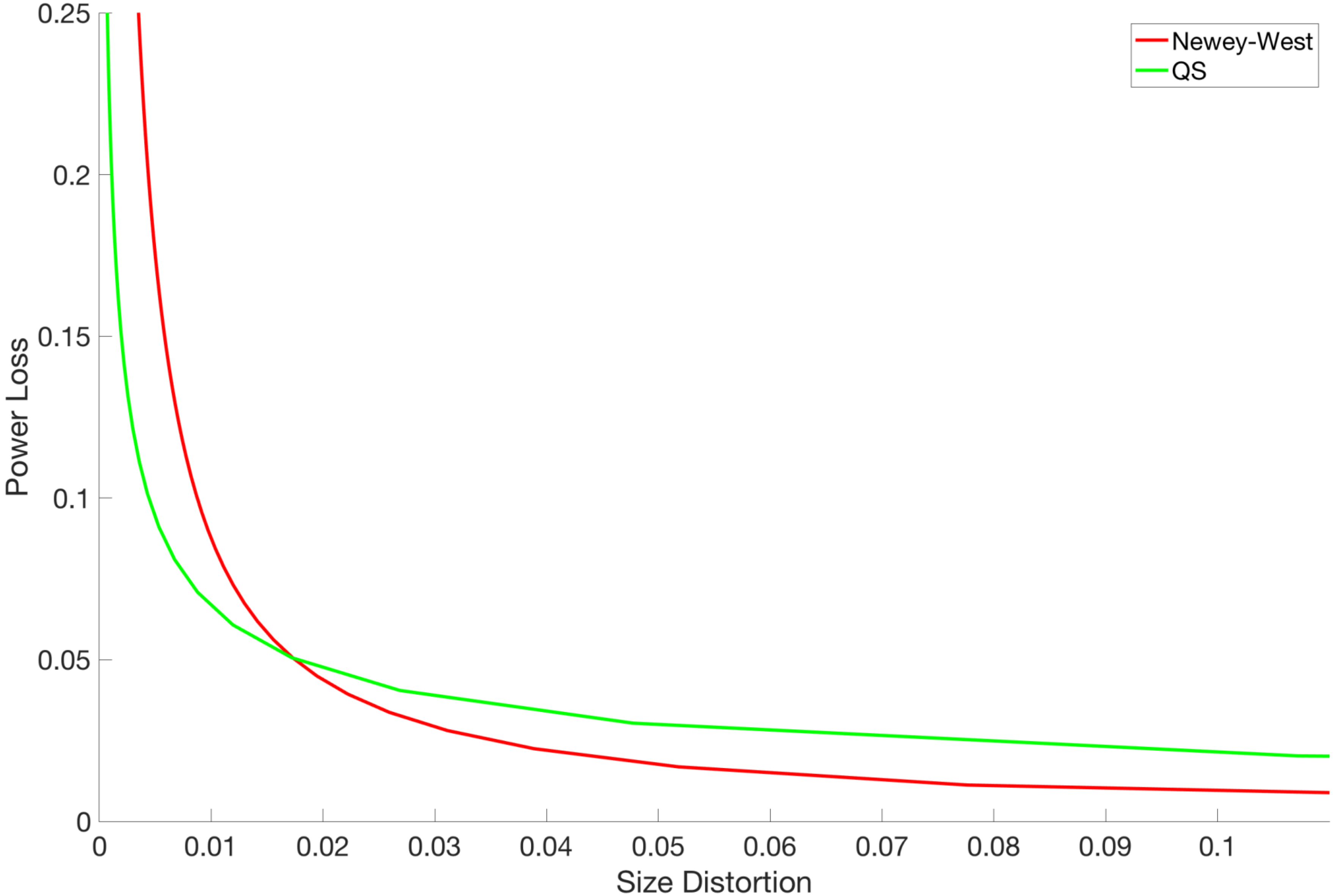
Gaussian location model:

$$\Delta_P^{\max}\left|\Delta_S\right|^{1/q} = \bar{a}_{m,\alpha,q}\left(\sqrt{k^{(q)}(0)}\int_{-\infty}^{\infty} k^2(x)dx\right)\left|\omega^{(q)}\right|^{1/q} T^{-1} + o(T^{-1})$$

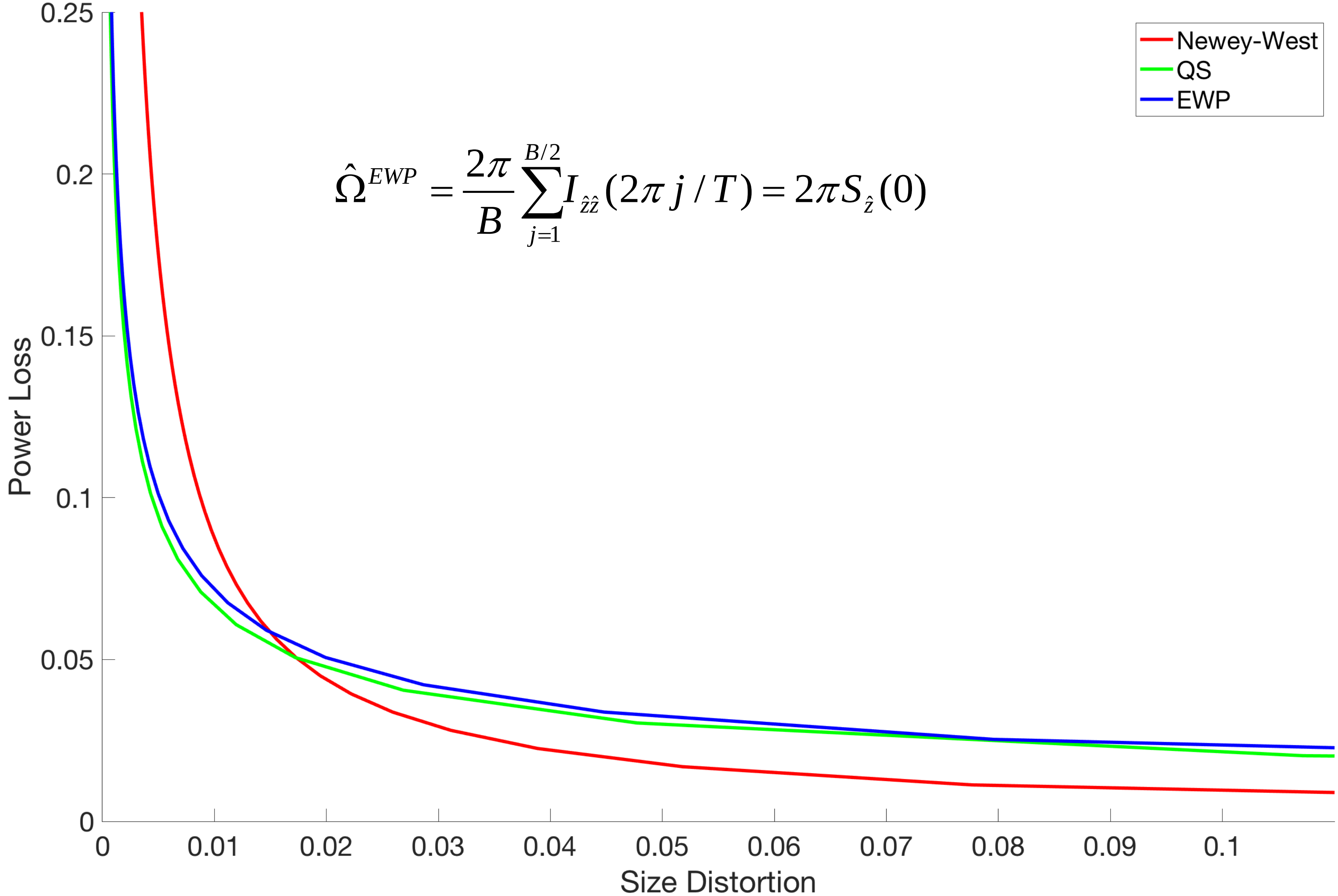
x axis: Δ_S = |rejection rate| – 0.05 = size distortion

y axis: $\Delta_P^{\max}$ = maximum power loss, compared to oracle (Ω known) test with same second-order size

…and for QS (Epanechnikov) kernel



…and for Equal-weighted periodogram (EWP) kernel



$$\hat{\Omega}^{EWP} = \frac{2\pi}{B}\sum_{j=1}^{B/2} I_{\hat{z}\hat{z}}(2\pi j / T) = 2\pi S_{\hat{z}}(0)$$

Summary: Asymptotic frontiers and NW tradeoff

LLS (2017) summary:

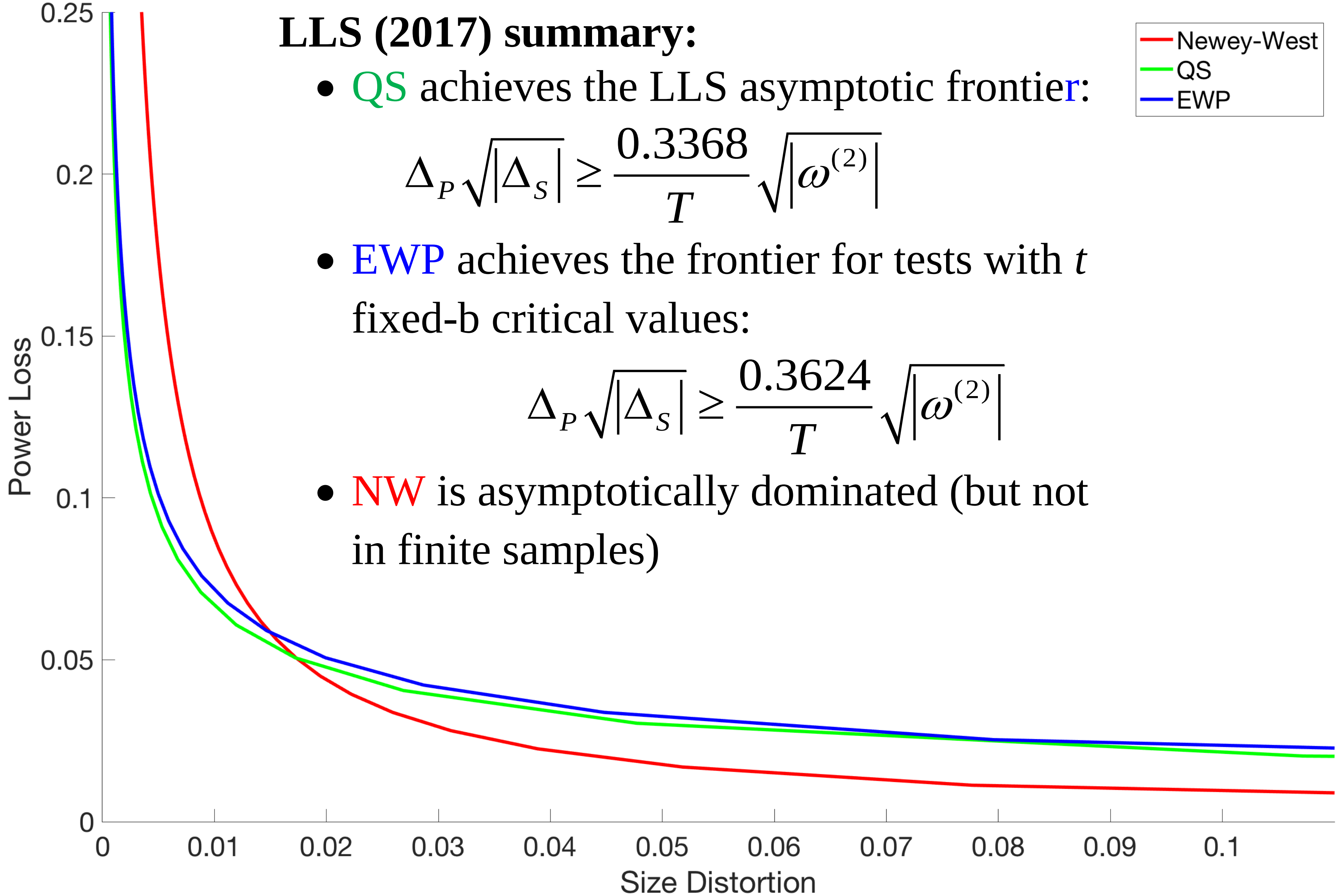
- QS achieves the LLS asymptotic frontier:

$$\Delta_P \sqrt{|\Delta_S|} \geq \frac{0.3368}{T} \sqrt{|\omega^{(2)}|}$$

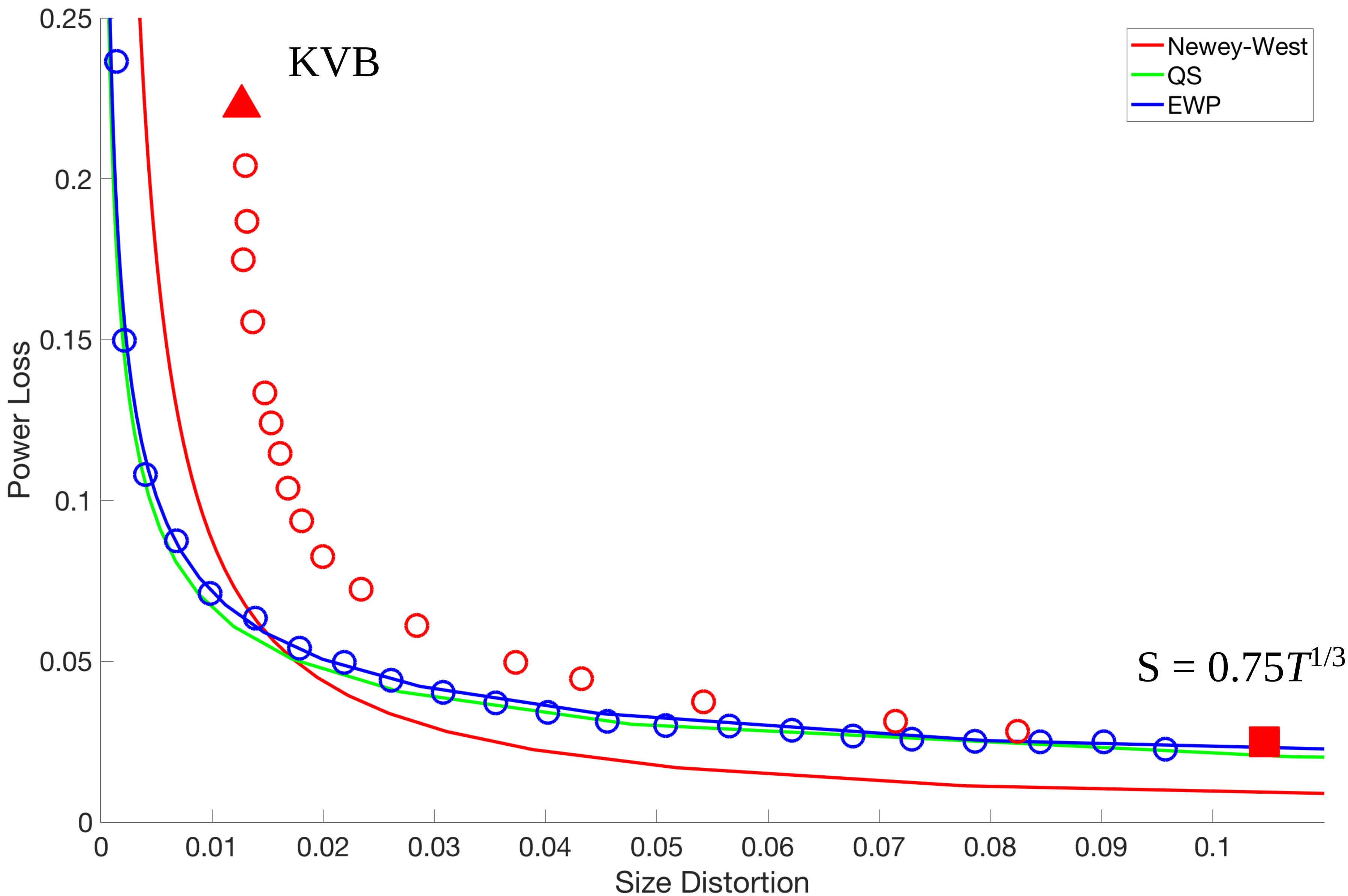
- EWP achieves the frontier for tests with *t* fixed-b critical values:

$$\Delta_P \sqrt{|\Delta_S|} \geq \frac{0.3624}{T} \sqrt{|\omega^{(2)}|}$$

- NW is asymptotically dominated (but not in finite samples)



How about finite sample performance? T = 200, AR(1), 0.7



Bandwidth rule?

Choosing a bandwidth entails choosing a point on the tradeoff curve, for a given kernel.

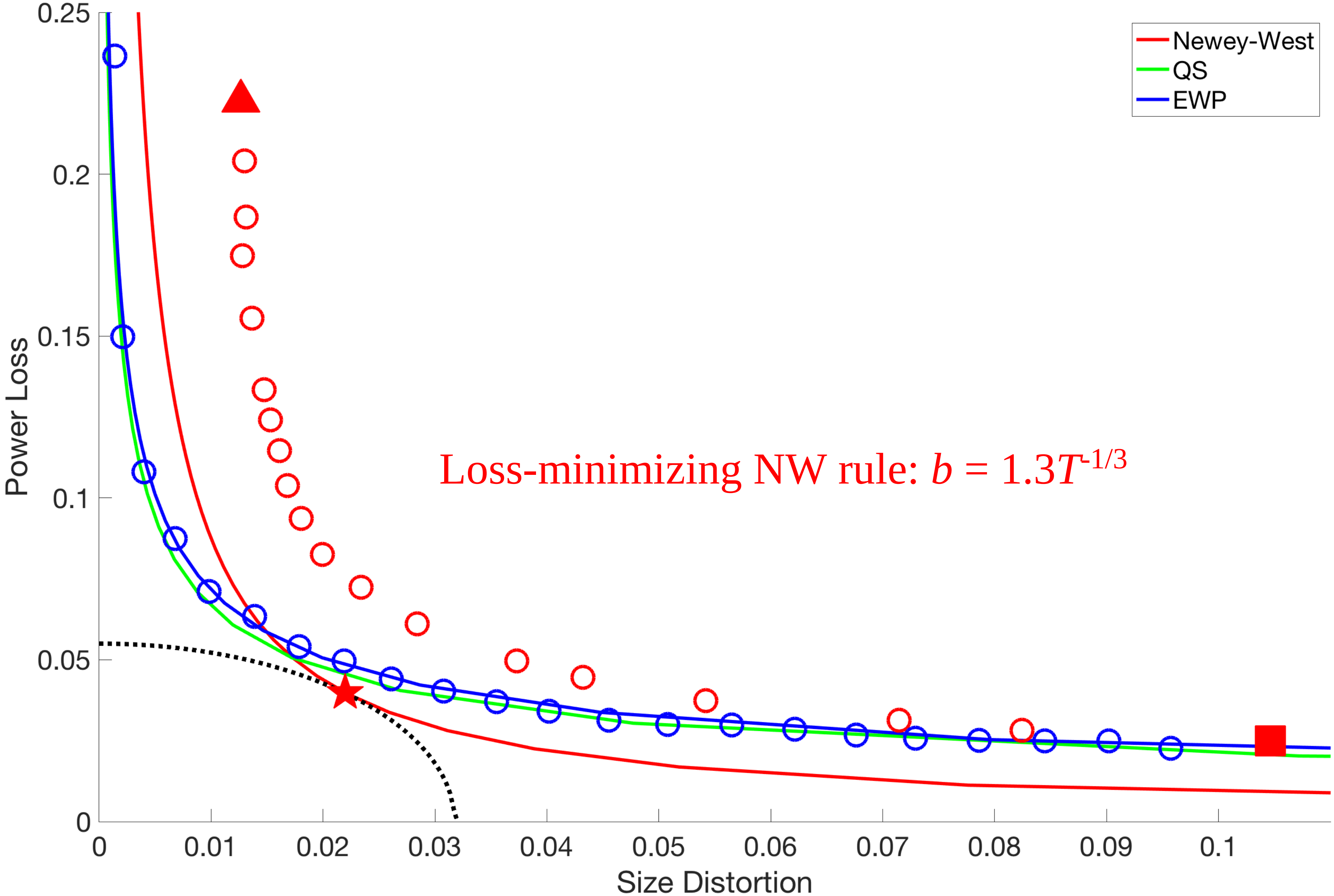
To make this choice, you need to make a decision – how in fact do you trade off size v. power loss?

LLSW propose minimizing the loss function,

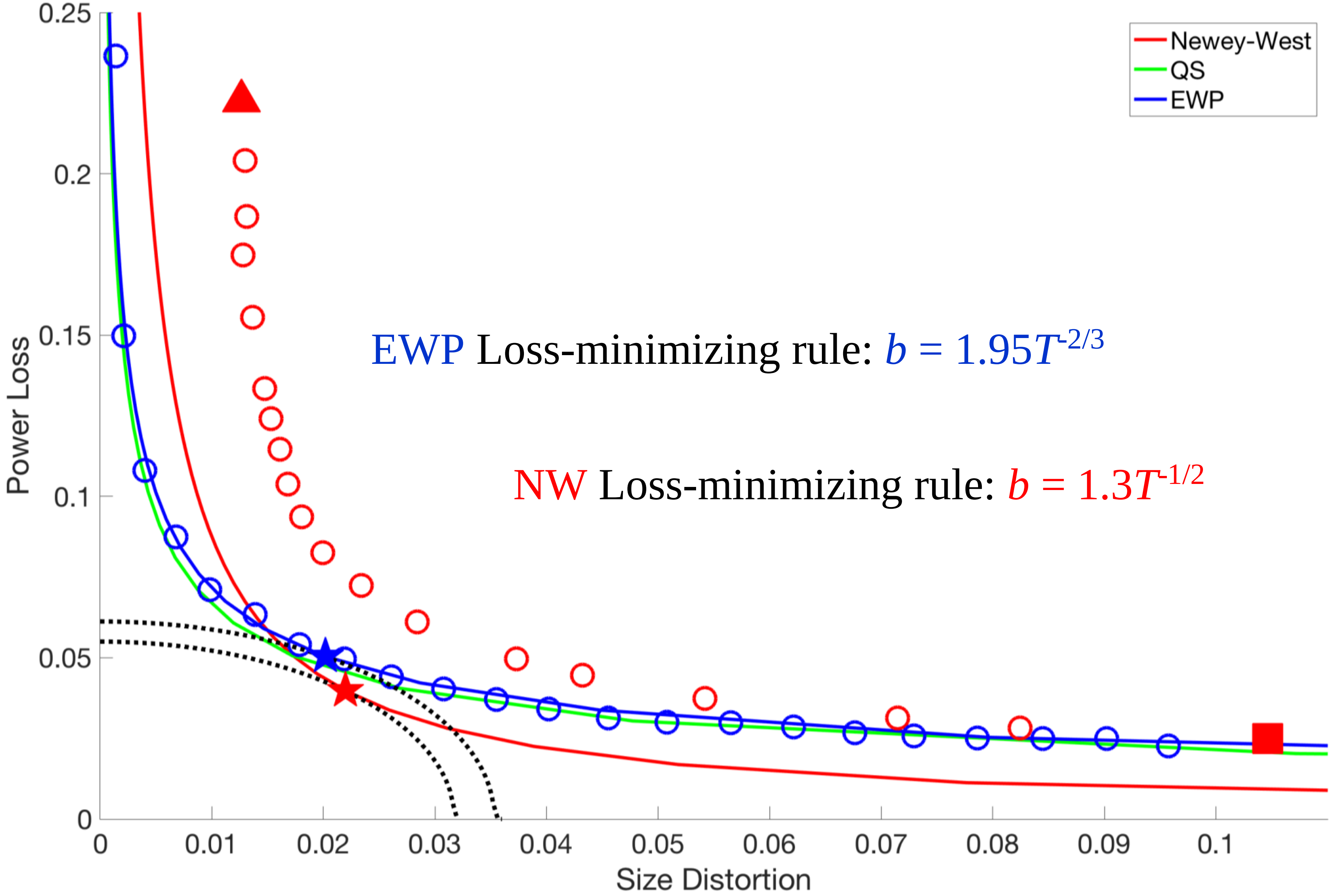
$$Loss = \kappa\left(\Delta_S\right)^2 + (1-\kappa)\left(\Delta_P^{\max}\right)^2.$$

with $\boldsymbol{\kappa} = \mathbf{0.9}$ (most concerned about size)
and $\boldsymbol{\rho} = \mathbf{0.7}$ (a large degree of persistence for most problems – Local projections, multistep ahead forecasts, distributed lags)

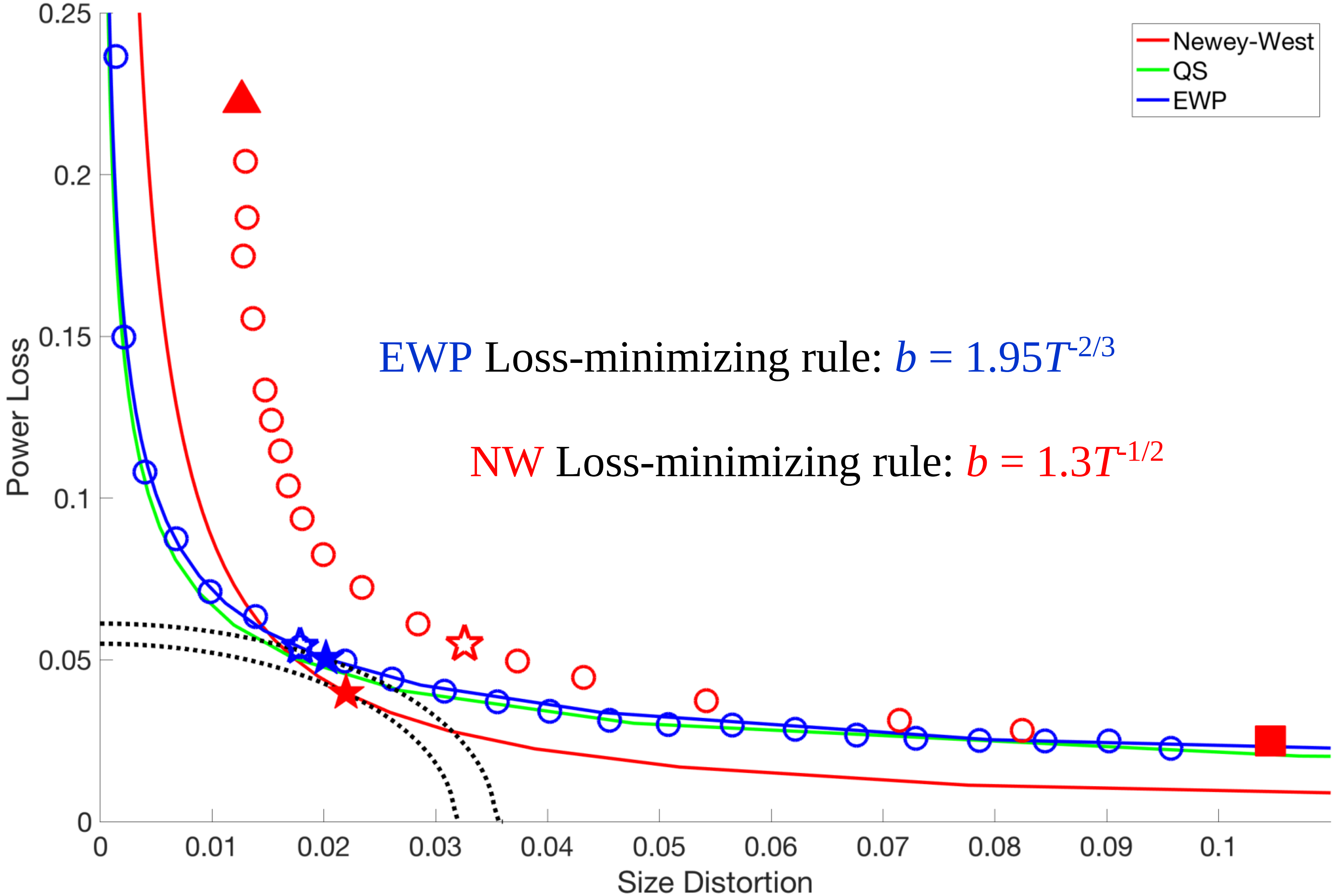
LLS propose using quadratic size/power loss to choose a point on the curve



… for both NW and EWP



How do they perform in finite samples? (open stars = MC)



7) Monte Carlo Results

LLSW Monte Carlo study with parametric models and data-based models (generate data from a DFM)

Main findings:

- In the location model
  - The approximations are good for QS and EWP, and OK for NW
  - Departures from normality change the location and shape of the frontier, but not by a lot (heavy tails result in a *more* favorable tradeoff…)

- In the regression model
  - The Edgeworth approximation to the frontier deteriorates substantially: the finite-sample frontier is less favorable than the asymptotic means-case frontier
  - Still, the qualitative findings go through:
    - NW, EWP have similar size distortions and crossing tradeoffs
    - The larger-*S* rules and fixed-*b* critical values improve size

MC results, regression: different rules, restricted/unrestricted

Rejection rates of HAR tests with nominal level 5% (*b* = *S*/*T*)

$y_t = \beta_0 + \beta_1 x_t + u_t$, x_t & u_t Gaussian AR(1), $\rho_x = \rho_u = 0.7^{1/2}$, $T = 200$

Estimator	Truncation rule for *b*	Critical values	Null imposed?	ρ = 0.3	ρ = 0.5	ρ = 0.7
NW	$0.75T^{-2/3}$	N(0,1)	No	**0.079**	**0.105**	**0.164**
NW	$1.3T^{-1/2}$	fixed-*b* (nonstandard)	No	0.067	0.080	0.107
EWP	$1.95T^{-2/3}$	fixed-*b* (t_v)	No	0.063	0.074	0.100
NW	$1.3T^{-1/2}$	fixed-*b* (nonstandard)	Yes	0.057	0.062	0.073
EWP	$1.95T^{-2/3}$	fixed-*b* (t_v)	Yes	0.052	0.056	0.066
Theoretical bound based on Edgeworth expansions for the Gaussian location model						
NW	$1.3T^{-1/2}$	fixed-*b* (nonstandard)	No	0.054	0.058	0.067
EWP	$1.95T^{-2/3}$	fixed-*b* (t_v)	No	0.052	0.056	0.073

Note: x_t and u_t are independent Gaussian AR(1)'s, single regressor.

8) Summary

Topics not covered here:

- Tests based on orthogonal series estimators (these include split sample or "batch means estimator" tests)
  - However these are covered in LLS and shown to be dominated by the EWP.
- Tests that are not psd (e.g. flat-top kernels)
- Tests that do not have known fixed-*b* asymptotic distributions
- Tests with data-dependent rules for *S*
- Tests with feasible size-adjusted critical values
  - However LLSW looked at these and found that they worked poorly in MCs
- Bootstrap tests
  - However current theory shows they are asymptotically equivalent to using fixed-*b* critical values.

Summary of current state of knowledge:

- Gaussian location model is well understood
- Regression model has some open puzzles
- Still, theory and MC results strongly point towards:
  - Larger bandwidths
  - Fixed-*b* critical values
  - NW kernel works well in typical sample sizes, with $S = 1.3T^{1/2}$
- **Software:** fixed-*b* critical values are available for NW from Vogelsang's web site – hopefully will get into STATA at some point...

Related literature

- *Classic spectral estimation:* Tukey (1949), Parzen (1957), Grenander & Rosenblatt (1957), Brillinger (1975), Priestley (1981)
- *Classic econometrics papers:* Newey-West (1987), Andrews (1991)
- *VAR-HAC:* Parzen, Berk (1974), den Han and Levin (1994)
- *Fixed-b:* Kiefer, Vogelsang, Bunzel (2000), Kiefer and Vogelsang (2002, 2005)
- *Small-b Edgeworth expansions*: Velasco and Robinson (2001), Jansson (2004), Sun, Phillips, & Jin (2008), Sun (2014)
- *Batch means estimator:* Blackman & Tukey (1958), Conway, Johnson, & Maxwell (1959), Ibragimov and Müller (2010)
- *Orthogonal series:* Grenander & Rosenblatt (1957), Foley & Goldsman (1988), Phillips (2005), Sun (2011, 2013)
- *Higher-order kernels (not psd):* Politis (2011)
- *Bootstrap:* Gonçalves & Vogelsang (2011), Zhang and Shao (2013)
- Recent survey: Müller (2014)

Selected References

Andrews, D.W. K. (1991), "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–858

Brillinger, D.R. (1975), *Time Series Data Analysis and Theory*. New York: Holt, Rinehart and Winston.

den Haan, W.J. and A. Levin (1994), "Vector Autoregressive Covariance Matrix Estimation," manuscript, Board of Governors of the Federal Reserve.

den Haan, W.J. and A. Levin (1997), "A Practitioners Guide to Robust Covariance Matrix Estimation," *Handbook of Statistics* 15, ch. 12, 291-341.

den Haan, W.J. and A. Levin (2000), "Robust Covariance Matrix Estimation with Data-Dependent VAR Prewhitening Order," NBER Technical Working Paper #255.

Hwang, J. and Y. Sun (2017). "Asymptotic *F* and *t* Tests in an Efficient GMM Setting," *Journal of Econometrics* 198(2), 2017, 277-295.

Ibragimov, R. and Müller, U.K. (2010), "*t*-statistic based correlation and heterogeneity robust inference," *Journal of Business and Economic Statistics* 28, 453-468.

Jansson, M. (2004), "The Error in Rejection Probability of Simple Autocorrelation Robust Tests," *Econometrica*, 72, 937-946.

Kiefer, N., T.J. Vogelsang, and H. Bunzel (2000), "Simple Robust Testing of Regression Hypotheses," *Econometrica*, 69, 695-714.

Kiefer, N. and T.J. Vogelsang (2002), "Heteroskedasticity-Autocorrelation Robust Standard Errors Using the Bartlett Kernel Without Truncation," *Econometrica*, 70, 2093-2095.

Kiefer, N. and T.J. Vogelsang (2005), "A New Asymptotic Theory for Heteroskedasticity-Autocorrelation Robust Tests," *Econometric Theory*, 21, 2093-2095.

Lazarus, E., D.J. Lewis and J.H. Stock (2017), "The Size-Power Tradeoff in HAR Inference," manuscript, Harvard University.

Lazarus, E., D.J. Lewis, J.H. Stock, and M.W. Watson (2018), "HAR Inference: Recommendations for Practice" (with discussion), *Journal of Business & Economic Statistics* 36(4), 541-575.

Müller, U. (2004), “A Theory of Robust Long-Run Variance Estimation,” working paper, Princeton University.

Müller, U. (2014), “HAC Corrections for Strongly Autocorrelated Time Series,” *Journal of Business and Economic Statistics* 32, 311-322.

Newey, W.K. and K.D. West (1987), “A Simple Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix,” *Econometrica* 55, 703-708.

Phillips, P. C. B. (2005), “HAC Estimation by Automated Regression,” *Econometric Theory* 21, 116–142.

Priestley, M.B. (1981), *Spectral Analysis and Time Series*. London: Academic Press.

Stock, J.H. and M.W. Watson (2008), “Heteroskedasticity-Robust Standard Errors for Fixed Effects Regression,” *Econometrica* 76, 155-174.

Stock, J.H. and M.W. Watson (2015), *Introduction to Econometrics*, *3rd Edition - Update*. Boston: Addison-Wesley.

Stoica, P. and R. Moses (2005), *Spectral Analysis of Signals*. Englewood Cliffs, NJ: Pearson Prentice Hall.

Sun Y., P.C.B. Phillips, and S. Jin (2008), “Optimal Bandwidth Selection in Heteroskedasticity-Autocorrelation Robust Testing,” *Econometrica*, 76(1): 175-194.

Sun, Y. (2013), “Heteroscedasticity and Autocorrelation Robust F Test Using Orthonormal Series Variance Estimator,” *The Econometrics Journal*, 16, 1–26.

Sun, Y. (2014a), “Let’s Fix It: Fixed-*b* Asymptotics versus Small-*b* Asymptotics in Heteroskedasticity and Autocorrelation Robust Inference,” *Journal of Econometrics* 178, 659-677.

Sun, Y. (2014b), “Fixed-smoothing Asymptotics in a Two-step GMM Framework,” *Econometrica* 82, 2327-2370.

Velasco, C. and P.M. Robinson (2001), “Edgeworth Expansions for Spectral Density Estimates and Studentized Sample Mean,” *Econometric Theory* 17, 497-539.

Westhoff, Frank (2013), *An Introduction to Econometrics: A Self-Contained Approach*. Cambridge: MIT Press.

Wooldridge, Jeffrey M. (2012). *Introductory Econometrics: A Modern Approach, 4th Edition*. Thomson.

AEA Continuing Education Course
Time Series Econometrics

Lecture 3 (two parts)

Identification and Estimation of Dynamic Causal Effects

James H. Stock
Harvard University

January 7, 2019, 10:00-11:45am and 12:45-2:15pm

Outline

Part A

1. **Dynamic causal effects: Overview**
    a. Definition and conceptual framework
    b. Estimation when the shock is observed
    c. Identification and estimation when the shock is unobserved
2. Multivariate methods with internal identification: SVARs

Part B

3. Single equation methods with internal identification: LP
4. Multivariate methods with external instruments: SVAR-IV
5. Single-equation methods with external instruments: LP-IV
6. Summary

Three real-world questions that economists are paid to answer

1. The President has criticized the Fed for its interest rate increases. What is the effect of a 25 bp increase in the FF rate, vs. keeping the FF rate constant, on price inflation, GDP growth, employment, and stock prices?

2. What is (will be) the effect of the TCJA on GDP growth, employment, the deficit, and wage and price inflation?

3. In its RIA justifying EPA's Aug, 2018 rollback of the 2023-2027 CAFÉ fuel economy standards (the SAFE rule), EPA cited the safety hazards of fuel economy standards: fuel economy standards will increase the price of cars, so drivers will purchase fewer new cars, and thus drive older, less safe cars for longer, thereby increasing traffic fatalities. What is the effect of a permanent increase in new car prices by (say) 1% on new vehicle sales?

Common theme: The answer to each question is a *dynamic causal effect*

Let $Y_{j,t}$ be a variable of interest (inflation, the deficit, new vehicle sales)

$\varepsilon_{1,t}$ be an unexpected policy-induced change ("shock") (a monetary policy shock to the FF rate, the JTCA "shock" to tax rates, the increase in new vehicle prices from an inward shift of the auto supply curve)

$\varepsilon_{2:m,t}$ = all other shocks/unexpected developments (m might be large!!)

Four flavors of dynamic causal effects

1. Potential outcomes:
$$\Theta_{h,j1} = Y_{j,t+h}\left(\varepsilon_{1,t}=1, \varepsilon_{1,s}=0(s \neq t), \varepsilon_{2:m,t}=0\right) - Y_{j,t+h}\left(\varepsilon_{1,t}=0, \varepsilon_{1,s}=0(s \neq t), \varepsilon_{2:m,t}=0\right)$$

2. *Ceteris paribus* (nonstochastic):
$$\Theta_{h,j1} = \left.\frac{\partial Y_{j,t+h}}{\partial \varepsilon_{1,t}}\right|_{\varepsilon_{1,s}(s\neq t),\varepsilon_{2:m,t}}, \; h = 1, 2, 3, \ldots.$$

3. Conditional expectations:
$$\Theta_{h,j1} = E_t\left(Y_{j,t+h} \mid \varepsilon_{1,t}=1, \varepsilon_{1,s}=0(s \neq t), \varepsilon_{2:m,t}=0\right) - E_t\left(Y_{j,t+h} \mid \varepsilon_{1,t}=0, \varepsilon_{1,s}=0(s \neq t), \varepsilon_{2:m,t}=0\right)$$

4. Conditional expectations with independent shocks:
$$\Theta_{h,j1} = E_t\left(Y_{j,t+h} \mid \varepsilon_{1,t}=1\right) - E_t\left(Y_{j,t+h} \mid \varepsilon_{1,t}=0\right)$$

Terminology

Conditional expectations with independent shocks:

$$\Theta_{h,j1} = E_t\left(Y_{j,t+h} \mid \varepsilon_{1,t} = 1\right) - E_t\left(Y_{j,t+h} \mid \varepsilon_{1,t} = 0\right)$$

- $\Theta_{h,j1}$ is the ***h*-period dynamic causal effect** of $\varepsilon_{1,t}$ on $Y_{j,t}$
- $\Theta_{0,j1}$ is the **(causal) impact effect** of $\varepsilon_{1,t}$ on $Y_{j,t}$
- $\{\Theta_{h,j1}\}, h = 0, 1, 2, \ldots$ is the **impulse response function** of $Y_{j,t}$ to $\varepsilon_{1,t}$

<u>Recent literature on dynamic causal effects from primitives</u>

Lechner (2009), Angrist, Jordà, and Kuersteiner (2017), Jordà, Schularick, and Taylor (2017), Bojinov and Shephard (2017)

Refresher on potential outcomes

Textbook references: Imbens (2014); Angrist & Pischke (2009), Stock & Watson (2018b)

$Y_i(1)$ = outcome if treatment received
$Y_i(0)$ = outcome if treatment not received
Y_i = observed outcome
X_i = treatment (binary)

From potential outcomes to regression:

$$
\begin{aligned}
Y_i &= Y_i(1)X_i + Y_i(0)(1-X_i) \\
&= EY_i(0) + [Y_i(1) - Y_i(0)]X_i + [Y_i(0) - EY_i(0)] \\
&= \alpha + \beta_i X_i + u_i
\end{aligned}
$$

where

u_i = no-treatment baseline for individual i
$\beta_i = Y_i(1) - Y_i(0)$ = treatment effect for individual i
$E\beta_i = E[Y_i(1) - Y_i(0)]$ = average treatment effect (ATE)

The OLS estimand is the ATE if:

$X_i \perp (Y_i(0), Y_i(1))$ (random assignment of treatment)

$\Leftrightarrow \; X_i \perp u_i \quad (\Rightarrow E(u_iX_i) = 0)$

Dynamic causal effects with linearity and independent (uncorrelated) shocks

Linearity:

$$\overset{n\times 1}{Y_t} = \Theta_t(L)\overset{m\times 1}{\varepsilon_t}$$

$$= \overset{n\times 1}{\Theta_{1,t}}(L)\varepsilon_{1t} + \overset{n\times(m-1)}{\Theta_{2:m,t}}(L)\varepsilon_{2:m,t} = \overset{n\times 1}{\Theta_{1,t}}(L)\varepsilon_{1t} + u_t$$

Linearity + stationarity

- Potential outcomes analog: homogeneous treatment effects

$$\overset{n\times 1}{Y_t} = \Theta(L)\overset{m\times 1}{\varepsilon_t}$$

Linearity + stationarity + independence [uncorrelatedness] of shocks:

- Potential outcomes analog: treatment randomly or as-if randomly assigned

Structural Moving Average

$$\overset{n\times 1}{Y_t} = \Theta(L)\overset{m\times 1}{\varepsilon_t},\ E(\varepsilon_t\varepsilon_t') = \text{diagonal},\ E(\varepsilon_t\varepsilon_t') = 0,\ t \neq s$$

Singling out first shock, putting $\varepsilon_{2:m,t}$ in the error term

$$\overset{n\times 1}{Y_t} = \overset{n\times 1}{\Theta_1}(L)\varepsilon_{1t} + u_t \text{ where } \{\varepsilon_{1t}\} \perp \{u_t\} \Rightarrow E\left(u_t \mid \varepsilon_{1t}, \varepsilon_{1t-1}, \ldots\right) = 0$$

Estimation of DCEs When the Shock is Observed

Henceforth, we assume linearity + stationarity + uncorrelated shocks

- If $\varepsilon_{1,t}$ is observed, estimating the DCE is a straightforward regression problem, aside from the technical difficulty of infinitely many lags
- The regression can be implemented as a single regression or as separate regressions, one for each horizon:

Three variants:

(a) $\overset{n\times 1}{Y_t} = \overset{n\times 1}{\Theta_1}(L)\varepsilon_{1t} + u_t$ with $E\left(u_t \mid \varepsilon_{1t}, \varepsilon_{1t-1}, \ldots\right) = 0$

(b) $Y_{j,t+h} = \Theta_{h,j1}\varepsilon_{1t} + u_{j,t+h}^{(h)}$, $h = 1, \ldots$

(c) $Y_{j,t+h} = \Theta_{h,j1}\varepsilon_{1t} + \delta(L)Y_{t-1} + u_{j,t+h}^{(h)}$ (additional variables to get smaller SEs)

Technical notes:

- All three directly trace their roots to the four flavors of DCE
- (a) is a distributed lag regression
- (c) is also called a "direct" forecasting regression
- All three require HAR SEs (in general)

Estimation of DCEs: case of an observed shock, ctd.

$$Y_{j,t+h} = \Theta_{h,j1}\varepsilon_{1t} + \delta(L)Y_{t-1} + u^{(h)}_{j,t+h} \text{ (additional variables to get smaller SEs)}$$

- The measured shock approach has been popular in the monetary shock literature, where the monetary policy shock is measured as the surprise change in an interest rate around announcement window (press conference window)
  - Kuttner (2001)
    - Cochrane and Piazessi (2002) aggregates daily Eurodollar rate changes after FOMC announcements to a monthly shock series
    - Faust, Swanson, and Wright (2003, 2004) estimate monetary policy shock estimate from futures markets
  - Bernanke and Kuttner (2005)

- The conditional mean independence condition provides a framework for evaluating the internal validity of the regression:

$$E\left(u^{(h)}_{j,t+h} \mid \varepsilon_{1,t}, Y_{t-1}, Y_{t-2}, \ldots\right) = E\left(u^{(h)}_{j,t+h} \mid Y_{t-1}, Y_{t-2}, \ldots\right)$$

which will hold if ε_{1t} is in fact a structural shock (but is it, in a given application)

Estimation of DCEs When the Shock is Unobserved

Two estimation methods:

- Multiple equation: Structural VARs (SVARs)
- Single equation: direct multistep regressions (called Local projections in this literature)

Two identification frameworks:

- Internal identification – restrictions on coefficients
- External identification – external instruments

Remainder of this lecture:

- Will go through the estimation methods and identification frameworks (four cases)
- The literature treats the identification requirements of the two methods as different. A major theme of this lecture is that, in general, they are not.
  - References: Stock & Watson (2018), Plagborg-Møller and Wolf (2018)

Outline

Part A

1. Dynamic causal effects: Overview
   a. Definition and conceptual framework
   b. Estimation when the shock is observed
   c. Identification and estimation when the shock is unobserved
2. **Multivariate methods with internal identification: SVARs**

Part B

3. Single equation methods with internal identification: LP
4. Multivariate methods with external instruments: SVAR-IV
5. Single-equation methods with external instruments: LP-IV
6. Summary

SVARs with Internal Identification: Setup and Maintained Assumptions

Assumptions: **Linearity + stationarity + uncorrelated shocks + invertibility**

Structural MA: $Y_t = \Theta(\mathrm{L})\varepsilon_t$

VAR: $\mathrm{A(L)}Y_t = \nu_t$, where $\nu_t = Y_t - \mathrm{Proj}(Y_t | Y_{t-1}, Y_{t-2},\ldots)$ = Wold errors

Invertibility: $\varepsilon_t = \mathrm{Proj}(\varepsilon_t \mid Y_t, Y_{t-1}, \ldots)$

which implies: $\nu_t = \Theta_0 \varepsilon_t$, where $m = n$ (i.e., # ε's = # Y's) and Θ_0^{-1} exists

SVAR IRFs: $Y_t = A(\mathrm{L})^{-1}\nu_t = C(\mathrm{L})\Theta_0\varepsilon_t$, where C(L) = A(L)$^{-1}$

so $$\Theta_{h,i1} = C_h \Theta_{0,i1} \qquad (*)$$

The expression (*) is the payoff of SVARs!

Under the SVAR assumptions, if you can estimate the impact effect $\Theta_{0,1}$, you can estimate the entire dynamic effect for all the variables in the system

Invertibility – what does it mean?

Invertibility is a critical assumption in getting the SVAR payoff (*)

Digression: Proof that invertibility implies that Θ_0^{-1} exists

Start with the structural MA: $\quad Y_t = \Theta(\mathrm{L})\varepsilon_t$

From the definition of the innovation,

$$\begin{aligned}
\nu_t &= Y_t - \mathrm{Proj}\left(Y_t \mid Y_{t-1}, Y_{t-2}, \ldots\right) \\
&= \Theta(\mathrm{L})\varepsilon_t - \mathrm{Proj}\left(\Theta(\mathrm{L})\varepsilon_t \mid Y_{t-1}, Y_{t-2}, \ldots\right) \qquad \text{(structural MA)} \\
&= \Theta_0\varepsilon_t + \sum_{i=1}^{\infty}\Theta_i\left[\varepsilon_{t-i} - \mathrm{Proj}\left(\varepsilon_{t-i} \mid Y_{t-1}, Y_{t-2}, \ldots\right)\right] \qquad \text{(rearranging)} \\
&= \Theta_0\varepsilon_t \qquad \text{(using definition of invertibility)}
\end{aligned}$$

so,

$$\begin{aligned}
\mathrm{Proj}\left(\varepsilon_t \mid Y_t, Y_{t-1}, \ldots\right) &= \mathrm{Proj}\left(\varepsilon_t \mid \nu_t, \nu_{t-1}, \ldots\right) \qquad \text{(follows from defintion of innovations)} \\
&= \mathrm{Proj}\left(\varepsilon_t \mid \Theta_0\varepsilon_t, \Theta_0\varepsilon_{t-1}, \ldots\right) \qquad \text{(from above)} \\
&= \mathrm{Proj}\left(\varepsilon_t \mid \Theta_0\varepsilon_t\right) \qquad (\varepsilon_t \text{ serially uncorrelated})
\end{aligned}$$

from which it follows that Θ_0^{-1} exists

Invertibility as no OVB

Invertibility can be interpreted as “no omitted variables” (Fernández-Villaverde et al (2007)):

$$\begin{aligned}\text{Proj}(Y_t \mid Y_{t-1}, Y_{t-2}, ..., \varepsilon_{t-1}, \varepsilon_{t-2}, ...) &= \text{Proj}\left(Y_t \mid \nu_{t-1}, \nu_{t-2}, ..., \varepsilon_{t-1}, \varepsilon_{t-2}, ...\right) \\ &= \text{Proj}\left(Y_t \mid \nu_{t-1}, \nu_{t-2}, ...\right) \qquad (\varepsilon_{t-1} = \Theta_0^{-1}\nu_{t-1} \text{ by invertibility}) \\ &= \text{Proj}(Y_t \mid Y_{t-1}, Y_{t-2}, ...)\end{aligned}$$

This is a very strong condition! If you were a forecaster and could download the true shock history, would you do so?

Invertibility references

Lippi and Reichlin (1993, 1994), Sims and Zha (2006b), Fernandez-Villaverde, Rubio-Ramirez, Sargent, and Watson (2007), Hansen and Sargent (2007), E. Sims (2012), Blanchard, L’Huillier, and Lorenzoni (2012), Forni and Gambetti (2012), and Gourieroux and Monfort (2014), Plagborg-Møller (2016), Plagborg-Møller and Wolf (2018a, 2018b), Miranda-Aggripino and Ricco(2018), Stock and Watson (2018)

The SVAR identification problem

Assumptions: **Linearity + stationarity + uncorrelated shocks + invertibility**

VAR: $A(L)Y_t = \nu_t$,

SVAR IRFs: $Y_t = A(L)^{-1}\nu_t = C(L)\Theta_0\varepsilon_t = \Theta(L)\varepsilon_t$, so $\Theta_j = C_j\Theta_0$

and in particular: $\Theta_{h,i1} = C_h\Theta_{0,i1}$

The SVAR identification problem under the assumption of invertibility is the requirement that $\Theta_{0,i1}$(first column of Θ_0) be identified, or (if one is interested in IRFs for all shocks) that Θ_0 be identified.

The SVAR identification problem, ctd

System identification. In general, the SVAR is fully identified if

$$\Theta_0 \Sigma_v \Theta_0' = \Sigma_\varepsilon, \text{ where } \Sigma_\varepsilon = \text{diagonal}$$

can be solved for the unknown elements of R and Σ_ε.. Recall that Σ_u is identified.

- There are $n(n+1)/2$ distinct equations in the matrix equation above, so the order condition says that you can estimate (at most) $k(k+1)/2$ parameters.
- Normalization of the scale of ε delivers n parameters
  - Unit standard deviation normalization: $\Sigma_\varepsilon = I$
  - Unit effect normalization: $\Theta_{0,ii} = 1$
- So we need $n^2 - n(n+1)/2 = n(n-1)/2$ restrictions on Θ_0.
- If $n = 2$, then $n(n-1)/2 = 1$, which is delivered by imposing a single restriction (commonly, that Θ_0 is lower or upper triangular).
- This ignores rank conditions, which can matter.

SVAR Identification by Short Run Restrictions

Example: Dynamic effect of new vehicle prices on sales.

- In its August 2018 Proposed Regulatory Impact Analysis of the SAFE rule (CAFÉ rollback), NHTSA modeled the effect of a one-time permanent change in the price level on new vehicle sales.
- Here, the DCE is just a time path of elasticities
- NHTSA variables:

q_t = log(Sales$_t$) (Sales = number of vehicles sold)
p_t = log(average vehicle price)
Emp_t = log(Payroll Employment$_t$)
GDP_t = GDP growth (percent at annual rate, SA)

SVAR Identification by short run restrictions, ctd.

Let $W_t = Emp_t, GDP_t$. Three modeling options:

Distributed lag (DL): $q_t = \Theta_{qp}(L)p_t + \gamma(L)W_t + u_t$

exogeneity reqm't: $E\left(u_t \mid p_t, p_{t-1},\ldots,W_t,W_{t-1},\ldots\right) = E\left(u_t \mid W_t,W_{t-1},\ldots\right)$

(conditional weak exogeneity)

Autoregressive DL (ADL): $q_t = \alpha(L)q_{t-1} + \beta(L)p_t + \gamma(L)W_t + u_t$

where $\Theta_{qp}(L) = \left(1-\alpha(L)\right)^{-1}\beta(L)$

exogeneity reqm't: $E\left(u_t \mid p_{t+1}, p_t, p_{t-1},\ldots,W_t,W_{t-1},\ldots\right) = E\left(u_t \mid W_t,W_{t-1},\ldots\right)$

(conditional strict exogeneity)

Both conditional weak exogeneity and conditional strict exogeneity are strong assumptions in this application – effectively they say that car dealerships don't hold sales if they have excess inventories, and don't raise prices if they see or expect strong demand.

SVAR Identification by short run restrictions, ctd.

SVAR: $Y_t = A(L)Y_{t-1} + v_t$

q equation: $q_t = \Theta_{0,qp} p_t + \Theta_{0,qW} W_t + \gamma(L)Y_{t-1} + \varepsilon_t^p$ (*)

exogeneity reqm't: $E\left(\varepsilon_t^q \mid p_t, W_t, Y_{t-1}, Y_{t-2}, \ldots\right) = E\left(\varepsilon_t^q \mid 1Y_{t-1}, Y_{t-2}, \ldots\right) = 0$

(contemporaneous conditional exogeneity)

(contemporaneous conditional mean independence)

- In words: dealers can cut prices based on last quarter's sales but not on unexpected current-quarter demand surges or drops, except as related to overall economic conditions.
- This is a weaker requirement than for the DL and ADL models

Rewrite as a VAR: First, rewrite (*) in terms of innovations:

$$v_t^p = p_t - \text{Proj}\left(p_t \mid Y_{t-1}\right)$$

so (*) becomes

$$v_t^q = \Theta_{0,qp} v_t^p + \Theta_{0,qW} v_t^W + \varepsilon_t^p$$

SVAR Identification by short run restrictions, ctd.

In VAR notation, $E\left(\varepsilon_t^q \mid p_t, W_t, Y_{t-1}, Y_{t-2}, ...\right) = 0$

is $E\left(\varepsilon_t^q \mid \nu_t^p, \nu_t^W\right) = 0$

This implies that $\left(\nu_t^p, \nu_t^W\right)$ don't depend on ε_t^q, so ε_t^q is ordered last:

$$\begin{pmatrix} \nu_t^W \\ \nu_t^p \\ \nu_t^q \end{pmatrix} = \begin{pmatrix} \Theta_{0,WW} & 0 & 0 \\ \Theta_{0,pW} & \Theta_{0,pp} & 0 \\ \Theta_{0,qW} & \Theta_{0,qp} & \Theta_{0,qq} \end{pmatrix} \begin{pmatrix} \varepsilon_t^W \\ \varepsilon_t^p \\ \varepsilon_t^q \end{pmatrix}$$

Normalization? The ADL, DL equations use the unit effect normalization:

$$\begin{pmatrix} \nu_t^W \\ \nu_t^p \\ \nu_t^q \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ \Theta_{0,pW} & 1 & 0 \\ \Theta_{0,qW} & \Theta_{0,qp} & 1 \end{pmatrix} \begin{pmatrix} \varepsilon_t^W \\ \varepsilon_t^p \\ \varepsilon_t^q \end{pmatrix}$$

SVAR Identification by short run restrictions, ctd.

SVAR sales model: $A(L)Y_t = \nu_t$

Identification of $\Theta_{0,p}$ (column of impulse effects of a price shock):

$$\begin{pmatrix} \nu_t^W \\ \nu_t^p \\ \nu_t^q \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ \Theta_{0,pW} & 1 & 0 \\ \Theta_{0,qW} & \Theta_{0,qp} & 1 \end{pmatrix} \begin{pmatrix} \varepsilon_t^W \\ \varepsilon_t^p \\ \varepsilon_t^q \end{pmatrix}$$

SVAR IRFs: $Y_t = A(L)^{-1}\nu_t = C(L)\Theta_0\varepsilon_t = \Theta(L)\varepsilon_t$, so $\Theta_j = C_j\Theta_0$

and in particular: $\Theta_{h,p} = C_h\Theta_{0,p}$

- Note: the ordering of *W*, *p* is arbitrary – get the same results for IRF for *q*.
- In this model, we have identified the price and quantity shocks, not the separate employment and GDP shocks

Identification by Long Run Restrictions

This approach identifies Θ by imposing restrictions on the long run effect of one or more ε's on one or more Y's.

Reduced form VAR: $A(L)Y_t = \nu_t$

Structural VAR: $\nu_t = \Theta_0\varepsilon_t$

Long-run effect of ε on Y: $Y_t = C(1)\Theta_0\varepsilon_t$

Typical long-run restriction: ε_{1t} has no long-run effect on $Y_{j,t}$

This imposes zero restrictions on $C(1)\Theta_0$ and thus restrictions on Θ_0

Digression: $A(1)^{-1} = C(1)$ is the long-run effect on Y_t of ν_t; this can be seen using the Beveridge-Nelson decomposition,

$$\sum_{s=1}^{t} Y_s = C(1)\sum_{s=1}^{t} \nu_s + C^*(L)\varepsilon_t, \text{ where } C_i^* = -\sum_{j=i+1}^{\infty} C_j$$

***Comments*:**

- If the zero restrictions on $C(1)\Theta_0$ make $C(1)\Theta_0$ lower triangular and the unit standard deviation normalization is used (so $\Sigma_\varepsilon = I$), then $C(1)\Theta_0$ is the Cholesky factorization of $\Omega = A(1)^{-1}\Sigma_v A(1)^{-1}$, so $\Theta_0 = A(1)Chol(\Omega)$.
- Blanchard-Quah (1989) had 2 variables (unemployment and output), with the restriction that the demand shock has no long-run effect on the unemployment rate. This imposed a single zero restriction, which is all that is needed for system identification when $k = 2$.
- King, Plosser, Stock, and Watson (1991) work through system and partial identification (identifying the effect of only some shocks), things are analogous to the partial identification using short-run timing.
- This approach was at the center of a debate about whether technology shocks lead to a short-run decline in hours, based on long-run restrictions (Galí (1999), Christiano, Eichenbaum, and Vigfusson (2004, 2006), Erceg, Guerrieri, and Gust (2005), Chari, Kehoe, and McGrattan (2007), Francis and Ramey (2005), Kehoe (2006), and Fernald (2007))
- The theoretical grounding of long-run restrictions is often questionable; for a case in favor of this approach, see Giannone, Lenza, and Primiceri (2014)

Long run restrictions, ctd.

In this literature, Ω is estimated using the VAR-HAC estimator,

VAR-HAC estimator of Ω: $\hat{\Omega} = \hat{A}(1)^{-1}\hat{\Sigma}_{\hat{v}}\hat{A}(1)^{-1\prime}$

Estimator of Θ_0 under unit std. dev. normalization: $\hat{\Theta}_0 = \hat{A}(1)Chol\left(\hat{\Omega}\right)$

Comments:

- This confronts the problem of estimating the LRV so not surprisingly encounters sampling distribution problems.
- A recurring theme is the sensitivity of the results to apparently minor specification changes, in Chari, Kehoe, and McGrattan's (2007) example results are sensitive to the lag length. It is unlikely that $\hat{\Sigma}_u$ is sensitive to specification changes, but $\hat{A}(1)$ is much more difficult to estimate.
- These observations are closely linked to the critiques by Faust and Leeper (1997), Pagan and Robertson (1998), Sarte (1997), Cooley and Dwyer (1998), Watson (2006), and Gospodinov (2008), which are essentially weak instrument concerns.

Identification from Heteroskedasticity

Simplest case: Discrete break in heteroskedasticity at a known date

Suppose:

(a) The structural shock variance breaks at date s: $\Sigma_{\varepsilon,1}$ before, $\Sigma_{\varepsilon,2}$ after.

(b) Θ_0 doesn't change between variance regimes.

(c) Adopt the unit effect normalization.

First period: $\Theta_0\Sigma_{u,1}\Theta_0' = \Sigma_{\varepsilon,1}$ $\quad$ $k(k+1)/2$ equations, k^2 unknowns

Second period: $\Theta_0\Sigma_{u,2}\Theta_0' = \Sigma_{\varepsilon,2}$ $\quad$ $k(k+1)/2$ equations, k more unknowns

Number of equations = $k(k+1)/2 + k(k+1)/2 = k(k+1)$

Number of unknowns = $k^2 - k + k + k = k(k+1)$

Rigobon (2003), Rigobon and Sack (2003, 2004)

ARCH version by Sentana and Fiorentini (2001)

General time-varying cond'l variances or stoch. volatility: Lewis (2018)

Identification from Heteroskedasticity, ctd.

Comments:

1. There is a rank condition here too – for example, identification will not be achieved if $\Sigma_{\varepsilon,1}$ and $\Sigma_{\varepsilon,2}$ are proportional.
2. The break date need not be known as long as it can be estimated consistently
3. Different intuition: suppose only one structural shock is homoskedastic. Then find the linear combination without any heteroskedasticity!
4. **Major generalization:** Lewis (2018) – don't need to identify regimes or the volatility process (!)
5. But, some cautionary notes:
   a. Θ_0 must remain constant despite change in Σ_ε
   b. Shocks are identified only up to order – i.e. they are not "named". Lewis's (2018) result implies that with time-varying variances, SVARs are generically identified up to the names of the shocks.
   c. Strong identification will come from large differences in variances

Example: Wright (2012), Monetary Policy at ZLB

Identification by Sign Restrictions

Consider restrictions of the form: a monetary policy shock…

- does not decrease the FF rate for months 1,…,6
- does not increase inflation for months 6,..,12

These are restrictions on the sign of elements of $\Theta(L)$.

Sign restrictions can be used to set-identify $\Theta(L)$. Let $\boldsymbol{\Theta}$ denote the set of $\Theta(L)$'s that satisfy the restriction. There are currently three ways to handle sign restrictions:

1. Faust's (1998) quadratic programming method
2. Uhlig's (2005) Bayesian method
3. Uhlig's (2005) penalty function method

I will describe #2, which is the most popular method (the first steps are the same as #3; #1 has only been used a few times)

Sign restrictions, ctd.

It is useful to rewrite the identification problem after normalizing by a Cholesky factorization (and setting $\Sigma_\varepsilon = I$):

SVAR identification: $\Theta_0\Sigma_v\Theta_0' = \Sigma_\varepsilon$

Normalize $\Sigma_\varepsilon = I$; then $\Sigma_v = \Theta_0^{-1}\,\Theta_0^{-1\prime} = R_c^{-1}QQ'R_c^{-1\prime}$

Where $R_c^{-1} = Chol(\Sigma_v)$ and Q is a $n\times n$ orthonormal matrix so $QQ' = I$. Then

Structural errors: $u_t = R_c^{-1}Q\varepsilon_t$

Structural IRF: $\Theta(L) = C(L)R_c^{-1}Q$

Let $\boldsymbol{\Theta}$ denote the set of acceptable IRFs (IRFs that satisfy the sign restrictions)

Sign restrictions, ctd.

Structural IRF: $\Theta(L) = C(L)R_c^{-1}Q$

Uhlig's algorithm (slightly modified):

(i) Draw $\tilde{Q}$ randomly from the space of orthonormal matrices

(ii) Compute the IRF $\tilde{\Theta}(L) = \Theta(L) = C(L)R_c^{-1}\tilde{Q}$

(iii) If $\tilde{\Theta}(L) \notin \mathbf{\Theta}$, discard this trial $\tilde{Q}$ and go to (i). Otherwise, if $\tilde{\Theta}(L) \in \mathbf{\Theta}$, retain $\tilde{Q}$ then go to (i)

(iv) Compute the posterior (using a prior on A(L) and Σ_v, plus the retained $\tilde{Q}$'s) and conduct Bayesian inference, e.g. compute posterior mean (integrate over A(L), Σ_v, and the retained $\tilde{Q}$'s), compute credible sets (Bayesian confidence sets), etc.

This algorithm implements Bayes inference using a prior proportional to

$$\pi(A(L), \Sigma_v)\times\mathbf{1}(\tilde{\Theta}(L) \in \mathbf{\Theta})\mu(Q)$$

where $\mu(Q)$ is the distribution from which Q is drawn.

Sign restrictions, n = 2 example

Consider a $n = 2$ VAR: $A(L)Y_t = u_t$ and structural IRF

$$\Theta(L) = \begin{pmatrix} \Theta_{11}(L) & \Theta_{12}(L) \\ \Theta_{21}(L) & \Theta_{22}(L) \end{pmatrix} = A(L)^{-1} R_c^{-1} Q.$$

The sign restriction is $\Theta_{21,I} \geq 0$, $I = 1,\ldots, 4$ (shock 1 has a positive effect on variable 2 for the first 4 quarters).

Suppose the population reduced form VAR is $A(L)Y_t = u_t$ where

$$A(L) = \begin{pmatrix} (1-\alpha_1 L)^{-1} & 0 \\ 0 & (1-\alpha_2 L)^{-1} \end{pmatrix} \text{ and } \Sigma_v = I \text{ so } R_c^{-1} = I.$$

What does set-identified Bayesian inference look like for this problem, in a large sample?

- With point-identified inference and nondogmatic priors, it looks like frequentist inference (Bernstein-von Mises theorem)

Sign restrictions, n = 2 example, ctd.

Step 1: use n =2 to characterize Q

In the n = 2 case, the restriction QQ' = I implies that there is only one free parameter in Q, so that all orthonormal Q can be written,

$$Q = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \text{ [check: } \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} = I]$$

- The standard method, used here, is to draw Q by drawing $\theta \sim$ U$[0,2\pi]$
- The main point of this example is that the uniform prior on θ ends up being informative for what matters, D(L), so much so that the prior induced a Bayesian posterior coverage region strictly inside the identified set.

Step 2: Condition for checking whether Q is retained:

$$\hat{\Theta}_{21}(L) = \left[\hat{A}(L)^{-1}\hat{R}_c^{-1}Q\right]_{21} \geq 0 \text{ for first 4 lags}$$

Sign restrictions, n = 2 example, ctd.

Step 3: In a very large sample, A(L) and Σ_n will be essentially known (WLLN), so that

$$\hat{A}(L)^{-1}\hat{R}_c^{-1}Q \approx \begin{pmatrix} (1-\alpha_1 L)^{-1} & 0 \\ 0 & (1-\alpha_2 L)^{-1} \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

$$= \begin{pmatrix} (1-\alpha_1 L)^{-1}\cos\theta & -(1-\alpha_1 L)^{-1}\sin\theta \\ (1-\alpha_2 L)^{-1}\sin\theta & (1-\alpha_2 L)^{-1}\cos\theta \end{pmatrix}$$

so $$\hat{\Theta}_{21}(L) = \left[\hat{A}(L)^{-1}\hat{R}_c^{-1}Q\right]_{21} \approx (1-\alpha_2 L)^{-1}\sin\theta$$

Thus the step, keep Q if $\hat{\Theta}_{21,i} \geq 0$, $i = 1,\ldots,4$ reduces to keep Q if $\sin\theta \geq 0$, which is equivalent to $0 \leq \theta \leq \pi$.

Thus, in large samples the posterior of $\hat{\Theta}_{21}(L)$ is $\approx (1-\alpha_2 L)^{-1}\sin\theta$, for $\theta \sim U[0,\pi]$.

Sign restrictions, n = 2 example, ctd.

Characterization of posterior

A draw from the posterior (for a retained θ is): $\Theta_{21}(L) = (1-\alpha_2 L)^{-1}\sin\theta$

Posterior mean for $D_{21,i}$:

$$\begin{aligned} E[\Theta_{21,i}] &= E\left(\alpha_2^i \sin\theta\right) \\ &= \alpha_2^i E\left(\sin\theta\right) \\ &= \alpha_2^i \int_0^\pi \frac{1}{\pi}\sin\theta d\theta \\ &= \frac{\alpha_2^i}{\pi}(-\cos\theta\big|_0^\pi) = \frac{2}{\pi}\alpha_2^i \approx .637\alpha_2^i \end{aligned}$$

Posterior distribution: drop scaling by α_2^i and focus on sinθ part

$$\Pr[\sin\theta \le x] = \Pr[\theta \le \mathrm{Sin}^{-1}(x)] \text{ for } \theta \sim U[0,\pi/2]$$

$$= 2\mathrm{Sin}^{-1}(x)/\pi$$

So the pdf of x is: $f_X(x) = \dfrac{d}{dx}\dfrac{2}{\pi}\mathrm{Sin}^{-1}(x) = \dfrac{2}{\pi\sqrt{1-x^2}}$

Characterization of posterior, ctd.

So the posterior of $\hat{D}_{21,i}$ is: $p(\hat{\Theta}_{21,i}|Y) \propto \dfrac{2}{\pi\sqrt{1-x^2}}\alpha_2^i$

67% posterior probability interval with equal mass in each tail:

Lower cutoff:

$$\Pr[\sin\theta \le x] = 1/6 \rightarrow x_{lower} = \sin(\pi/12) = .259$$

$$\Pr[\sin\theta \le x] = 5/6 \rightarrow x_{upper} = \sin(5\pi/12) = .966$$

so 67% posterior coverage interval is $[.259\alpha_2^i, .966\alpha_2^i]$, with mean $.637\alpha_2^i$

What's wrong with this picture?

- Posterior coverage interval: $[.259\alpha_2^i, .966\alpha_2^i]$, with mean $.637\alpha_2^i$
- Identified set is $[0, \alpha_2^i]$
- What is the frequentist confidence interval here?
- Why don't Bayesian and frequentist coincide?

Sign restrictions, ctd.

Recent references on sign-restriction VARs:

Baumeister and Hamilton (*ECMA*, 2015)

Fry and Pagan (2011)

Kilian and Murphy (*JEEA*, 2012)

Moon and Schorfheide (*ECMA*, 2012)

Moon, Schorfheide, and Granziera (*QE*, 2018)

Giacomini and Kitagawa (ms, 2015)

Outline

Part A

1. Dynamic causal effects: Overview
   a. Definition and conceptual framework
   b. Estimation when the shock is observed
   c. Identification and estimation when the shock is unobserved
2. Multivariate methods with internal identification: SVARs

Part B

3. **Single equation methods with internal identification: LP**
4. Multivariate methods with external instruments: SVAR-IV
5. Single-equation methods with external instruments: LP-IV
6. Summary

Local Projections

Local projections estimate the dynamic causal effect of interest one equation at a time. The term "local projections" dates to Jordà (2005)

Some useful notation: {.} = linear combination of the variables in brackets

Algebra leading to LP regression:

Linearity + stationarity + independence [uncorrelatedness] of shocks:

- Potential outcomes analog: treatment randomly or as-if randomly assigned

Start with Structural Moving Average:

$$\underset{n\times 1}{Y_t} = \Theta(L)\underset{m\times 1}{\varepsilon_t}, \; E(\varepsilon_t\varepsilon_t') = \text{diagonal}, \; E(\varepsilon_t\varepsilon_t') = 0, \; t \neq s$$

Single out first shock, putting $\varepsilon_{2:m,t}$ in the error term

$$\underset{n\times 1}{Y_t} = \underset{n\times 1}{\Theta_1}(L)\varepsilon_{1t} + \{\varepsilon_{2:n,t}, \varepsilon_{2:n,t-1}, \ldots\}$$

Local projections algebra, ctd.

DL: $$\overset{n\times 1}{Y_t} = \overset{n\times 1}{\Theta_1}(L)\varepsilon_{1t} + \{\varepsilon_{2:n,t}, \varepsilon_{2:n,t-1}, ...\}$$

This DL can be implemented as separate regressions, one for each horizon:

$$Y_{j,t+h} = \Theta_{h,j1}\varepsilon_{1t} + \{\varepsilon_{t+h}, ..., \varepsilon_{t+1}, \varepsilon_{2:n,t}, \varepsilon_{t-1}, ...\},\ h = 1,2,\ldots$$

Lagged *Y*'s can be added to get smaller SEs:

$$Y_{j,t+h} = \Theta_{h,j1}\varepsilon_{1t} + \delta(L)Y_{t-1} + \{\varepsilon_{t+h}, ..., \varepsilon_{t+1}, \varepsilon_{2:n,t}\} \qquad (*)$$

(*) is the LP regression for ε_{1t} observed.

Local projections when ε_{1t} is not observed:

$$Y_{j,t+h} = \Theta_{h,j1}\varepsilon_{1t} + \delta(L)Y_{t-1} + \{\varepsilon_{t+h}, ..., \varepsilon_{t+1}, \varepsilon_{2:n,t}\} \qquad (*)$$

Must use some assumptions/restrictions to identify $\varepsilon_{1,t}$ and thus $\Theta_{h,j1}$

Local projections when ε_{1t} is not observed: Timing restrictions

$$Y_{j,t+h} = \Theta_{h,j1}\varepsilon_{1t} + \delta(L)Y_{t-1} + \{\varepsilon_{t+h},\ldots,\varepsilon_{t+1},\varepsilon_{2:n,t}\} \qquad (*)$$

LP with timing restrictions

New car sales example: $\begin{pmatrix} \nu_t^W \\ \nu_t^p \\ \nu_t^q \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ \Theta_{0,pW} & 1 & 0 \\ \Theta_{0,qW} & \Theta_{0,qp} & 1 \end{pmatrix} \begin{pmatrix} \varepsilon_t^W \\ \varepsilon_t^p \\ \varepsilon_t^q \end{pmatrix}$

so $\qquad \varepsilon_t^p = \nu_t^p - \text{Proj}\left(\nu_t^p \mid \nu_t^W\right)$

Denote the error $\{\varepsilon_{t+h},\ldots,\varepsilon_{t+1},\varepsilon_{2:n,t}\} = u_{j,t+h}^{(h)}$ so the LP equation becomes,

$$\begin{aligned} Y_{j,t+h} &= \Theta_{h,j1}\left[\nu_t^p - \text{Proj}\left(\nu_t^p \mid \nu_t^W\right)\right] + \delta(L)Y_{t-1} + u_{j,t+h}^{(h)} \\ &= \Theta_{h,j1}\nu_t^p + \gamma'\nu_t^W + \delta(L)Y_{t-1} + u_{j,t+h}^{(h)} \\ &= \Theta_{h,j1}p_t + \gamma' W_t + \delta(L)Y_{t-1} + u_{j,t+h}^{(h)} \end{aligned}$$

which is estimated separately at every horizon.

LP vs. SVAR

Purported pros and cons of LP vs. IV

Claimed pros and cons of LP regression (claims from various papers):

Pros of LP

- Can model nonlinearities
- Doesn't make lag length assumptions for full VAR
- Doesn't require invertibility
- Robust to misspecification

Cons of LP

- Less efficient asymptotically if the VAR restrictions are correct
- Provides non-smooth IRFs (the horizons aren't tied together) (see Smoothness constraints (Barnichon-Brownless (2016), Plagborg-Møller (2016), Miranda-Agrippino and Ricco (2017))

LP vs. SVAR, ctd

$$Y_{j,t+h} = \Theta_{h,j1} p_t + \gamma' W_t + \delta(L) Y_{t-1} + u_{j,t+h}^{(h)}$$

This derivation started with the SVAR and derived the LP. Is the LP method more generally applicable?

Consider the special case of $h = 0$ (impact effect):

$$Y_{j,t} = \Theta_{0,j1} p_t + \gamma_j{}' W_t + \delta(L) Y_{t-1} + u_{j,t}$$

or

$$\nu_{j,t} = \Theta_{0,j1} \nu_t^p + \gamma_j{}' \nu_t^W + u_{j,t}^{\perp}$$

The conditional mean independence condition for this impact regression is:

$$E\left(u_{j,t}^{\perp} \mid \nu_t^p, \nu_t^W\right) = E\left(u_{j,t}^{\perp} \mid \nu_t^W\right)$$

Reference: Kim and Kilian (2011), Plagborg-Møller and Wolf (2018b)

Outline

Part A

1. Dynamic causal effects: Overview
   a. Definition and conceptual framework
   b. Estimation when the shock is observed
   c. Identification and estimation when the shock is unobserved
2. Multivariate methods with internal identification: SVARs

Part B

3. Single equation methods with internal identification: LP
4. **Multivariate methods with external instruments: SVAR-IV**
5. Single-equation methods with external instruments: LP-IV
6. Summary

Identification by External Instruments: SVAR-IV

The external instrument approach entails finding some external information (outside the model) that is relevant (correlated with the shock of interest) and exogenous (uncorrelated with the other shocks).

***Example* 1**: The Cochrane- Piazessi (2002) shock (Z^{CP}) measures the part of the monetary policy shock revealed around a FOMC announcement – but not the shock revealed at other times. If CP's identification is sound, $Z^{CP} \neq \varepsilon_t^r$ but

(i) corr(ε_t^r,Z^{CP}) ≠ 0 (relevance)

(ii) corr(other shocks, Z^{CP}) = 0 (exogeneity)

***Example* 2**: Romer and Romer (1989, 2004, 2008); Ramey and Shapiro (1998); Ramey (2009) use the narrative approach to identify moments at which fiscal/monetary shocks occur. If identification is sound, $Z^{RR} \neq \varepsilon_t^r$ but

(i) corr(ε_t^r,Z^{RR}) ≠ 0 (relevance)

(ii) corr(other shocks, Z^{RR}) = 0 (exogeneity)

Some empirical papers that can be reinterpreted as external instruments

- **Monetary shock**: Cochrane and Piazzesi (2002), Faust, Swanson, and Wright (2003. 2004), Romer and Romer (2004), Bernanke and Kuttner (2005), Gürkaynak, Sack, and Swanson (2005)

- **Fiscal shock**: Romer and Romer (2010), Fisher and Peters (2010), Ramey (2011)

- **Uncertainty shock**: Bloom (2009), Baker, Bloom, and Davis (2011), Bekaert, Hoerova, and Lo Duca (2010), Bachman, Elstner, and Sims (2010)

- **Liquidity shocks**: Gilchrist and Zakrajšek's (2011), Bassett, Chosak, Driscoll, and Zakrajšek's (2011)

- **Oil shock**: Hamilton (1996, 2003), Kilian (2008a), Ramey and Vine (2010)

SVAR-IV

VAR: $Y_t = A(\mathrm{L})Y_{t-1} + \nu_t, \quad \nu_t = \Theta_0 \varepsilon_t$

Suppose you have an instrument Z_t which is correlated with ε_{1t} and uncorrelated with other shocks; specifically, Z_t satisfies,

Condition SVAR-IV

(i) $E\varepsilon_{1t}Z_t' = \alpha' \neq 0$ (relevance)

(ii) $E\varepsilon_{2:n,t}Z_t' = 0$ (exogeneity w.r.t. other current shocks)

Then

$$E\nu_t Z_t = E\left(\Theta_0 \varepsilon_t Z_t\right) = \Theta_0 E\begin{pmatrix} \varepsilon_{1t}Z_t' \\ \varepsilon_{2:n,t}Z_t' \end{pmatrix} = \Theta_0 \begin{pmatrix} \alpha' \\ 0 \end{pmatrix} = \begin{pmatrix} \Theta_{0,11}\alpha' \\ \Theta_{0,2:n,1}\alpha' \end{pmatrix} = \begin{pmatrix} \alpha' \\ \Theta_{0,2:n,1}\alpha' \end{pmatrix}.$$

so, with a single instrument, $\Theta_{0,2:n,1} = \dfrac{E\left(\nu_{2:n,t}Z_t\right)}{E\left(\nu_{1,t}Z_t\right)}$ and $\hat{\Theta}_{0,2:n,1}^{SVAR-IV} = \dfrac{\sum_{t=1}^{T} \nu_{2:n,t}Z_t}{\sum_{t=1}^{T} \nu_{1,t}Z_t}$

SVAR-IV, ctd.

IV interpretation:

$$\nu_{i,t} = \Theta_{0,i1}\varepsilon_{1,t} + \Theta_{0,i2}\varepsilon_{2,t} + \ldots + \Theta_{0,in_Y}\varepsilon_{n_Y,t}$$

$$= \Theta_{0,i1}\nu_{1,t} + \{\varepsilon_{2:n_Y,t}\} \quad \text{(using unit effect normalization)}$$

or

$$Y_{i,t} = \Theta_{0,i1}Y_{1,t} + \gamma_i(\mathrm{L})Y_{t-1} + \{\varepsilon_{2:n_Y,t}\}$$

SVAR-IV estimator: $\hat{\Theta}_{h,1}^{SVAR-IV} = \hat{\mathrm{C}}_h\hat{\Theta}_{0,1}^{SVAR-IV}$, $\hat{\mathrm{C}}(\mathrm{L}) = \hat{\mathrm{A}}(\mathrm{L})^{-1}$

Strong instrument asymptotics and inference

- Conventional delta method and IV formulas go through
- One implementation is the parametric bootstrap, where a time series process for Z_t is estimated, see Stock and Watson (2018)

Weak instrument asymptotics and inference

(Montiel Olea, Stock, and Watson (2018)) Weak IV asymptotic setup. Obtain weak instrument distribution, conduct robust inference

SVAR-IV, ctd.

***Example*: New car sales**

Recall the estimation equations:

$$p_t = \Theta_{0,pW} W_t + \delta_p(L) Y_{t-1} + \varepsilon_t^p$$

$$q_t = \Theta_{0,qp} p_t + \Theta_{0,pW} W_t + \delta_q(L) Y_{t-1} + \varepsilon_t^q$$

- The critique of this specification is that it requires auto dealers not to hold a clearance event if sales are weak this quarter – the dealer needs to wait until next quarter to hold the sale.
- Technically, the condition is $E\left(\varepsilon_t^q \mid \nu_t^p, \nu_t^W\right) = 0$
- If this condition fails, then p_t is correlated with ε_t^q, which is in the error term.

Instead, if there is an instrument for new car prices that satisfies Condition SVAR-IV (correlated with price shock, uncorrelated with other shocks), then $\Theta_{0,p}$ can be estimated using the instrument.

Outline

Part A

1. Dynamic causal effects: Overview

    a. Definition and conceptual framework

    b. Estimation when the shock is observed

    c. Identification and estimation when the shock is unobserved

2. Multivariate methods with internal identification: SVARs

Part B

3. Single equation methods with internal identification: LP

4. Multivariate methods with external instruments: SVAR-IV

5. **Single-equation methods with external instruments: LP-IV**

6. Summary

Single-equation IV estimation: LP-IV

Start with the basic LP regression for *h*-period effect of shock 1 on variable *j*:

$$Y_{i,t+h} = \Theta_{h,i1}\varepsilon_{1,t} + \{\varepsilon_{t+h},...,\varepsilon_{t+1},\varepsilon_{2:n_\varepsilon,t},\varepsilon_{t-1},\varepsilon_{t-2},...\} \qquad (*)$$

With the unit effect normalization $\nu_{1,t} = \varepsilon_{1,t} + \{\varepsilon_{2:n,t}\}$, so (*) can be written,

$$Y_{j,t+h} = \Theta_{h,j1}\nu_{1t} + \{\varepsilon_{t+h},...,\varepsilon_{t+1},\varepsilon_{2:n_\varepsilon,t},\varepsilon_{t-1},\varepsilon_{t-2},...\} \qquad (**)$$

In general, v_{1t} is correlated with the error term in (**) (simultaneous equations bias). But suppose you have an instrument Z_t that satisfies:

Condition LP-IV (Mertens' conditions)

(i) $E\left(\varepsilon_{1,t}Z_t'\right) = \alpha' \neq 0$ (relevance)

(ii) $E\left(\varepsilon_{2:n,t}Z_t'\right) = 0$ (contemporaneous exogeneity)

(iii) $E\left(\varepsilon_{t+j}Z_t'\right) = 0$ for $j \neq 0$ (lag & **lead** exogeneity).

Then $\Theta_{h,j1}$ in (**) can be estimated using Z_t as an IV (with HAR SEs).

LP-IV, ctd.

IV estimation of: $$Y_{j,t+h} = \Theta_{h,j1}\nu_{1t} + \{\ldots,\varepsilon_{t+1},\varepsilon_{2:n,t}\}$$

IV estimator: $$\hat{\Theta}_{h,j1}^{LP-IV} = \frac{\sum_{t=1}^{T} Y_{j,t+h} Z_t}{\sum_{t=1}^{T} \nu_{1t} Z_t}$$

Interpretation: Regress inflation 4 quarters hence on the FF rate, using the MP surprise as an instrument

Relation between SVAR-IV and LP-IV

SVAR-IV and LP-IV produce identical impact effects, but differ for $h \geq 1$:

$h = 0$: $$\hat{\Theta}_{0,1}^{LP-IV} = \begin{pmatrix} 1 \\ \dfrac{\sum_{t=1}^{T} \nu_{2:n,t} Z_t}{\sum_{t=1}^{T} \nu_{1,t} Z_t} \end{pmatrix} = \hat{\Theta}_{0,1}^{SVAR-IV}$$

$h \geq 1$: $$\hat{\Theta}_{h,1}^{SVAR-IV} = \hat{C}_h \hat{\Theta}_{0,1}^{SVAR-IV}, \quad \hat{C}(L) = \hat{A}(L)^{-1}$$

LP-IV with controls

Condition LP-IV(iii) is (very) strong, not even satisfied in the Gertler-Karadi application, however it might be satisfied after including some control variables.

Notation: $x_t^{\perp} = x_t - \mathrm{Proj}(x_t \mid W_t)$

LP-IV regression with controls: $Y_{i,t+h} = \Theta_{h,i1} Y_{1,t} + \gamma_h{}' W_t + u_{i,t+h}^{h\perp}$

Conditions for instrument validity with controls:

Condition LP-IV$^{\perp}$

(i) $E\left(\varepsilon_{1,t}^{\perp} Z_t^{\perp\prime}\right) = \alpha' \neq 0$

(ii) $E\left(\varepsilon_{2:n,t}^{\perp} Z_t^{\perp\prime}\right) = 0$

(iii) $E\left(\varepsilon_{t+j}^{\perp} Z_t^{\perp\prime}\right) = 0$ for $j \neq 0$.

LP-IV with controls, ctd

$$Y_{i,t+h} = \Theta_{h,i1} Y_{1,t} + \gamma_h{}' W_t + u_{i,t+h}^{h\perp}$$

Condition LP-IV$^{\perp}$ (iii) $E\left(\varepsilon_{t+j}^{\perp} Z_t^{\perp\prime}\right) = 0$ for $j \neq 0$.

What are the controls?

- ***Specific controls***: Gertler-Karadi construction of Z induced MA(1) structure:
  $Z_t = \delta_1 \varepsilon_{1,t} + \delta_1 \varepsilon_{1,t\text{-}1}$
  so $W_t = (Z_{t\text{-}1}, Z_{t\text{-}2},\ldots)$ and $Z_t^{\perp} = \delta_1 \varepsilon_{1,t}$

- ***Generic controls***: if Z_t depends on $\varepsilon_{1,t}$ and $\varepsilon_{t\text{-}1}$, $\varepsilon_{t\text{-}2},\ldots$
  $W_t = (Y_{t\text{-}1}, Y_{t\text{-}2},\ldots)$ (lagged observable macro variables)
  $W_t = (F_{t\text{-}1}, F_{t\text{-}2},\ldots)$ (lagged estimated factors – like FAVAR, but single-equation)

SVAR-IV vs. LP-IV

The purported pros and cons of SVAR-IV vs, LP-IV parallel those of SVAR vs. LP (e.g., LP-IV purportedly does not require correct specification of the VAR, LP-IV does not require invertibility, LP-IV is valid under nonlinearities).

However, there is a strong equivalency result available which basically says that in general, if lagged *Y*'s are required as controlled variables for an instrument to be valid, then Conditions SVAR-IV and LP-IV$^{\perp}$ are equivalent.

Invertibility and assumptions LP-IV, LP-IV$^{\perp}$, and SVAR-IV,

	LP-IV	LP-IV$^{\perp}$	SVAR-IV
	$Y_{i,t+h} = \Theta_{h,i1} Y_{1,t} + u_{t+h}^{h}$	$Y_{i,t+h} = \Theta_{h,i1} Y_{1,t} + \gamma_h{}' W_t + u_{i,t+h}^{h\perp}$	$Y_t = A(\mathrm{L}) Y_{t-1} + \nu_t$ $\varepsilon_t = Proj(\varepsilon_t \mid Y_t, Y_{t-1}, \ldots)$
(i)	$E\left(\varepsilon_{1,t} Z_t{}'\right) = \alpha' \neq 0$	$E\left(\varepsilon_{1,t}^{\perp} Z_t^{\perp\prime}\right) = \alpha' \neq 0$	$E\left(\varepsilon_{1t} Z_t{}'\right) = \alpha' \neq 0$
(ii)	$E\left(\varepsilon_{2:n,t} Z_t{}'\right) = 0$	$E\left(\varepsilon_{2:n,t}^{\perp} Z_t^{\perp\prime}\right) = 0$	$E\left(\varepsilon_{2:n,t} Z_t{}'\right) = 0$
(iii)	$E\left(\varepsilon_{t+j} Z_t{}'\right) = 0$ for $j \neq 0$	$E\left(\varepsilon_{t+j}^{\perp} Z_t^{\perp\prime}\right) = 0$ for $j \neq 0$	

Equivalence result (Stock & Watson (2018))

In LP-IV, let $W_t = (Y_{t-1}, Y_{t-2}, \ldots)$, and let $\boldsymbol{Z}$ denote the set of stochastic processes (candidate instruments) that satisfies LP-IV (i), (ii), and (iii) for $j > 0$. Then:

(a) Condition SVAR-IV is satisfied

(b) LP-IV$^{\perp}$ is satisfied for all $Z \in \boldsymbol{Z}$ if and only if invertibility holds.

i.e. LP-IV$^{\perp}$ = SVAR-IV + invertibility

Sketch of proof

In LP-IV, let $W_t = (Y_{t-1}, Y_{t-2},\ldots)$, and let $\boldsymbol{Z}$ denote the set of stochastic processes (candidate instruments) that satisfies LP-IV (i), (ii), and (iii) for $j > 0$. Then:

(a) Condition SVAR-IV is satisfied

(b) LP-IV$^{\perp}$ is satisfied for all $Z \in \boldsymbol{Z}$ if and only if invertibility holds.

(a) This is immediate because $E\left(\varepsilon_t^{\perp} Z_t^{\perp}\right) = E\left[\varepsilon_t\left(Z_t - \text{Proj}(Z_t \mid Y_{t-1}, Y_{t-2},\ldots)\right)\right] = E\left(\varepsilon_t Z_t\right)$

(b) Invertibility implies LP-IV$^{\perp}$(iii):

Under invertibility $\text{Proj}\left(\varepsilon_{t-j} \mid Y_{t-1}, Y_{t-2},\ldots\right) = \varepsilon_{t-j}$ so

$\varepsilon_{t-j}^{\perp} = \varepsilon_{t-j} - \text{Proj}\left(\varepsilon_{t-j} \mid Y_{t-1}, Y_{t-2},\ldots\right) = 0$ so $E\left(\varepsilon_{t-j}^{\perp} Z_t^{\perp}\right) = 0$.

LP-IV$^{\perp}$(iii) implies invertibility:

Consider AR(1) instrument case, $Z_t = \text{Proj}\left(Z_t \mid \varepsilon_{t-1}, \varepsilon_{t-2},\ldots\right) + \zeta_t = \Phi\varepsilon_{t-1} + \zeta_t$,

where ζ_t satisfies LP-IV$^{\perp}$(iii). If LP-IV$^{\perp}$(iii) is satisfied, then

$$0 = E\left(\varepsilon_{t-1}^{\perp} Z_t^{\perp}\right) = E\left[\varepsilon_{t-1}^{\perp}\left(\zeta_t + \Phi_1 \varepsilon_{t-1}\right)^{\perp}\right] = E\left(\varepsilon_{t-1}^{\perp} \varepsilon_{t-1}^{\perp\prime}\right)\Phi_1'$$

so $\varepsilon_{t-1}^{\perp} = \varepsilon_{t-1} - \text{Proj}\left(\varepsilon_{t-1} \mid Y_{t-1}, Y_{t-2},\ldots\right) = 0$.

Hausman-type test for invertibility

Suppose an instrument satisfies LP-IV, or LP-IV$^{\perp}$ with specific controls (not generic controls). Then SVAR-IV is satisfied.

If invertibility holds, then SVAR-IV is more efficient

Hausman-type test statistic: $T(\hat{\theta}^{LP-IV}-\hat{\theta}^{SVAR-IV})'\hat{V}^{-1}(\hat{\theta}^{LP-IV}-\hat{\theta}^{SVAR-IV})$, $h > 1$

Idea that invertible and noninvertible IRFs can be close to each other (Beaudry et al (2015), Plagbog-Møller (2016)) suggest the null and local alternative,

$$\Theta_{h,1} = C_h \Theta_{0,1} + T^{-1/2} d_h$$

Under the local alternative, $\sqrt{T}\left(\hat{\theta}^{LP-IV}-\hat{\theta}^{SVAR-IV}\right) \xrightarrow{d} \mathrm{N}(d,V)$

Variance matrix computed by parametric bootstrap

Forecast error variance decompositions and historical decompositions

FEVD is fraction of *h*-step ahead forecast error for variable *i* that is explained by shock 1.

$$FEVD_{h,i1} = \frac{\sum_{k=0}^{h-1} \Theta_{k,i1}^2 \sigma_{\varepsilon_1}^2}{\operatorname{var}\left(Y_{i,t+h} \mid \varepsilon_t, \varepsilon_{t-1}, \ldots\right)}$$

- This answers: How important is shock 1?
- Computing this requires identification of $\sigma_{\varepsilon_1}^2$ and $\operatorname{var}\left(Y_{i,t+h} \mid \varepsilon_t, \varepsilon_{t-1}, \ldots\right)$ in addition to SIRF.
- Sufficient condition is invertibility and identification of $\Theta_{0,1}$:

$$\varepsilon_{1,t} = \lambda' \nu_t \text{, where } \lambda = \Theta_{0,1}{}' \Sigma_{\nu\nu}^{-1} \Big/ \left(\Theta_{0,1}{}' \Sigma_{\nu\nu}^{-1} \Theta_{0,1}\right)$$

$$\sigma_{\varepsilon_1}^2 = \left(\Theta_{0,1}{}' \Sigma_{\nu\nu}^{-1} \Theta_{0,1}\right)^{-1}$$

and $\operatorname{var}\left(Y_{i,t+h} \mid \varepsilon_t, \varepsilon_{t-1}, \ldots\right) = \operatorname{var}\left(Y_{i,t+h} \mid Y_t, Y_{t-1}, \ldots\right)$

[First line: $\Theta_{0,1}{}' \Sigma_{\nu\nu}^{-1} \nu_t = \Theta_{0,1}{}' \left(\Theta_0 \Sigma_{\varepsilon\varepsilon} \Theta_0{}'\right)^{-1} \nu_t = e_1{}' \Sigma_{\varepsilon\varepsilon}^{-1} \varepsilon_t = \varepsilon_{1,t} / \sigma_{\varepsilon_1}^2$]

- Estimation: Gorodnichenko and Lee (2017)
- Bounds: Plagborg-Møller and Wolf (2017)

SVAR-IV and LP-IV Example: Gertler-Karadi (2015)

$Y_t = (\Delta\ln IP_t, \Delta\ln CPI_t, \text{1Yr Treasury rate}_t, \text{EBP}_t)$

EBP_t = Gilchrist-Zakrajšek (2012) Excess Bond Premium

z_t = "Announcement surprise"
= change in 4-week Fed Funds Futures around FOMC announcement windows

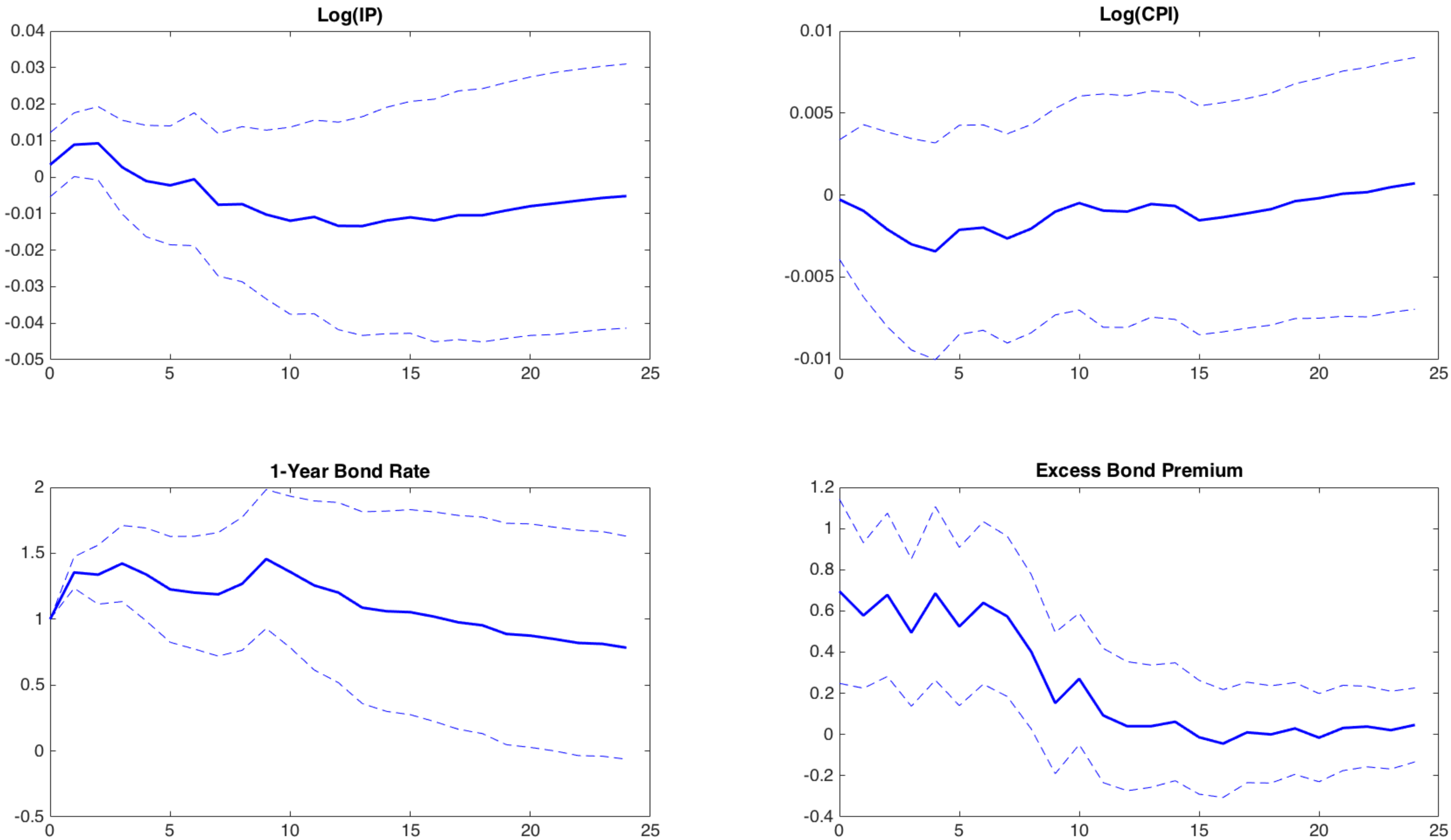
Sample period: 1990m1-2012m6 (monthly)

SVAR-IV: GK specification: 12 lag VAR

LP-IV: $W_t = Y_{t\text{-}1},\ldots, Y_{t\text{-}4}, z_{t\text{-}1},\ldots, z_{t\text{-}4}$

Gertler-Karadi example, ctd.

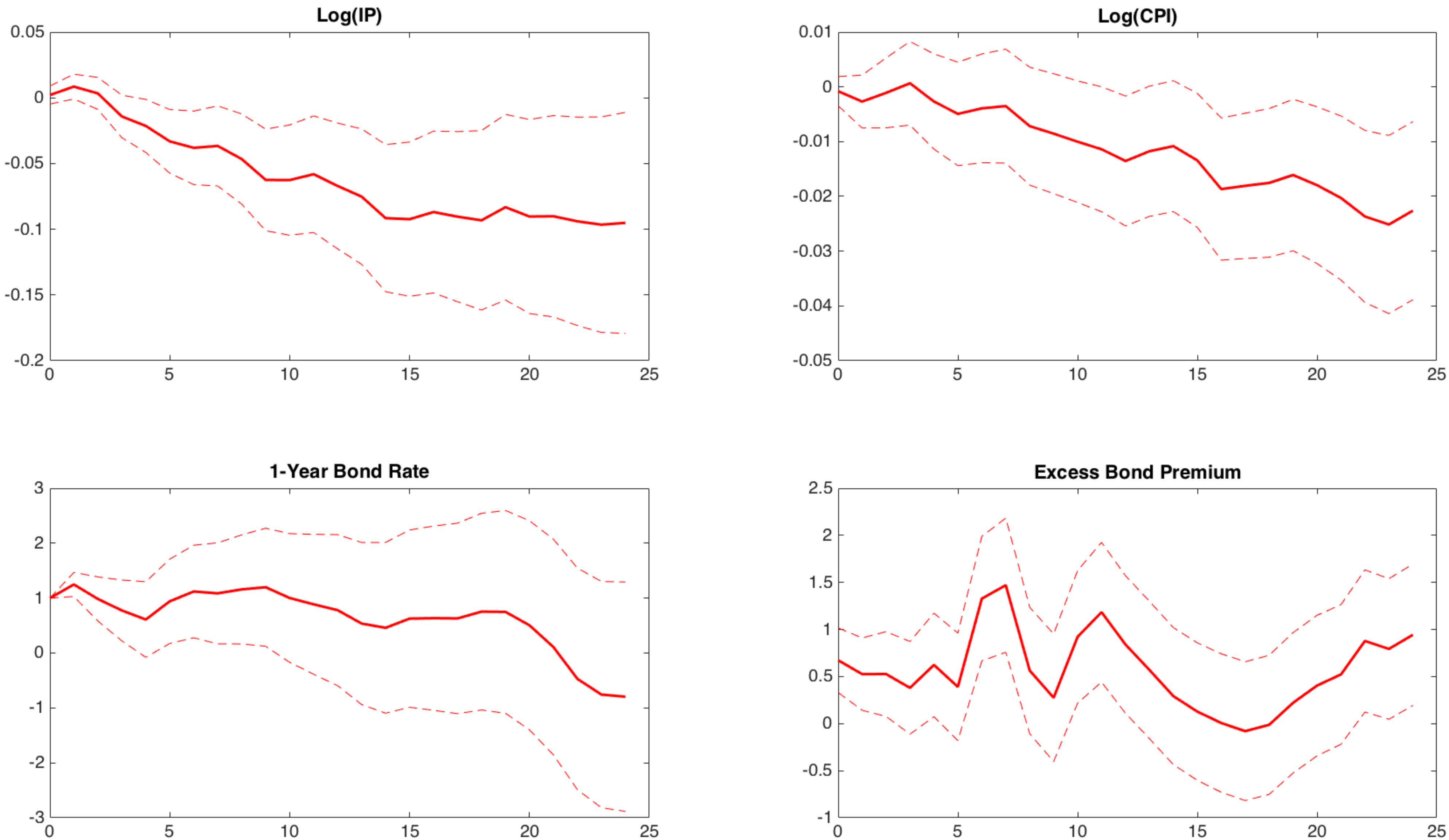
Cumulative IRFs: **SVAR-IV** with ±1 SE bands



Gertler-Karadi example, ctd.

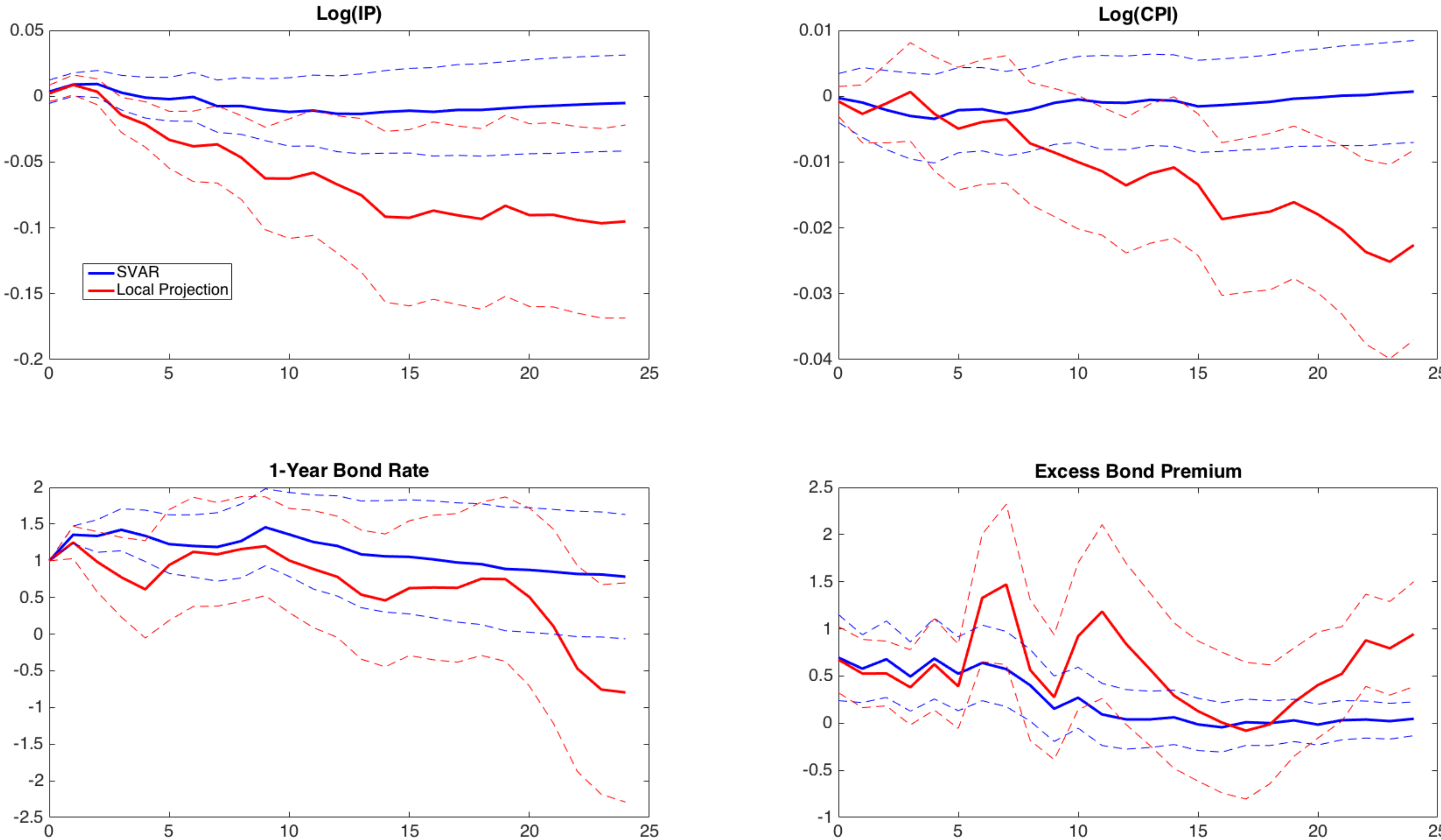
Cumulative IRFs: **LP-IV** with ±1 SE bands

W = 4 lags of Y, z



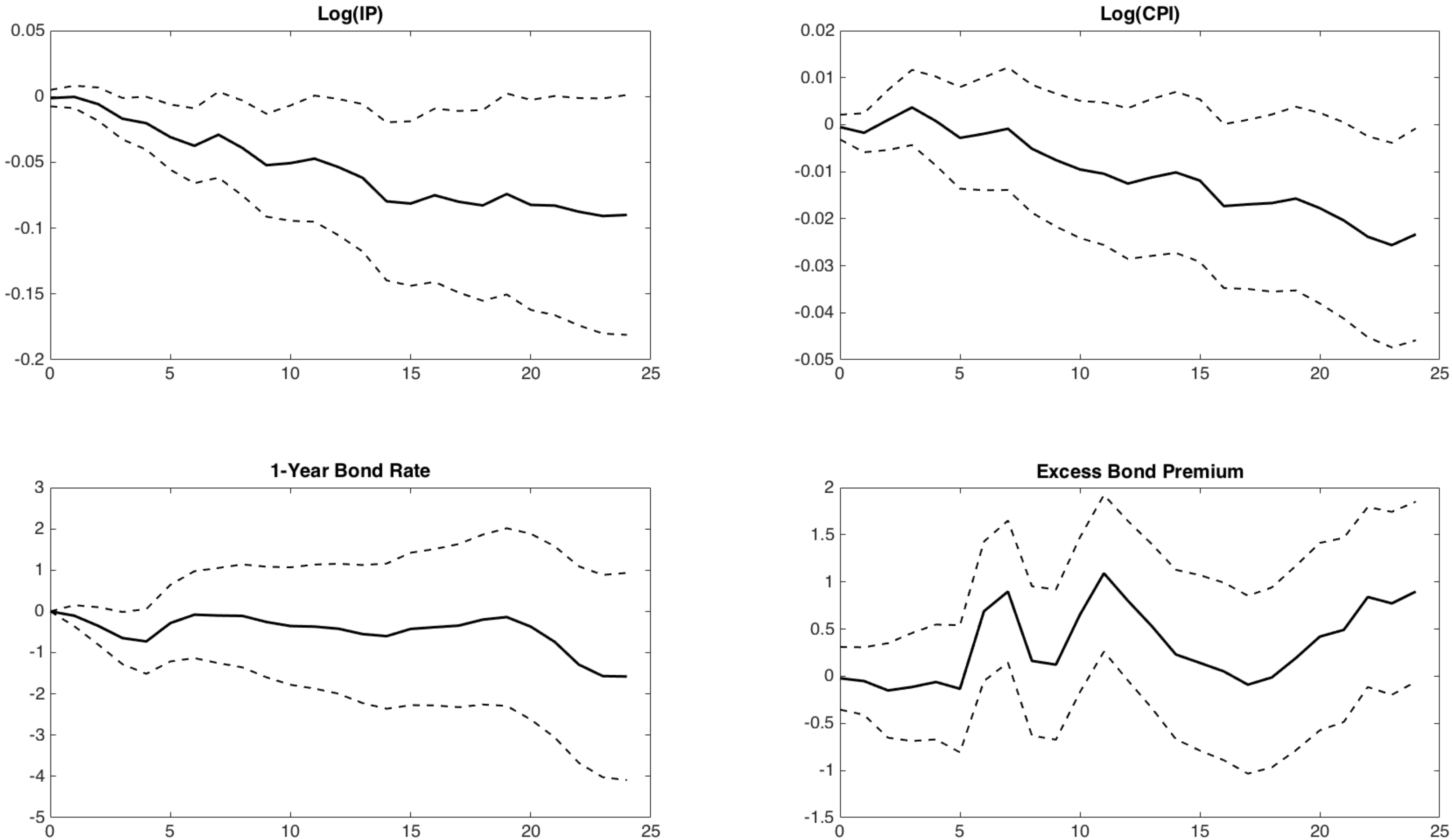
Gertler-Karadi example, ctd.

Cumulative IRFs: **SVAR-IV** and **LP-IV** and ±1 SE bands (parametric bootstrap)



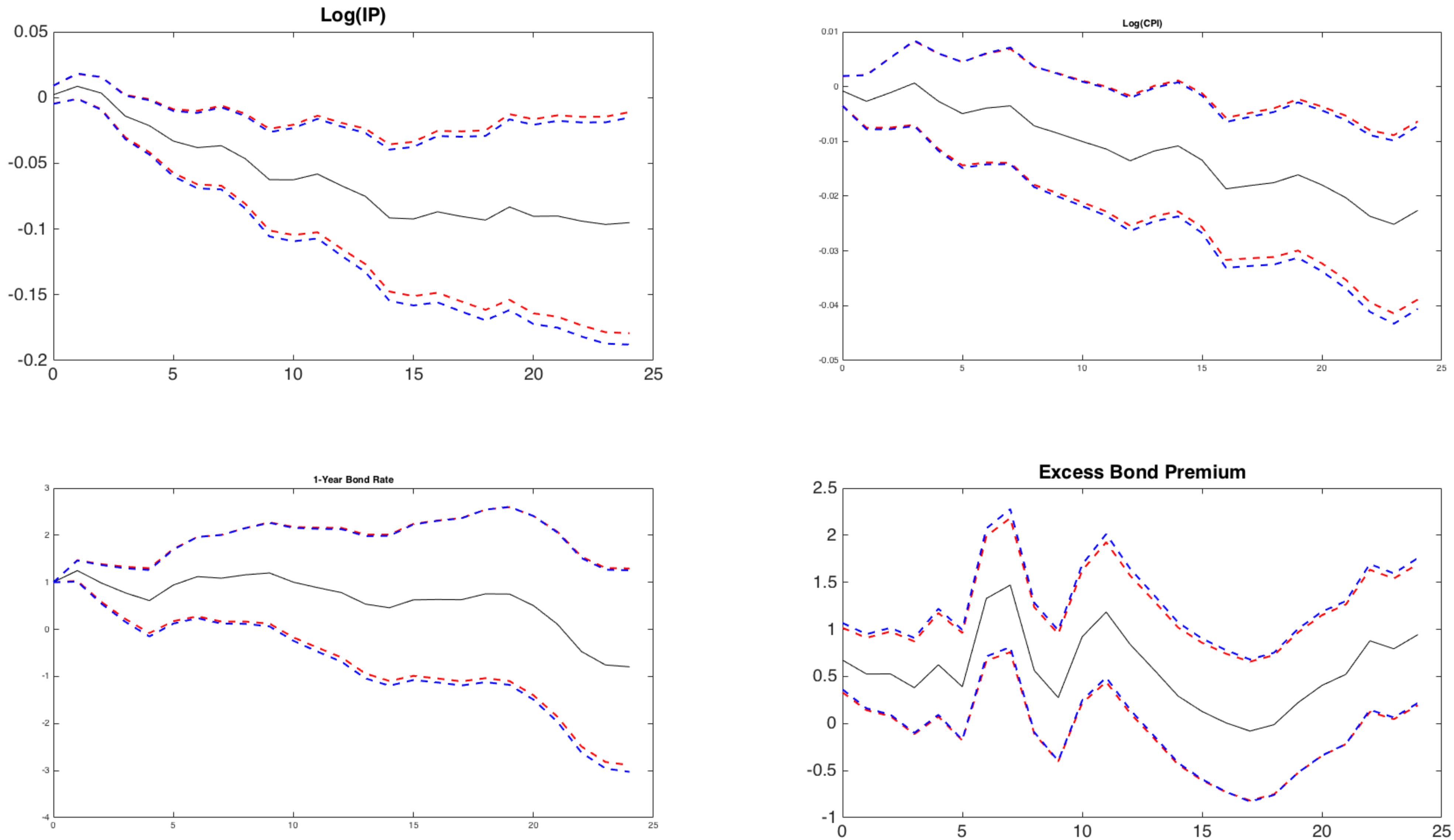
Gertler-Karadi example, ctd.

Test statistics by horizon by variable: entries are *t*-statistics $\Psi_T / \sqrt{\hat{V}_h}$



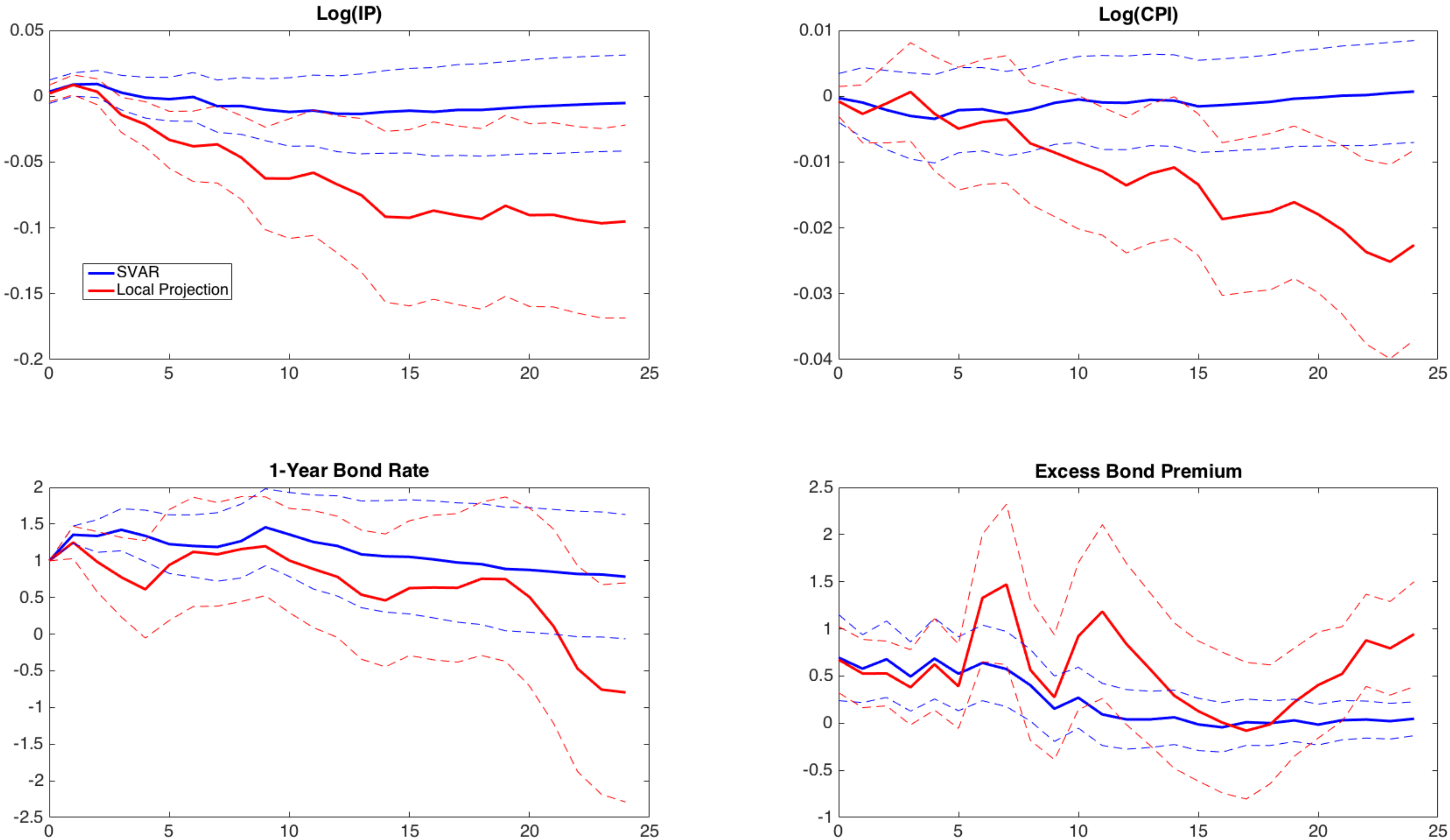
Gertler-Karadi example, ctd.

LP-IV 68% bands: **±1 SE** and **Anderson-Rubin Confidence Interval**



Gertler-Karadi example, ctd.

Cumulative IRFs: **SVAR-IV** and **LP-IV** and ±1 SE bands (parametric bootstrap)



Gertler-Karadi example, ctd.

Table 2: Tests for VAR Invertibility (*p*-values)

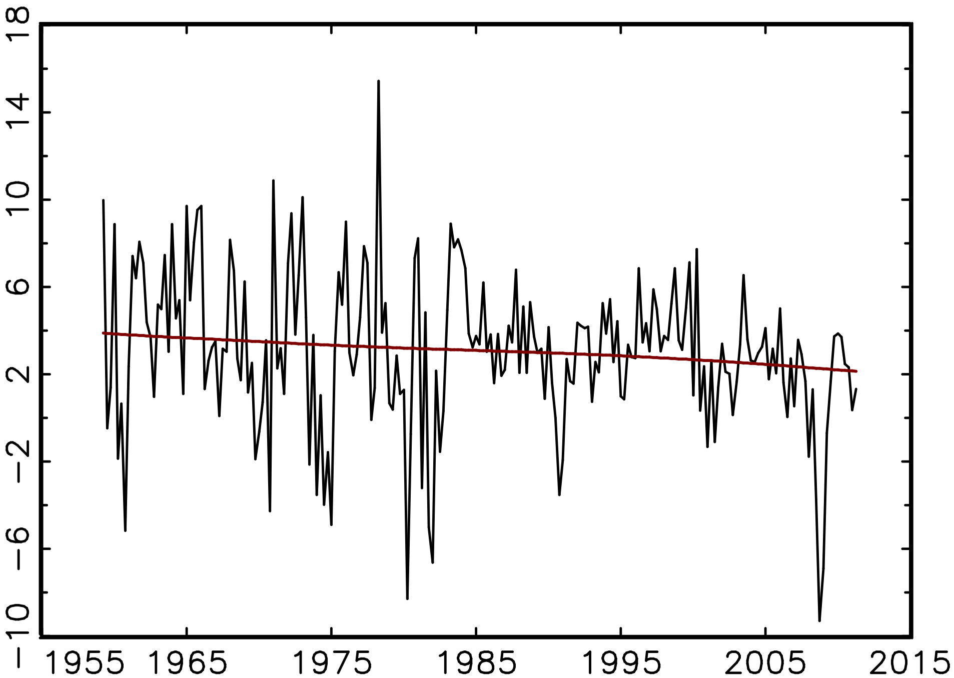
	1Year Rate	**ln(IP)**	**ln(CPI)**	**GZ EBP**
VAR-LP difference (lags 0,6,12,24)	0.95	0.55	0.75	0.26
VAR Z-GC test	0.16	0.09	0.38	0.97

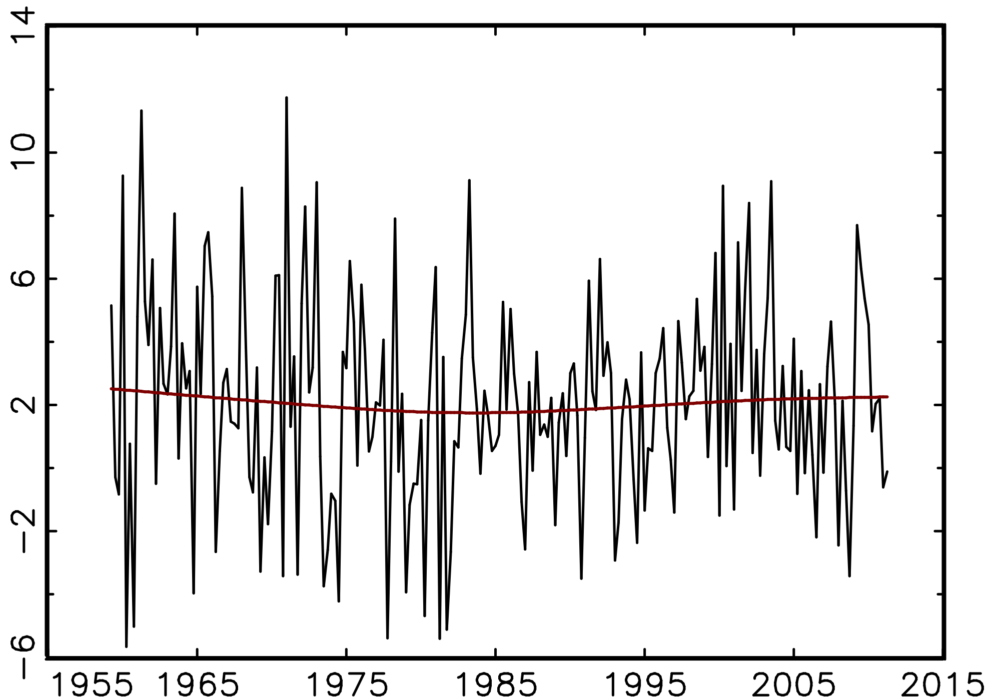
Notes: The first row is the bootstrap *p*-value for the test of the null hypothesis that IV-LP and IV-SVAR causal effects are same for h = 0, 6, 12, and 24 (test for invertibility). The second row shows *p*-values for the *F*-statistic testing the null hypothesis that the coefficients on four lags of *Z* are jointly equal to zero in each of the VAR equations.

SVAR-IV Empirical Application #2: Stock-Watson (BPEA, 2012)

Dynamic factor model identified by external instruments:

- U.S., quarterly, 1959-2011Q2, 200 time series
- Almost all series analyzed in changes or growth rates
- All series detrended by local demeaning – approximately 15 year centered moving average:





Quarterly GDP growth (a.r.) Quarterly productivity growth

Trend: 3.7% → 2.5% 2.3% → 1.8% → 2.2%

Instruments

1. Oil Shocks
   a. Hamilton (2003) net oil price increases
   b. Killian (2008) OPEC supply shortfalls
   c. Ramey-Vine (2010) innovations in adjusted gasoline prices

2. Monetary Policy
   a. Romer and Romer (2004) policy
   b. Smets-Wouters (2007) monetary policy shock
   c. Sims-Zha (2007) MS-VAR-based shock
   d. Gürkaynak, Sack, and Swanson (2005), FF futures market

3. Productivity
   a. Fernald (2009) adjusted productivity
   b. Gali (200x) long-run shock to labor productivity
   c. Smets-Wouters (2007) productivity shock

Instruments, ctd.

4. Uncertainty
   a. VIX/Bloom (2009)
   b. Baker, Bloom, and Davis (2009) Policy Uncertainty

5. Liquidity/risk
   a. Spread: Gilchrist-Zakrajšek (2011) excess bond premium
   b. Bank loan supply: Bassett, Chosak, Driscoll, Zakrajšek (2011)
   c. TED Spread

6. Fiscal Policy
   a. Ramey (2011) spending news
   b. Fisher-Peters (2010) excess returns gov. defense contractors
   c. Romer and Romer (2010) “all exogenous” tax changes.

"First stage": F_1: regression of Z_t on u_t, F_2: regression of u_{1t} on Z_t

Structural Shock	F_1	F_2
1. Oil		
Hamilton	2.9	**15.7**
Killian	1.1	1.6
Ramey-Vine	1.8	0.6
2. Monetary policy		
Romer and Romer	4.5	**21.4**
Smets-Wouters	9.0	5.3
Sims-Zha	6.5	**32.5**
GSS	0.6	0.1
3. Productivity		
Fernald TFP	**14.5**	**59.6**
Smets-Wouters	7.0	**32.3**
Structural Shock	F_1	F_2
4. Uncertainty		
Fin Unc (VIX)	**43.2**	**239.6**
Pol Unc (BBD)	**12.5**	73.1

5. Liquidity/risk	F_1	F_2
GZ EBP Spread	4.5	**23.8**
TED Spread	**12.3**	**61.1**
BCDZ Bank Loan	4.4	4.2
6. Fiscal policy		
Ramey Spending	0.5	1.0
Fisher-Peters Spending	1.3	0.1
Romer-Romer Taxes	0.5	2.1

Correlations among selected structural shocks

	O_K	M_{RR}	M_{SZ}	P_F	U_B	U_{BBD}	S_{GZ}	B_{BCDZ}	F_R	F_{RR}
O_K	1.00									
M_{RR}	**0.65**	1.00								
M_{SZ}	0.35	0.93	1.00							
P_F	0.30	0.20	0.06	1.00						
U_B	-0.37	-0.39	-0.29	0.19	1.00					
U_{BBD}	0.11	-0.17	-0.22	-0.06	0.78	1.00				
L_{GZ}	-0.42	-0.41	-0.24	0.07	**0.92**	**0.66**	1.00			
L_{BCDZ}	0.22	0.56	0.55	-0.09	**-0.69**	-0.54	-0.73	1.00		
F_R	**-0.64**	**-0.84**	**-0.72**	-0.17	0.26	-0.08	0.40	-0.13	1.00	
F_{RR}	0.15	**0.77**	**0.88**	0.18	0.01	-0.10	0.02	0.19	-0.45	1.00

Oil_{Kilian}	oil – Kilian (2009)
M_{RR}	monetary policy – Romer and Romer (2004)
M_{SZ}	monetary policy – Sims-Zha (2006)
P_F	productivity – Fernald (2009)
U_B	Uncertainty – VIX/Bloom (2009)
U_{BBD}	uncertainty (policy) – Baker, Bloom, and Davis (2012)
L_{GZ}	liquidity/risk – Gilchrist-Zakrajšek (2011) excess bond premium
L_{BCDZ}	liquidity/risk – BCDZ (2011) SLOOS shock
F_R	fiscal policy – Ramey (2011) federal spending
F_{RR}	fiscal policy – Romer-Romer (2010) federal tax

Selected literature on external instruments

<u>SVAR-IV</u>

Stock (2008), Stock and Watson (2012), Montiel Olea, Stock and Watson (2018), Mertens and Ravn (2013, 2014), Gertler and Karadi (2015), Caldera and Kamps (2017), Lumsford (2015), Carriero, Momtaz, Theodoridis, and Theophilopoulou (2015), Jentsch and Lunsford (2016), …

<u>Local-projections (LP-IV)</u>

Jordà, Schularick, and Taylor (2015), Ramey and Zubairy (2017), Ramey (2016), Mertens (2015), Fieldhouse, Mertens, Ravn (2017), Mertens (2015) lecture notes, Fieldhouse, Mertens, Ravn (2017), Plagborg-Møller and Wolf (2017), Gorodnichenko and Lee (2017)

Outline

Part A

1. Dynamic causal effects: Overview
   a. Definition and conceptual framework
   b. Estimation when the shock is observed
   c. Identification and estimation when the shock is unobserved
2. Multivariate methods with internal identification: SVARs

Part B

3. Single equation methods with internal identification: LP
4. Multivariate methods with external instruments: SVAR-IV
5. Single-equation methods with external instruments: LP-IV
6. **Summary**

Summary

1. Within the context of SVAR identification using internal restrictions (requiring invertibility), recent work has focused on inference with sign restrictions and on identification by heteroskedasticity.
   a. In the case of sign restrictions, the standard Bayesian algorithm produces results that are troubling from both a Bayesian and frequentist perspective, in which informative priors are imposed through a nonlinear transformation of flat priors on the space of orthonormal matrices.
   b. In the case of identification by heteroskedasticity, such identification seems to be generic, if there is heteroskedasticity (e.g., ARCH or stochastic volatility), without knowing the driving process. The shocks are then identified up to their “name”.

2. Otherwise, recent work on identification and estimation of dynamic causal effects methods have focused on bringing the rigor of the microeconometric identification revolution to macroeconomics.

Summary, ctd.

3. One branch is formulating shocks and dynamic causal effects from the perspective of potential outcomes and experimental treatments.

4. Another branch, which has already produced a substantial number of papers, is using IV methods to identify and to estimate dynamic causal effects.

5. At the heart of these new IV methods (SVAR-IV and LP-IV) is finding external information – an instrument that is correlated with the shock of interest, but not other shocks.

6. LP and LP-IV initially appear to require weaker assumptions than SVAR and SVAR-IV, but this is not the case. Outside of the measured shock case, identification by timing or by instrument conditions in general requires invertibility for the LP conditioning set to deliver conditional mean independence. But if invertibility holds, SVAR and SVAR-IV are valid and are more efficient asymptotically.

References

Aikman, D., O. Bush, and A.M. Taylor (2016), "Monetary versus Macroprudential Policies: Causal Impacts of Interest Rates and Credit Controls in the Era of the UK Radcliffe Report." *NBER Working Paper* 22380.

Angrist, J.D., Ò. Jordà, and G.M. Kuersteiner (2017). "Semiparametric Estimates of Monetary Policy Effects: String Theory Revisited," *Journal of Business and Economic Statistics*, forthcoming.

Barnichon, R. and C. Brownlees (2016). "Impulse Response Estimation by Smooth Local Projections," CEPR Discussion paper DP11726.

Beaudry, P., P. Fève, A. Guay, and F. Portier (2015). "When is Nonfundamentalness in VARs a Real Problem? An Application to News Shocks," manuscript, University of British Columbia.

Beaudry, P. and M. Saito (1998), "Estimating the Effects of Monetary Policy Shocks: An Evaluation of Different Approaches," *Journal of Monetary Economics* 42, 241-260.

Bernanke, B.S. and K.N. Kuttner (2005). "What Explains the Stock Market's Reaction to Federal Reserve Policy?", *The Journal of Finance* 40, 1221-1257.

Bojinov, I. and N. Shephard (2017). "Time Series Experiments, Causal Estimands, Exact *p*-values, and Trading." manuscript, Harvard University.

Caldara, D. and C. Kamps (2107). "The Analytics of SVARs: A Unified Framework to Measure Fiscal Multipliers," forthcoming, *The Review of Economic Studies*.

Chahrour, R. and K. Jurado (2017). "Recoverability," manuscript, Duke University.

Cochrane, J.H., and M. Piazzesi (2002). "The Fed and Interest Rates: A High-Frequency Identification," *American Economic Review* 92 (May), 90-95.

Faust, J., Rogers, J.H., Swanson, E., and J.H. Wright (2003). "Identifying the Effects of Monetary Policy Shocks on Exchange Rates Using High Frequency Data," *Journal of the European Economic Association* 1(5), 1031-57.

Fernández-Villaverde, J., J.F. Rubio-Ramírez, T.J. Sargent, and M.W. Watson (2007). “The ABCs (and Ds) of Understanding VARs” *American Economic Review*, Vol. 97, No. 3, 1021-1026.

Fieldhouse, A., K. Mertens, and M.O. Ravn (2017). “The Macroeconomic Effects of Government Asset Purchases: Evidence from Postwar U.S. Housing Credit Policy,” NBER Working Paper 23154.

Forni, M. and L. Gambetti (2014). “Sufficient Information in Structural VARs,” *Journal of Monetary Economics* 66C), 124-136.

Gertler, M. and P. Karadi (2015). “Monetary Policy Surprises, Credit Costs, and Economic Activity.” *American Economic Journal: Macroeconomics* 7, 44-76.

Gilchrist, S. and E. Zakrajšek (2012). “Credit Spreads and Business Cycle Fluctuations,” *American Economic Review* 102 (4), 1692-1720.

Gürkaynak, R.S., Sack, B., and E. Swanson (2005). “The sensitivity of long-term interest rates to economic news: Evidence and implications for macroeconomic models,” *American Economic Review* 95, 425–436.

Hamilton, J. D. (2003) “What Is an Oil Shock?” *Journal of Econometrics* 113: 363–98.

Hausman, J.A. (1978), "Specification Tests in Econometrics," *Econometrica*, 46(6), 1251-1271.

Hayashi, F., *Econometrics*. Princeton: Princeton University Press, 2000.

Imbens, G. (2014). “Instrumental Variables: An Econometrician’s Perspective.” *Statistical Science* 29, 323-58.

Jordà, Ò. (2005). “Estimation and Inference of Impulse Responses by Local Projections,” *American Economic Review* 95(1), 161-182.

Jordà, Ò., M. Schularick, and A.M. Taylor (2015). “Betting the House,” *Journal of International Economics* 96, S2-S18.

Jordà, Ò., M. Schularick, and A.M. Taylor (2017). “Large and State-Dependent Effects of Quasi-Random Monetary Experiments,” NBER Working Paper 23074.

Kilian, L. (2008). “Exogenous Oil Supply Shocks: How Big Are They and How Much Do They Matter for the U.S. Economy?” *Review of Economics and Statistics* 90, 216–40.

Kilian, L. and H. Lütkepohl (2018). *Structural Vector Autoregressive Analysis*. Cambridge UK: Cambridge University Press.

Kim, Y.J. and L. Kilian (2011). "How Reliable are Local Projection Estimators of Impulse Responses?" *The Review of Economics and Statistics* 93, 1460-1466.

Kuttner, K.N. (2001). "Monetary Policy Surprises and Interest Rates: Evidence from the Fed Funds Futures Market." *Journal of Monetary Economics* 47, 523-544.

Lechner, M. (2009). "Sequential Causal Models for the Evaluation of Labor Market Programs," *Journal of the American Statistical Association* 27, 71-83.

Mertens, K. (2015). "Bonn Summer School – Advances in Empirical Macroeconomics, Lecture 2" (slide deck).

Mertens, K. and M.O. Ravn (2013). "The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States," *American Economic Review* 103: 1212-1247.

Miranda-Agrippino, S. and G. Ricco (2017). "The Transmission of Monetary Policy Shocks." Manuscript, Department of Economics, University of Warwick.

Montiel Olea, J.L. and C. Pflueger (2013). "A Robust Test for Weak Instruments," *Journal of Business and Economic Statistics* 31, 358-369.

Montiel Olea, J., J.H. Stock, and M.W. Watson (2018). "Inference in Structural VARs Identified with an External Instrument," manuscript, Harvard University

Plagborg-Møller, M. (2016a), "Estimation of Smooth Impulse Response Functions," manuscript, Princeton University.

Plagborg-Møller, M. (2016b), "Bayesian Inference on Structural Impulse Response Functions," manuscript, Princeton University.

Plagborg-Møller, M. and C. Wolf (2017), "On Structural Inference with External Instruments," manuscript, Princeton University.

Plagborg-Møller, M. and C. Wolf (2018), "Local Projections and VARs Estimate the Same Impulse Response," manuscript, Princeton University.

Ramey, V. (2011). "Identifying government spending shocks: it's all in the timing," *Quarterly Journal of Economics* 126, 1–50

Ramey, V. (2016). "Macroeconomic Shocks and their Propagation," *Handbook of Macroeconomics, Vol. 2A.* Amsterdam: Elsevier, pp. 71-162.

Ramey, V. and S. Zubairy, (2017). "Government Spending Multipliers in Good Times and in Bad: Evidence from U.S. Historical Data," *Journal of Political Economy*, forthcoming.

Romer, C.D. and D.H. Romer (1989). "Does Monetary Policy Matter? A New Test in the Spirit of Friedman and Schwartz" (with discussion), in O.J. Blanchard and S. Fischer (eds.), *NBER Macroeconomics Annual* 1989. Cambridge, MA: MIT Press, 121-70.

Romer, C.D. and D.H. Romer (2010). "The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Shocks," *American Economic Review* 100, 763-801.

Sims, C.A. (1980). "Macroeconomics and Reality," *Econometrica* 48, 1-48.

Stock, J.H. (2008). *What's New in Econometrics: Time Series, Lecture 7*. Short course lectures, NBER Summer Institute, at http://www.nber.org/minicourse_2008.html.

Stock, J.H. and M.W. Watson (2012). "Disentangling the Channels of the 2007-09 Recession." *Brookings Papers on Economic Activity*, no. 1: 81-135.

Stock, J.H. and M.W. Watson (2016). "Dynamic Factor Models, Factor-Augmented Vector Autoregressions, and Structural Vector Autoregressions in Macroeconomics," *Handbook of Macroeconomics, Vol. 2A.* Amsterdam: Elsevier, pp. 415-525.

Stock, J.H. and M.W. Watson (2018). "Identification and Estimation of Dynamic Causal Effects in Macroeconomics Using External Instruments," *Economic Journal* 128, 917-948.

Stock, J.H. and M.W. Watson (2018b). *Introduction to Econometrics*, 4th Edition. Pearson.

AEA Continuing Education Course
Time Series Econometrics

Lecture 4

Weak Instruments: Beyond the i.i.d. Homoskedastic Case*

James H. Stock
Harvard University

January 7, 2019, 2:30-4:00pm

*This lecture draws on the 2018 NBER Summer Institute mini-course, "Weak Instruments and What To Do About Them" taught by Isaiah Andrews and James Stock, and the associated paper by Andrews, Stock, and Sun (2018), "Weak Instruments in IV Regression: Theory and Practice" at https://scholar.harvard.edu/stock/publications/weak-instruments-iv-regression-theory-and-practice.

Overview and Summary

Topic: IV regression with a single included endogenous regressor, control variables, and non-homoskedastic errors.

- This covers heteroskedasticity, HAC, cluster, etc.
- We assume that consistent robust SEs exist for the reduced form & first stage regressions (*Note*: This means HAC, not HAR! old bandwidth rule.)
- Early literature (through ~2006): homoskedastic case
- This lecture focuses on weak instruments in the non-homoskedastic case (i.e., the relevant case).

Outline

1. So what?
2. Detecting weak instruments
3. Estimation (brief)
4. Weak-instrument robust inference
5. Extensions

1. So What?

So what? (1) Theory

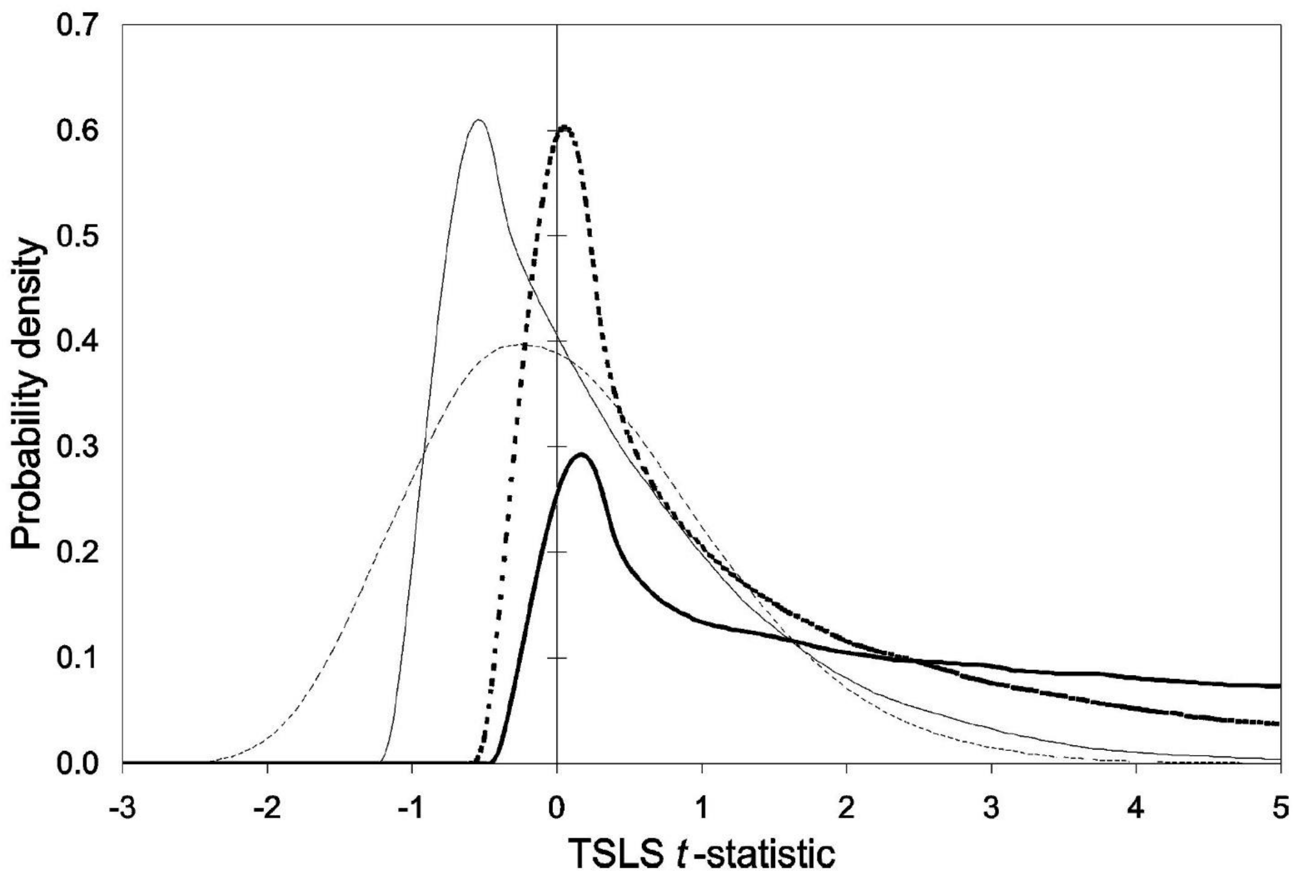
An instrumental variable is weak if its correlation with the included endogenous regressor is "small".

- "small" depends on the inference problem at hand, and on the sample size

With weak instruments, TSLS is biased towards OLS, and TSLS tests have the wrong size.

Distribution of the TSLS *t*-statistic (Nelson-Startz (1990a,b))

- Dark line = irrelevant instruments
- dashed light line = strong instruments
- intermediate cases = weak instruments

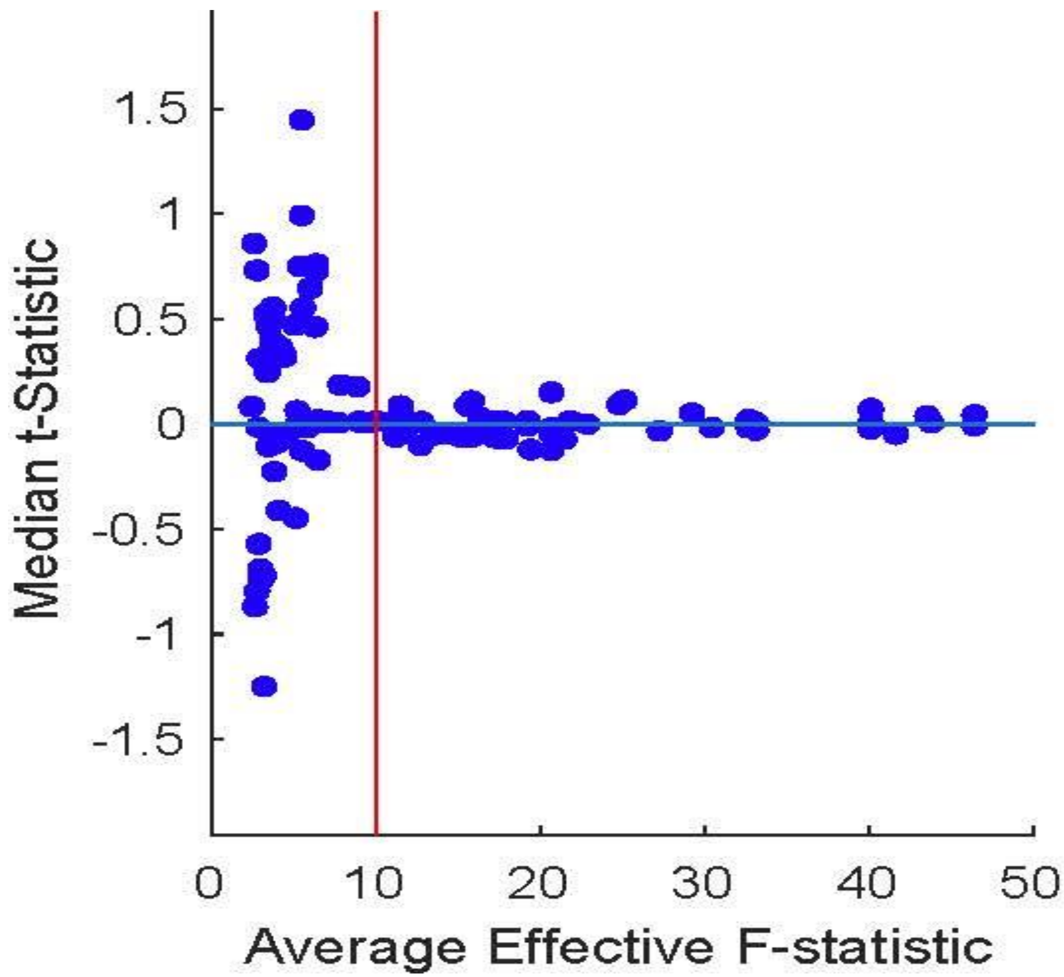


So what? (2) Simulation

DGP: 8 AER papers 2014-2018

(Sample: 17 that use IV; 16 with a single X; 8 in simulation sample)

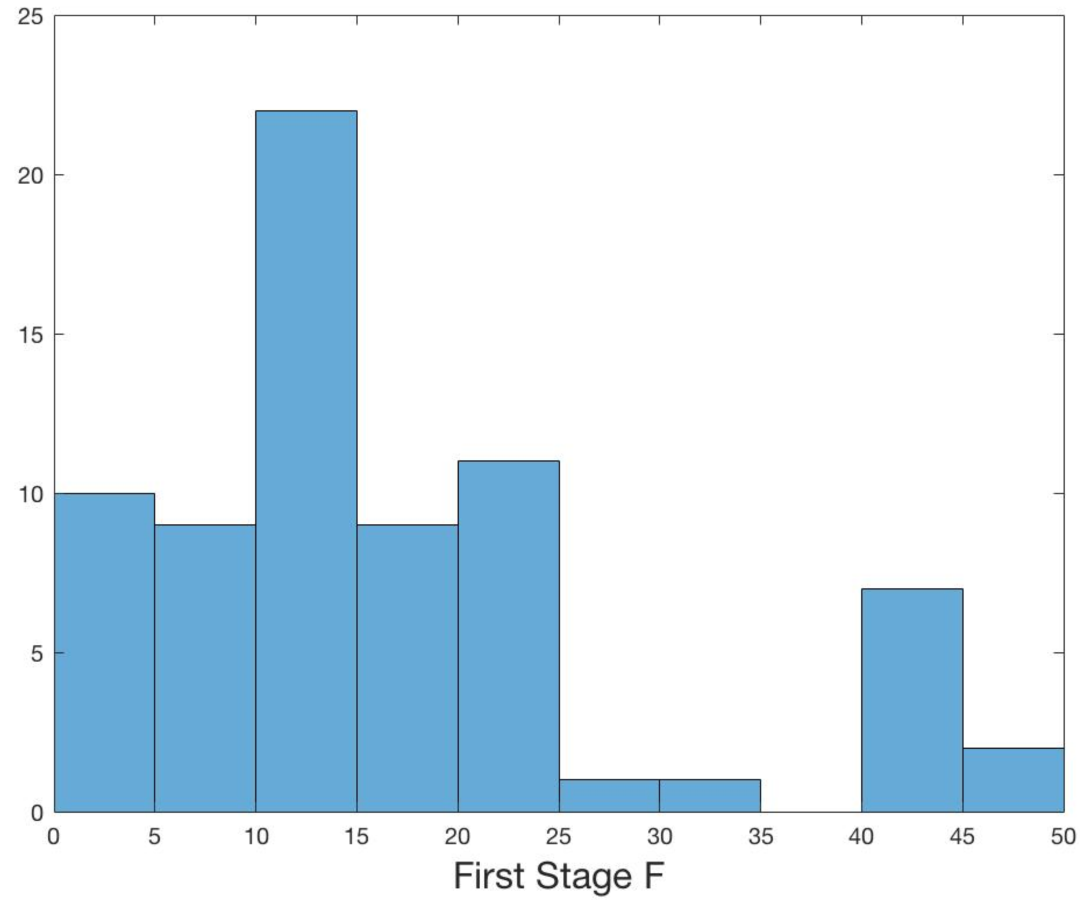
Median of TSLS t-statistic under the null



So what? (3) Practice

Histogram of first-stage *F*s in AER papers (108 specifications), 2014-2018

- The first-stage *F* tests the hypothesis that the first-stage coefficients are zero.
- Of the 17 papers, all but 1 report first-stage *F*s for at least one specification; the histogram is of the 108 specifications that report a first-stage *F* (72 of which are <50 and are in the plot).
- ***Great*** that authors/editors/referees are aware of the potential importance of weak instruments, as evidence by nearly all papers reporting first stages Fs!
- The spike at $F = 10$ is "interesting"



2. Detecting Weak Instruments

It is convenient to have a way to decide if instruments are strong (TSLS "works") or weak (use weak-instrument robust methods).

The standard method is "the" first-stage F. Candidates:

F^N – nonrobust

F^R – robust (HR, HAC, cluster), also called Kleibergen-Paap (2006)

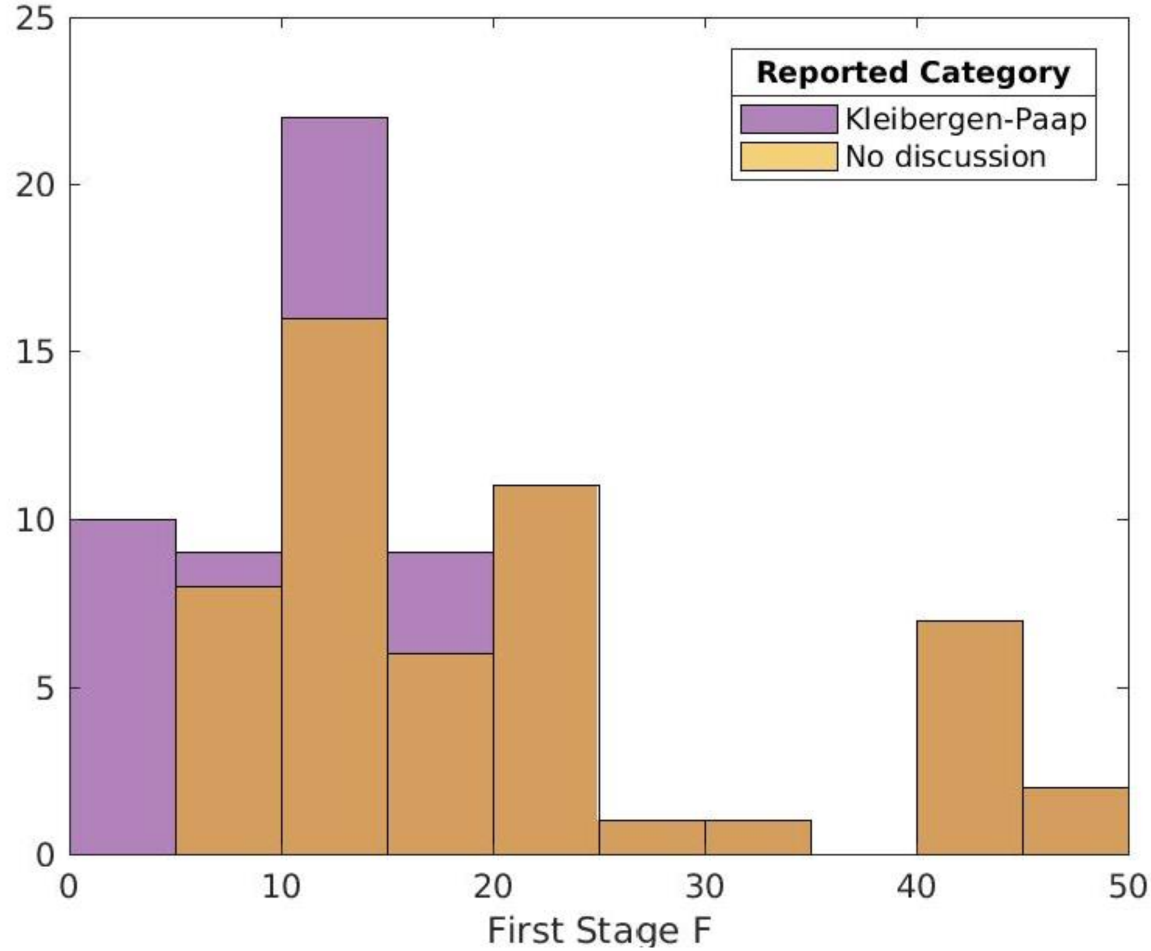
F^E – Effective first-stage F statistic of Montiel Olea and Plueger (2013)

- There are other candidates too, but they are not used in practice (and should not be); these include Hahn-Hausman (2002), Shea's (1997) partial R^2

- Multivariate extension (multiple included endogenous regressors): the Cragg-Donald statistic and its robustified counterpart, Kleibergen-Paap.

Detecting weak instruments in practice

Reported first-stage F's: what authors say they use



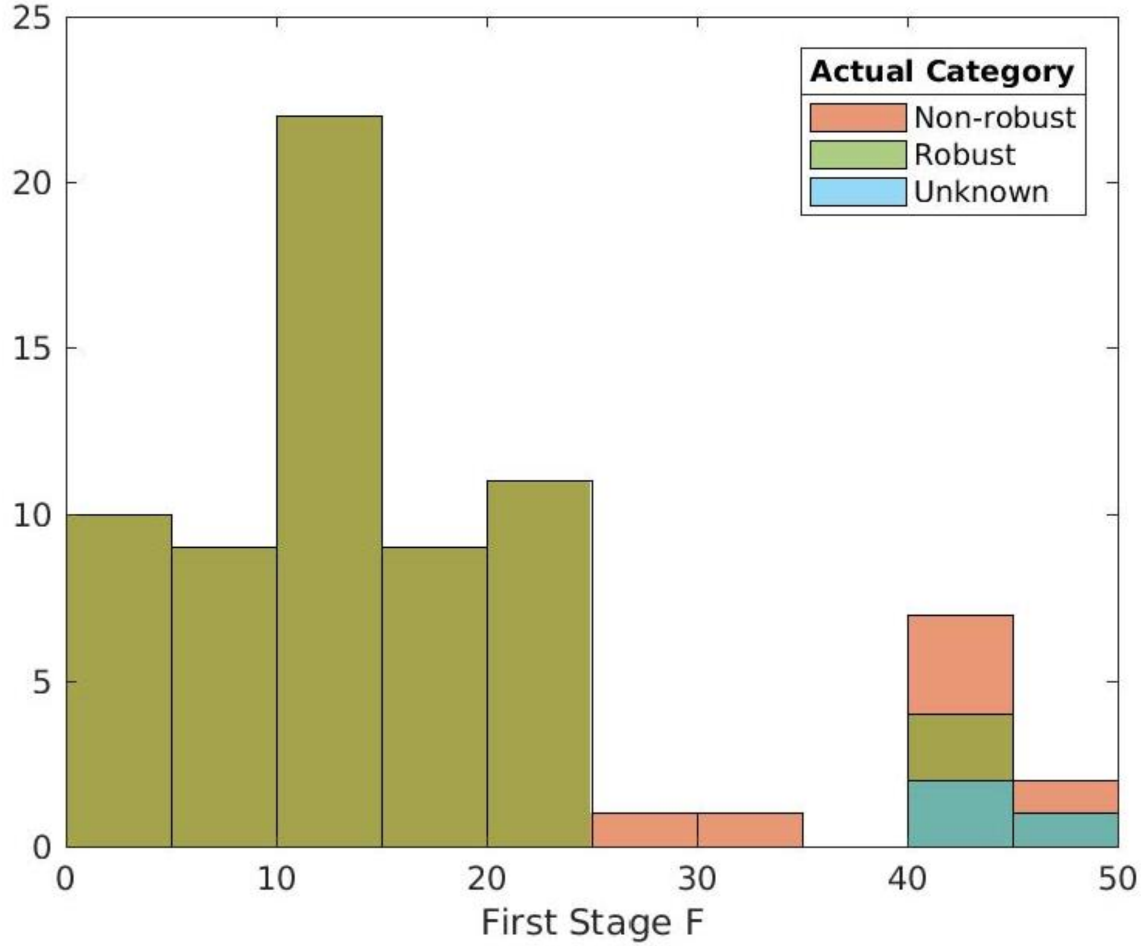
Candidates: F^N – nonrobust

F^R – robust (HR, HAC, cluster), also called Kleibergen-Paap (2006)

F^E – Effective first-stage F statistic of Montiel Olea and Plueger (2013)

Detecting weak instruments in practice, ctd

Actual first-stage *F*'s: what authors actually use



Candidates: F^N – nonrobust

F^R – robust (HR, HAC, cluster), also called Kleibergen-Paap (2006)

F^E – Effective first-stage F statistic of Montiel Olea and Plueger (2013)

Our recommendations (1 included endogenous regressor)

- Do:
  - Use the Montiel Olea-Pflueger (2013) effective first-stage F statistic

    $F^{Eff} = F^N \times$ correction factor for non-homoskedasticity
  - Report F^{Eff}
  - Compare F^{Eff} to MOP critical values (`weakivtest.ado`), or to 10.
  - If $F^{Eff} \geq$ MOP critical value, or ≥ 10 for rule-of-thumb method, use TSLS inference; else use weak-instrument robust inference.

- Don't
  - use/report p-values of test of $\pi = 0$ (null of irrelevant instruments)
  - use/report nonrobust first stage F (F^N)
  - use/report usual robust first-stage F (except OK for $k = 1$ where $F^R = F^{Eff}$)
  - use/report Kleibergen-Paap (2006) statistic (same thing).
  - compare HR/HAC/Kleibergen-Paap to Stock-Yogo critical values
  - reject a paper because $F^{Eff} < 10$!

    Instead, tell the authors to use weak-IV robust inference.

Notation and Review of IV Regression

IV regression model with a single endogenous regressor and *k* instruments

$$Y_i = X_i\beta + W_i'\gamma_1 + \varepsilon_i \qquad \text{(Structural equation)} \qquad (1)$$

$$X_i = Z_i'\pi + W_i'\gamma_2 + V_i \qquad \text{(First stage)} \qquad (2)$$

where W includes the constant. Substitute (2) into (1):

$$Y_i = Z_i'\delta + W_i'\gamma_3 + U_i \qquad \text{(Reduced form)} \qquad (3)$$

where $\delta = \pi\beta$ and $\varepsilon_i = U_i - \beta V_i$.

- OLS is in general inconsistent: $\hat{\beta}^{OLS} \xrightarrow{p} \beta + \dfrac{\sigma_{X\varepsilon}}{\sigma_X^2}$.
- β can be estimated by IV using the k instruments Z.
- By Frisch-Waugh, you can eliminate W by regressing Y, X, Z against W and using the residuals. This applies to everything we cover in the linear model so we drop W henceforth.

Setup:

$$Y_i = X_i\beta + \varepsilon_i \qquad \text{(Structural equation)} \qquad (1)$$

$$X_i = Z_i'\pi + V_i \qquad \text{(First stage)} \qquad (2)$$

$$Y_i = Z_i'\delta + U_i, \quad \delta = \pi\beta,\ \varepsilon = U - \beta V. \qquad \text{(Reduced form)} \qquad (3)$$

The two conditions for instrument validity

(i) Relevance: cov(Z,X) $\neq$ 0 or $\pi \neq 0$ (general k)

(ii) Exogeneity: cov(Z,ε) = 0

The IV estimator when k = 1 (Wright 1926)

$$\text{cov}(Z,Y) = \text{cov}(Z, X\beta + \varepsilon) = \text{cov}(Z, X)\beta + \text{cov}(Z,\varepsilon)$$

$$= \text{cov}(Z, X)\beta \qquad \text{by (i)}$$

so

$$\beta = \frac{\text{cov}(Z,Y)}{\text{cov}(Z,X)} \qquad \text{by (ii)}$$

IV estimator:

$$\hat{\beta}^{IV} = \frac{n^{-1}\sum_{i=1}^{n} Z_i Y_i}{n^{-1}\sum_{i=1}^{n} Z_i X_i} = \frac{\hat{\delta}}{\hat{\pi}}$$

Setup:

$$Y_i = X_i\beta + \varepsilon_i \qquad \text{(Structural equation)} \qquad (1)$$

$$X_i = Z_i'\pi + V_i \qquad \text{(First stage)} \qquad (2)$$

$$Y_i = Z_i'\delta + U_i, \quad \delta = \pi\beta, \varepsilon = U - \beta V. \qquad \text{(Reduced form)} \qquad (3)$$

---

k > 1: Two stage least squares (TSLS)

$$\hat{\beta}^{TSLS} = \frac{n^{-1}\sum_{i=1}^{n}\hat{X}_i Y_i}{n^{-1}\sum_{i=1}^{n}\hat{X}_i^2}, \qquad \text{where } \hat{X}_i = \text{predicted value from first stage}$$

$$= \frac{\mathbf{X'Z(Z'Z)^{-1}Z'Y}}{\mathbf{X'Z(Z'Z)^{-1}Z'X}}$$

$$= \frac{\hat{\pi}'\hat{Q}_{ZZ}\hat{\delta}}{\hat{\pi}'\hat{Q}_{ZZ}\hat{\pi}}, \qquad \text{where } \hat{Q}_{ZZ} = n^{-1}\sum_{i=1}^{n}Z_iZ_i'$$

The weak instruments problem is a "divide by zero" problem

- $cov(Z,X)$ is nearly zero; or π is nearly zero; or
- $\hat{\pi}'\hat{Q}_{ZZ}\hat{\pi}$ is noisy
- Weak IV is a subset of weak identification (Stock-Wright 2000, Nelson-Starts 2006, Andrews-Cheng 2012)

Statistics for measuring instrument strength

Non-robust: $$F^N = n\frac{\hat{\pi}'\hat{Q}_{ZZ}\hat{\pi}}{k\hat{\sigma}_V^2}$$

Robust: $$F^R = \frac{\hat{\pi}'\hat{\Sigma}_{\pi\pi}^{-1}\hat{\pi}}{k}$$

MOP Effective *F*: $$F^{Eff} = \frac{\hat{\pi}'\hat{Q}_{ZZ}\hat{\pi}}{tr\left(\hat{\Sigma}_{\pi\pi}^{1/2}\hat{Q}_{ZZ}\hat{\Sigma}_{\pi\pi}^{1/2\prime}\right)} = \frac{k\hat{\sigma}_V^2}{tr\left(\hat{\Sigma}_{\pi\pi}^{1/2}\hat{Q}_{ZZ}\hat{\Sigma}_{\pi\pi}^{1/2\prime}\right)}F^N$$

compare to TSLS: $$\hat{\beta}^{TSLS} = \frac{\hat{\pi}'\hat{Q}_{ZZ}\hat{\delta}}{\hat{\pi}'\hat{Q}_{ZZ}\hat{\pi}}$$

Intuition

- F^N measures the right thing ($\pi'Q_{ZZ}\pi$), but gets the SEs wrong
- F^R measures the wrong thing ($\pi'\Sigma_{\pi\pi}^{-1}\pi$), but gets the SEs right
- F^{Eff} measures the right thing and gets SEs right "on average"

Distributional assumptions

Setup: $X_i = Z_i'\pi + V_i$ (First stage) (2)

$Y_i = Z_i'\delta + U_i, \quad \delta = \pi\beta, \varepsilon = U - \beta V.$ (Reduced form) (3)

CLT: $\begin{pmatrix} \sqrt{n}\left(\hat{\delta}-\delta\right) \\ \sqrt{n}\left(\hat{\pi}-\pi\right) \end{pmatrix} \xrightarrow{d} N\left(0,\Sigma^*\right)$, Σ^* is HR/HAC/Cluster (henceforth, "HR")

(i) CLT limit holds exactly: $\begin{pmatrix} \hat{\delta} \\ \hat{\pi} \end{pmatrix} \sim N\left(\begin{pmatrix} \delta \\ \pi \end{pmatrix}, \Sigma\right)$, where $\Sigma = \begin{pmatrix} \Sigma_{\delta\delta} & \Sigma_{\delta\pi} \\ \Sigma_{\pi\delta} & \Sigma_{\pi\pi} \end{pmatrix} = n^{-1}\Sigma^*$

(ii) Reduced form variance & moment matrices are all known: Σ, Q_{ZZ}

<u>A lot is going on here!</u>

- HR/HAC/cluster variance estimators are consistent
- 1950s-1970s finite-sample normal (fixed Z's) literature

A lot is going on here, ctd

From $$\begin{pmatrix} \sqrt{n}\left(\hat{\delta}-\delta\right) \\ \sqrt{n}\left(\hat{\pi}-\pi\right) \end{pmatrix} \xrightarrow{d} N\left(0,\Sigma^*\right)$$

to $$\begin{pmatrix} \hat{\delta} \\ \hat{\pi} \end{pmatrix} \sim N\left(\begin{pmatrix} \delta \\ \pi \end{pmatrix},\Sigma\right), \text{ where } \Sigma = \begin{pmatrix} \Sigma_{\delta\delta} & \Sigma_{\delta\pi} \\ \Sigma_{\pi\delta} & \Sigma_{\pi\pi} \end{pmatrix} = n^{-1}\Sigma^*$$

- Weak IV asymptotics (Staiger-Stock 1997): $\pi = C/\sqrt{n}$.

$$kF^R = \hat{\pi}'\hat{\Sigma}_{\pi\pi}^{-1}\hat{\pi} = \left(\sqrt{n}\hat{\pi}\right)'\left(\hat{\Sigma}_{\pi\pi}^{-1}/n\right)\left(\sqrt{n}\hat{\pi}\right)$$

$$= \left(\sqrt{n}\left(\hat{\pi}-\pi\right)+\sqrt{n}\pi\right)'\hat{\Sigma}_{\pi\pi}^{*-1}\left(\sqrt{n}\left(\hat{\pi}-\pi\right)+\sqrt{n}\pi\right)$$

$$= \left(\sqrt{n}\left(\hat{\pi}-\pi\right)+C\right)'\hat{\Sigma}_{\pi\pi}^{*-1}\left(\sqrt{n}\left(\hat{\pi}-\pi\right)+C\right) \xrightarrow{d} \chi^2_{k;C'\Sigma^*_{\pi\pi}C}$$

- Limit experiment interpretation (Hirano-Porter 2015)
- Uniformity (D. Andrews-Cheng 2012)

Homework problem

Let $k = 2$ and $\hat{Q}_{ZZ} = I_2$. Suppose $\Sigma = \begin{pmatrix} \sigma_U^2 & \sigma_{UV} \\ \sigma_{UV} & \sigma_V^2 \end{pmatrix} \otimes \begin{pmatrix} \omega^2 & 0 \\ 0 & \omega^{-2} \end{pmatrix} / n$.

1) Let $\Sigma_{\pi\pi}^{*-1/2}\sqrt{n}\left(\hat{\pi} - \pi\right) \xrightarrow{d} z_\pi$. Show that:

a) $tr(\Sigma_{\pi\pi}^{1/2} Q_{ZZ} \Sigma_{\pi\pi}^{1/2\prime}) = \left(\omega^2 + \omega^{-2}\right)\sigma_V^2 / n$.

b) $F^N \cong \frac{1}{2}\left[\left(\lambda_1 + z_{\pi,1}\right)^2 \omega^2 + \left(\lambda_2 + z_{\pi,2}\right)^2 \omega^{-2}\right]$

c) $F^R \cong \frac{1}{2}\left(\lambda + z_\pi\right)'\left(\lambda + z_\pi\right)$

d) $F^{Eff} \cong \dfrac{\left(\lambda_1 + z_{\pi,1}\right)^2 \omega^2 + \left(\lambda_2 + z_{\pi,2}\right)^2 \omega^{-2}}{\omega^2 + \omega^{-2}}$

2) Adopt the weak instrument nesting $\pi = n^{-1/2}C$, where C_1, $C_2 \neq 0$. Show that as $\omega^2 \rightarrow \infty$:

a) "bias" of $\hat{\beta}^{TSLS} - \beta \cong \sigma_{\varepsilon V} / \sigma_V^2 = \text{plim}\left(\hat{\beta}^{OLS} - \beta\right)$

b) $F^N \xrightarrow{p} \infty$

c) $F^R \xrightarrow{p} \infty$

d) $F^{Eff} \xrightarrow{d} \chi_1^2$

3) Discuss

Work out the details for *k* = 1 first.

Preliminaries:

(a) Use distributional assumption (i)

$$\begin{pmatrix}\hat{\delta}\\ \hat{\pi}\end{pmatrix} \sim N\left(\begin{pmatrix}\delta\\ \pi\end{pmatrix},\Sigma\right),\ \text{ where } \Sigma = \begin{pmatrix}\Sigma_{\delta\delta} & \Sigma_{\delta\pi}\\ \Sigma_{\pi\delta} & \Sigma_{\pi\pi}\end{pmatrix} = n^{-1}\Sigma^*$$

to write,

$$\begin{aligned}\hat{\delta} &\cong \delta + \psi_\delta\\ \hat{\pi} &\cong \pi + \psi_\pi\end{aligned},\ \text{where } \begin{pmatrix}\psi_\delta\\ \psi_\pi\end{pmatrix} \sim N\left(0,\begin{pmatrix}\Sigma_{\delta\delta} & \Sigma_{\delta\pi}\\ \Sigma_{\pi\delta} & \Sigma_{\pi\pi}\end{pmatrix}\right)$$

(b) Connect to the structural regression:

$$\begin{aligned}(\mathbf{Z'Z})^{-1}\mathbf{Z'\varepsilon} &= \hat{\delta} - \hat{\pi}\beta \cong (\delta + \psi_\delta) - (\pi + \psi_\pi)\beta = (\delta - \pi\beta) + (\psi_\delta - \psi_\pi\beta)\\ &= \psi_\varepsilon,\ \text{ where }\ \psi_\varepsilon = \psi_\delta - \psi_\pi\beta\end{aligned}$$

(c) Standardize:

$$\begin{aligned}\hat{\pi} &\sim \pi + \psi_\pi = (\lambda + z_\pi)\Sigma_{\pi\pi}^{1/2}\\ \psi_\varepsilon &= z_\varepsilon\Sigma_{\varepsilon\varepsilon}^{1/2}\end{aligned},\ \text{where } \lambda = \Sigma_{\pi\pi}^{-1/2}\pi \text{ and } \begin{pmatrix}z_\varepsilon\\ z_\pi\end{pmatrix} \sim N\left(0,\begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix}\right)$$

(d) Project & orthogonalize:

$$z_\varepsilon = \rho z_\pi + \eta\ ,\ \text{where } \eta \sim N(0,1-\rho^2),\ \ \eta \perp z_v\ ,\ \rho = \Sigma_{\varepsilon\pi}\Big/\sqrt{\Sigma_{\varepsilon\varepsilon}\Sigma_{\pi\pi}}$$

What parameter governs departures from usual asymptotics (k = 1)?

$$\hat{\beta}^{IV} = \frac{\hat{\delta}}{\hat{\pi}}$$

$$= \frac{\hat{\pi}\beta + (\hat{\delta} - \hat{\pi}\beta)}{\hat{\pi}} \quad \text{add and subtract } \hat{\pi}\beta$$

$$\cong \beta + \frac{\psi_\varepsilon}{\pi + \psi_\pi} \quad \text{use representations in (a) and (b)}$$

$$= \beta + \frac{z_\varepsilon}{\lambda + z_\pi}\left(\frac{\Sigma_{\varepsilon\varepsilon}}{\Sigma_{\pi\pi}}\right)^{1/2} \text{standardize using representation in (c)}$$

$$= \beta + \underbrace{\frac{z_\pi}{\lambda + z_\pi}\left(\frac{\Sigma_{\varepsilon\pi}}{\Sigma_{\pi\pi}}\right)}_{\text{“bias”}} + \underbrace{\frac{\eta}{\lambda + z_\pi}\left(\frac{\Sigma_{\varepsilon\varepsilon}}{\Sigma_{\pi\pi}}\right)^{1/2}}_{\text{“noise”}} \quad \text{using projection (d)}$$

Parameter measuring instrument strength (k = 1) is $\lambda^2 = \pi^2 / \Sigma_{\pi\pi}$

"Bias" part of IV representation

$$\hat{\beta}^{IV} - \beta \cong \frac{z_\pi}{\lambda + z_\pi}\left(\frac{\Sigma_{\varepsilon\pi}}{\Sigma_{\pi\pi}}\right), \text{ where } \lambda = \Sigma_{\pi\pi}^{-1/2}\pi$$

Instrument strength depends on λ^2

- Strong instruments: $\lambda^2 \to \infty$, usual asymptotic distribution
- Irrelevant instruments: $\pi = 0$ so $\lambda = 0$:

$$\hat{\beta}^{IV} - \beta \cong \frac{\Sigma_{\varepsilon\pi}}{\Sigma_{\pi\pi}} + \frac{\eta}{z_\pi}\left(\frac{\Sigma_{\varepsilon\varepsilon}^{1/2}}{\Sigma_{\pi\pi}^{1/2}}\right) \sim \text{Cauchy centered at } \frac{\Sigma_{\varepsilon\pi}}{\Sigma_{\pi\pi}}$$

  - In homoskedastic case, $\frac{\Sigma_{\varepsilon\pi}}{\Sigma_{\pi\pi}} = \frac{\sigma_{\varepsilon V}}{\sigma_V^2} = \text{plim}\left(\hat{\beta}^{OLS} - \beta\right)$

- In the homoskedastic case, λ^2 = the concentration parameter (old Edgeworth expansion/finite sample distribution literature)

Instrument strength, k = 1, ctd.

How big does λ need to be? A "bias" heuristic:

$$\frac{E\left(\hat{\beta}^{IV}-\beta\right)}{\Sigma_{\varepsilon\pi}/\Sigma_{\pi\pi}} = E\,\frac{z_\pi}{\lambda+z_\pi}$$

$$= E\,\frac{z_\pi/\lambda}{1+z_\pi/\lambda}$$

$$\approx E\left(\frac{z_\pi}{\lambda}\right)\left(1-\frac{z_\pi}{\lambda}+\ldots\right) = -E\left(\frac{z_\pi^2}{\lambda^2}\right) = -\frac{1}{\lambda^2}$$

- For bias, relative to unidentified case, to be <0.1, need $\lambda^2 > 10$.
- But we don't know λ! So, we need a statistic with a distribution that depends on λ, which we can use to back out an estimate/test/rule of thumb.
- This is the Nagar (1959) expansion for the bias
- *How do the three candidate first-stage Fs fare?*

Distributions of the three first-stage *F*s, *k* = 1

First note that, when $k = 1$, $F^R = F^{Eff}$: $F^{Eff} = \dfrac{\hat{\pi}'\hat{Q}_{ZZ}\hat{\pi}}{tr\left(\hat{\Sigma}_{\pi\pi}^{1/2}\hat{Q}_{ZZ}\hat{\Sigma}_{\pi\pi}^{1/2\prime}\right)} = \dfrac{\hat{\pi}^2}{\hat{\Sigma}_{\pi\pi}} = F^R$

Distributions

$$F^{Eff}, F^R = \frac{\hat{\pi}^2}{\hat{\Sigma}_{\pi\pi}} \cong (\lambda + z_v)^2 \sim \chi^2_{1;\lambda^2}$$

$$F^N = n\frac{\hat{\pi}'\hat{Q}_{ZZ}\hat{\pi}}{\hat{\sigma}_V^2} = \frac{\hat{\pi}^2}{\hat{\sigma}_V^2 / n\hat{Q}_{ZZ}} \cong (\lambda + z_\pi)^2 \frac{\Sigma^*_{\pi\pi}}{\sigma_V^2 / Q_{ZZ}}$$

Implications

F^R, F^{Eff} can be used for inference about λ^2 when $k = 1$

- Estimation: $EF^{eff} = E\left(\lambda + z_V\right)^2 = \lambda^2 + 1$, so $\hat{\lambda}^2 = F^{Eff} - 1$
- Testing: H_0: "bias" ≤ 0.1. Reject H_0 if F^{Eff} > critical value.
- Rule of thumb: $F^{eff} < 10$ will detect weak IVs with probability that increases as λ^2 gets smaller

Implications, ctd.

$$F^N \cong (\lambda + z_V)^2 \frac{\Sigma^*_{\pi\pi}}{\sigma_V^2 / Q_{ZZ}}$$

$$F^{Eff}, F^R \cong \left(\lambda + z_V\right)^2$$

F^N is misleading in the HR case.

- Suppose $\Sigma^*_{\pi\pi}$ is large (i.e., first stage HR SEs are a lot bigger than NR SEs)

$$F^N \cong (\lambda + z_V)^2 \frac{\Sigma^*_{\pi\pi}}{\sigma_V^2 / Q_{ZZ}} \sim \frac{\Sigma^*_{\pi\pi}}{\sigma_V^2 / Q_{ZZ}} \times \chi^2_{1;\lambda^2}$$

where $\lambda^2 = \pi^2/\Sigma_{\pi\pi}$. For $\Sigma^*_{\pi\pi}$ large, $\lambda^2 \approx 0$, and $F^N \sim \frac{\Sigma^*_{\pi\pi}}{\sigma_V^2 / Q_{ZZ}} \times \chi^2_1 \to \infty$

i.e., Instruments are in the limit irrelevant – but $F_N \to \infty$.

In the k = 1 case, $F^R = F^{Eff}$. These differ in the k > 1 case, where F^{Eff} is preferred.

Homework problem

Let $k = 2$ and $\hat{Q}_{ZZ} = I_2$. Suppose $\Sigma = \begin{pmatrix} \sigma_U^2 & \sigma_{UV} \\ \sigma_{UV} & \sigma_V^2 \end{pmatrix} \otimes \begin{pmatrix} \omega^2 & 0 \\ 0 & \omega^{-2} \end{pmatrix} / n$.

1) Show that:

a) $tr(\Sigma_{\pi\pi}^{1/2} Q_{ZZ} \Sigma_{\pi\pi}^{1/2\prime}) = \left(\omega^2 + \omega^{-2}\right)\sigma_V^2 / n$.

b) $F^N \cong \frac{1}{2}\left[\left(\lambda_1 + z_{\pi,1}\right)^2 \omega^2 + \left(\lambda_2 + z_{\pi,2}\right)^2 \omega^{-2}\right]$

c) $F^R \cong \frac{1}{2}\left(\lambda + z_\pi\right)'\left(\lambda + z_\pi\right)$

d) $F^{Eff} \cong \dfrac{\left(\lambda_1 + z_{\pi,1}\right)^2 \omega^2 + \left(\lambda_2 + z_{\pi,2}\right)^2 \omega^{-2}}{\omega^2 + \omega^{-2}}$

2) Adopt the weak instrument nesting $\pi = n^{-1/2}C$, where C_1, $C_2 \neq 0$. Show that as $\omega^2 \to \infty$:

a) "bias" of $\hat{\beta}^{TSLS} - \beta \sim \sigma_{\varepsilon V} / \sigma_V^2 = \text{plim}\left(\hat{\beta}^{OLS} - \beta\right)$

b) $F^N \xrightarrow{p} \infty$

c) $F^R \xrightarrow{p} \infty$

d) $F^{Eff} \xrightarrow{d} \chi_1^2$

3) Discuss

Homework problem solution

Let $k = 2$ and $\hat{Q}_{ZZ} = I_2$. Suppose $\Sigma = \begin{pmatrix} \sigma_U^2 & \sigma_{UV} \\ \sigma_{UV} & \sigma_V^2 \end{pmatrix} \otimes \begin{pmatrix} \omega^2 & 0 \\ 0 & \omega^{-2} \end{pmatrix} / n$ and $\pi_1, \pi_2 \neq 0$

1(a) Direct calculation: $tr(\Sigma_{\pi\pi}^{1/2} Q_{ZZ} \Sigma_{\pi\pi}^{1/2\prime}) = \left(\omega^2 + \omega^{-2}\right)\sigma_V^2 / n$

1(b)-(d): We have already done the work to get the expressions below following "~", and the final expressions come from substitution of Q_{ZZ} and Σ:

(b)
$$F^N = \frac{n\hat{\pi}'\hat{Q}_{zz}\hat{\pi}}{k\sigma_V^2} \cong \frac{\left(\lambda + z_\pi\right)' n\Sigma_{\pi\pi}^{1/2}\hat{Q}_{zz}\Sigma_{\pi\pi}^{1/2\prime}\left(\lambda + z_\pi\right)}{k\sigma_V^2}$$

$$= \frac{1}{2}\left[\left(\lambda_1 + z_{\pi,1}\right)^2 \omega^2 + \left(\lambda_2 + z_{\pi,2}\right)^2 \omega^{-2}\right]$$

(c)
$$F^R = \frac{\hat{\pi}'\hat{\Sigma}_{\pi\pi}^{-1}\hat{\pi}}{k} \cong \frac{\left(\lambda + z_V\right)'\left(\lambda + z_V\right)}{k} = \frac{1}{2}\left(\lambda + z_\pi\right)'\left(\lambda + z_\pi\right)$$

(d)
$$F^{Eff} = \frac{\hat{\pi}'\hat{Q}_{ZZ}\hat{\pi}}{tr\left(\hat{\Sigma}_{\pi\pi}^{1/2}\hat{Q}_{ZZ}\hat{\Sigma}_{\pi\pi}^{1/2\prime}\right)} \cong \frac{\left(\lambda + z_\pi\right)'\Sigma_{\pi\pi}^{1/2}\hat{Q}_{zz}\Sigma_{\pi\pi}^{1/2\prime}\left(\lambda + z_\pi\right)}{tr\left(\Sigma_{\pi\pi}^{1/2}Q_{ZZ}\Sigma_{\pi\pi}^{1/2\prime}\right)}$$

$$= \frac{\left(\lambda_1 + z_{\pi,1}\right)^2 \omega^2 + \left(\lambda_2 + z_{\pi,2}\right)^2 \omega^{-2}}{\omega^2 + \omega^{-2}}$$

Homework problem solution, ctd.

2) Adopt the weak instrument nesting $\pi = n^{-1/2}C$, where C_1, $C_2 \neq 0$. Show that as $\omega^2 \to \infty$:

a) "bias" of $\hat{\beta}^{TSLS} - \beta \sim \sigma_{\varepsilon V}/\sigma_V^2 = \text{plim}\left(\hat{\beta}^{OLS} - \beta\right)$

Last part first: $\text{plim}\left(\hat{\beta}^{OLS} - \beta\right) = \sigma_{\varepsilon X}/\sigma_X^2 = \sigma_{\varepsilon V}/\sigma_V^2$ because $\pi = n^{1/2}C$.

Next obtain the expression (*several tedious steps*),

$$\text{"Bias" part } \hat{\beta}^{TSLS} - \beta \cong \frac{(\lambda + z_\pi)' HRz_\pi}{(\lambda + z_\pi)' H(\lambda + z_\pi)}$$

where $H = \Sigma_{\pi\pi}^{1/2} Q_{ZZ} \Sigma_{\pi\pi}^{1/2} = \begin{pmatrix} \omega^2 & 0 \\ 0 & \omega^{-2} \end{pmatrix} \sigma_V^2 / n$ and $R = \Sigma_{\pi\pi}^{-1/2} \Sigma_{\varepsilon\pi} \Sigma_{\pi\pi}^{-1/2\prime} = \frac{\sigma_{\varepsilon V}}{\sigma_V^2} I_2$.

For the weak instrument nesting,

$$\lambda = \Sigma_{\pi\pi}^{-1/2}\pi = \left[\sigma_V^2 \begin{pmatrix} \omega^2 & 0 \\ 0 & \omega^{-2} \end{pmatrix} / n \right]^{-/12} \pi$$

$$= \begin{pmatrix} \omega^{-1} & 0 \\ 0 & \omega \end{pmatrix} n^{1/2} \pi / \sigma_V = \begin{pmatrix} C_1 \omega^{-1} / \sigma_V \\ C_2 \omega / \sigma_V \end{pmatrix}$$

Homework problem solution, ctd.

Now substitute these expressions for λ, H, and R into the "bias" part:

$$\hat{\beta}^{TSLS} - \beta \cong \frac{(\lambda + z_\pi)'\begin{pmatrix} \omega^2 & 0 \\ 0 & \omega^{-2} \end{pmatrix} z_\pi}{(\lambda + z_\pi)'\begin{pmatrix} \omega^2 & 0 \\ 0 & \omega^{-2} \end{pmatrix}(\lambda + z_\pi)} \frac{\sigma_{\varepsilon V}}{\sigma_V^2}$$

$$= \frac{(C_1/\sigma_V + z_{\pi,1}\omega)z_{\pi,1}\omega + (C_2/\sigma_V + z_{\pi,2}\omega^{-1})z_{\pi,2}\omega^{-1}}{(C_1/\sigma_V + z_{\pi,1}\omega)^2 + (C_2/\sigma_V + z_{\pi,2}\omega^{-1})^2}\left(\frac{\sigma_{\varepsilon V}}{\sigma_V^2}\right)$$

$$= \left(1 + O_p(\omega^{-1})\right)\left(\frac{\sigma_{\varepsilon V}}{\sigma_V^2}\right)$$

Homework problem solution, ctd.

Remaining parts by substitution and taking limits:

(b) $$F^N \cong \frac{1}{2}\left[\left(\lambda_1 + z_{\pi,1}\right)^2 \omega^2 + \left(\lambda_2 + z_{\pi,2}\right)^2 \omega^{-2}\right]$$

$$= \frac{1}{2}\left[\left(C_1/\sigma_V + z_{\pi,1}\omega\right)^2 + \left(C_2/\sigma_V + z_{\pi,2}\omega^{-1}\right)^2\right] \sim \frac{1}{2}\omega^2\chi_1^2 + O_p(\omega)$$

(c) $$F^R \cong \frac{1}{2}\left(\lambda + z_\pi\right)'\left(\lambda + z_\pi\right)$$

$$= \frac{1}{2}\left[\left(C_1\omega^{-1}/\sigma_V + z_{\pi,1}\right)^2 + \left(C_2\omega/\sigma_V + z_{\pi,2}\right)^2\right]$$

$$= \frac{1}{2}\frac{C_2^2}{\sigma_V^2}\omega^2 + O_p(\omega) \to \infty$$

(d) $$F^{Eff} = \frac{F^N}{\omega^2 + \omega^{-2}} \cong \frac{\omega^2 z_{\pi,1}^2 + O_p(\omega)}{\omega^2 + \omega^{-2}} = z_{\pi,1}^2 + O_p(\omega^{-1}) \sim \chi_1^2$$

3) Discuss

OK, use F^{Eff} – but what cutoff?

$$F^{Eff} \cong (\lambda + z_\pi)' H (\lambda + z_\pi), \text{ where } H = \frac{\Sigma_{\pi\pi}^{1/2} Q_{ZZ} \Sigma_{\pi\pi}^{1/2}}{tr(\Sigma_{\pi\pi}^{1/2} Q_{ZZ} \Sigma_{\pi\pi}^{1/2})}$$

~ weighted average of noncentral χ^2's – depends on full matrix H,

$0 \leq$ eigenvalues (H) ≤ 1

Scalar case: Just use Stock-Yogo (2005) critical values

Hierarchy of options: overidentified case

1. **Testing approach**: test null of $\lambda' H \lambda \geq$ some threshold (e.g. 10% bias)
   - a) (MOP Monte Carlo method) Given $\hat{H}$, compute cutoff $\lambda' \hat{H} \lambda$; critical value by simulation
   - b) (MOP Paitnik-Nagar method) Approximate weighted average of noncentral χ^2's by noncentral χ^2; compute cutoff value of $\lambda' H \lambda$ using Nagar approximation to the bias, with some maximal allowable bias. Implemented in **`weakivtest.ado.`**
   - c) (MOP simple method) Pick a maximal allowable bias (or size distortion) and use their "simple" critical values (based on noncentral χ^2 bounding distribution). *These are simple, but conservative.*
2. **Consistent sequence approach**: "Weak" if $F^{Eff} < \kappa_n$, $\kappa_n \to \infty$ (but what is κ_n?)
3. **Rule-of-thumb approach**: "Weak" if $F^{Eff} < 10$

k=1 case, additional comments about F^{Eff} and F^R

$$\hat{\beta}^{IV} - \beta \cong \frac{z_\pi}{\lambda + z_\pi}\left(\frac{\Sigma_{\varepsilon\pi}}{\Sigma_{\pi\pi}}\right), \text{ where } \lambda = \Sigma_{\pi\pi}^{-1/2}\pi$$

$$t^{IV} = \frac{\hat{\beta}^{IV} - \beta_0}{SE(\hat{\beta}^{IV})} \cong \frac{z_\varepsilon}{\left[1 - 2\left(\frac{z_\varepsilon}{\lambda + z_\pi}\right)\rho + \left(\frac{z_\varepsilon}{\lambda + z_\pi}\right)^2\right]^{1/2}}, \text{ where } \rho = \frac{\Sigma_{\pi\varepsilon}}{\left(\Sigma_{\pi\pi}\Sigma_{\varepsilon\varepsilon}\right)^{1/2}}$$

$$F^R = F^{Eff} \cong \left(\lambda + z_\pi\right)'\left(\lambda + z_\pi\right)$$

- By maximizing over ρ you can find worst case size distortion for usual IV t-stat testing β_0. This depends on λ, which can be estimated from $F^R = F^{Eff}$.
- These are the same expressions, with different definition of λ, as in homoskedastic case (special to $k = 1$)
- Critical values for $k = 1$ – two choices:
- Nagar bias ≤ 10%: 23 (5% critical value from $\chi^2_{1;\lambda^2=10}$) (MOP)
- Maximum t^{IV} size distortion of 0.10: 16.4; of 0.15: 9.0
- But with $k = 1$ there are fully robust methods that are easy and have very strong theoretical properties (AR) (Lecture 3).

Detecting weak instruments with multiple included endogenous regressors

Methods are based on multivariate F: Cragg-Donald statistic and robust variants

- Nonrobust:
    - Minimum eigenvalue of Cragg-Donald statistic, Stock-Yogo (2005) critical values
    - Sanderson-Windmeijer (2016)
- HR: Main method used is Kleibergen-Paap statistic, which is HR Cragg-Donald.
    - But recall that this doesn't work (theory) for 1 X, and having multiple X's doesn't improve things.
- MOP Effective F: Hasn't been developed yet for the case of multiple included endogenous regressors.

More work is needed….

What if you want to use efficient 2-step GMM, not TSLS?

Everything above is tailored to TSLS!

- Suppose that, if you have strong instruments, you use efficient 2-step GMM:

$$\hat{\beta}^{GMM} = \frac{\hat{\pi}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\delta}}{\hat{\pi}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\pi}}\text{, where } \hat{\Sigma}_{\varepsilon\varepsilon} = \frac{1}{n}\sum_{i=1}^{n} Z_i Z_i'\left(\hat{\varepsilon}_i^{(1)}\right)^2$$

  where $\hat{\varepsilon}_i^{(1)}$ is the residual from a first-stage estimate of β, e.g. TSLS.

- Things get complicated because the first step (TSLS) isn't consistent with weak instruments.
  - $\hat{\Sigma}_{\varepsilon\varepsilon}$ converges in distribution to a random limit
  - If $\Sigma_{\varepsilon\varepsilon}$ were known (infeasible),

$$\hat{\beta}^{GMM} - \beta \Rightarrow \frac{\left(\lambda + z_\pi\right)'\Sigma_{\pi\pi}^{1/2}\Sigma_{\varepsilon\varepsilon}^{-1/2} z_\varepsilon}{\left(\lambda + z_\pi\right)'\Sigma_{\pi\pi}^{1/2}\Sigma_{\varepsilon\varepsilon}^{-1}\Sigma_{\pi\pi}^{1/2\prime}\left(\lambda + z_\pi\right)}$$

    In general none of the F's discussed so far get at the right object, $\lambda'\Sigma_{\pi\pi}^{1/2}\Sigma_{\varepsilon\varepsilon}^{-1}\Sigma_{\pi\pi}^{1/2\prime}\lambda / tr(\Sigma_{\pi\pi}^{1/2}\Sigma_{\varepsilon\varepsilon}^{-1}\Sigma_{\pi\pi}^{1/2\prime})$. (And this is "right" only if $\Sigma_{\varepsilon\varepsilon}$ is known.)

3. Estimation

What have we learned/state of knowledge: $k = 1$

- There does not exist an unbiased or asymptotically-unbiased IV estimator (folk theorem; Hirano and Porter 2015).
- Only one moment condition, so weighting (HR) isn't an issue
- LIML=TSLS=IV doesn't have moments…
- Fuller seems to have advantage over IV in terms of "bias" (location) in simulations (e.g., Hahn, Hausman, Kuersteiner (2004), I. Andrews and Armstrong 2017) (so should *k*-class).
- If you know *a-priori* the sign of π, then unbiased, strong-instrument efficient estimation is possible (I. Andrews and Armstrong 2017)

Estimation, ctd.

What have we learned/state of knowledge: $k > 1$

- There does not exist an unbiased or asymptotically-unbiased IV estimator (folk theorem; Hirano and Porter 2015).
- The IV estimators that were developed in the 60s-90s (LIML, k-class, double k-class, JIVE, Fuller) are special to the homoskedastic case, and in general lose their good properties in the HR case
- Different IV estimators place different weights on the moments, and thus in general have different LATEs
- With heterogeneity, the LIML estimand (Fuller too?) can be outside the convex hull of the LATEs of the individual instruments (Kolesár 2013)
- For GMM applications estimating a structural parameter (e.g. New Keynesian Phillips Curve, etc.), the LATE concerns don't apply, however when the moment conditions are nonlinear in θ, things get difficult.
- If you know *a-priori* the sign of π, then unbiased estimation is possible (I. Andrews and Armstrong 2017)

4. Weak-Instrument Robust Inference

OK – now what should you do if you have weak instruments?

Wrong answer: reject the paper.

Negative result:

Confidence intervals of the form $\hat{\beta} \pm \hat{\Delta}$ in general won't work

- Dufour (1997): if β is unidentified (irrelevant instrument), the confidence interval must be infinite with probability 95%
- We need a different approach

Instead, the weak IV-robust literature constructs confidence intervals by inverting tests

- Under H_0: $\beta = \beta_0$, a correctly-sized test rejects only 5% of the time.
- Thus, the acceptance region of a 5% test contains the true value of β in 95% of all draws.

The Anderson-Rubin Statistic

Setup: $Y_i = X_i\beta + \varepsilon_i$ (Structural equation) (1)

$X_i = Z_i'\pi + V_i$ (First stage) (2)

$Y_i = Z_i'\delta + U_i, \quad \delta = \pi\beta, \varepsilon = U - \beta V.$ (Reduced form) (3)

Under the null, $Y_i - X_i\beta_0 = \varepsilon_i$, so from the instrument exogeneity condition,

$$E\left[\left(Y_i - X_i\beta_0\right)Z_i\right] = E\left[\varepsilon_i Z_i\right] = 0$$

To test this, run the regression,

$$Y_i - X_i\beta_0 = \gamma_0 + \gamma' Z_i + e_{ix}.$$

The HR (HAC) *F*-statistic is,

$$AR(\beta_0) = N\hat{\gamma}_1\hat{\Omega}(\beta_0)^{-1}\hat{\gamma}_1$$

$$= N\left(Y - X\beta_0\right)' Z\left(Z'Z\right)^{-1}\hat{\Omega}(\beta_0)^{-1}\left(Z'Z\right)^{-1}Z'\left(Y - X\beta_0\right)$$

where $\hat{\Omega}(\beta_0) = \mathrm{var}\left(\hat{\gamma}_1\right)$, computed under $\beta = \beta_0$ (HR, cluster, HAR, etc)

The Anderson-Rubin Confidence Interval

Algorithm

1. Pick a value of β_0 and run the regression, $Y_i - X_i\beta_0 = \gamma_0 + \gamma' Z_i + e_{ix}$ using HR/HAR/Cluster SEs.
2. Reject if $AR(\beta_0) = N\hat{\gamma}_1\hat{\Omega}(\beta_0)^{-1}\hat{\gamma}_1$
3. If AR(β_0) < $\chi^2_{k;.05}$ (k = # instruments) then retain that value of β_0
4. Repeat for another value of β_0 (grid of β_0)

The set of retained values is the acceptance region of the test = 95% AR confidence interval for β.

Some strange properties of the AR interval

- Can be a closed interval, two open intervals, the real line, or empty in the homoskedastic case – and more complicated forms in the HR/HAC case.
- In the overidentified case (only), if one or more of the exogeneity conditions doesn't hold, it can be incorrectly small (rejecting the overid condition, not $\beta \neq \beta_0$).

Optimality of AR in Just-Identified Models

- In just-identified case with single endogenous regressor, AR is optimal
  - 101 out of 230 specifications in our AER sample are just-identified with a single endogenous regressor
- Moreira (2009) shows that AR test uniformly most powerful unbiased
- AR equivalent to two-sided t-test when instruments are strong
- In just-identified settings, strong case for using AR CS
  - Optimal among CS robust to weak instruments
  - No loss of power relative to t-test if instruments strong

Is it a problem that the AR interval is longer than the usual IV interval?

- Not in the exactly identified case.
- The two intervals will be the same if the instruments are strong, but if they are weak, the usual IV interval can be too short and in the wrong place.

Over-Identified Models

- With over-identification, AR is inefficient under strong instruments ("too many degrees of freedom being tested")
- In the homoskedastic case, Moreira's (2003) conditional likelihood ratio (CLR) test is essentially optimal (Andrews, Moreira, and Stock (2006)). No (single) way to generalize this to the HR/HAR case.
- The problem of optimal testing in the HR case is hard, with recent progress by Moreira and Moreira (2015), Montiel Olea (2017), and Moreira and Ridder (2018a)
- The tests currently in use are functions of HR/HAR AR and LM statistics.
- These tests can work well in some designs (I. Andrews (2016)) but can work poorly in others (Moreira and Ridder (2018b)).
    - Even under homoskedasticity, the LM statistic has non-monotonic power
- For now: use HR/HAR AR statistic or I. Andrews (2016) linear combination test with plug-in weight function. Work on improving these is ongoing.

In the homoskedastic case, CLR has desirable properties, but LM does not

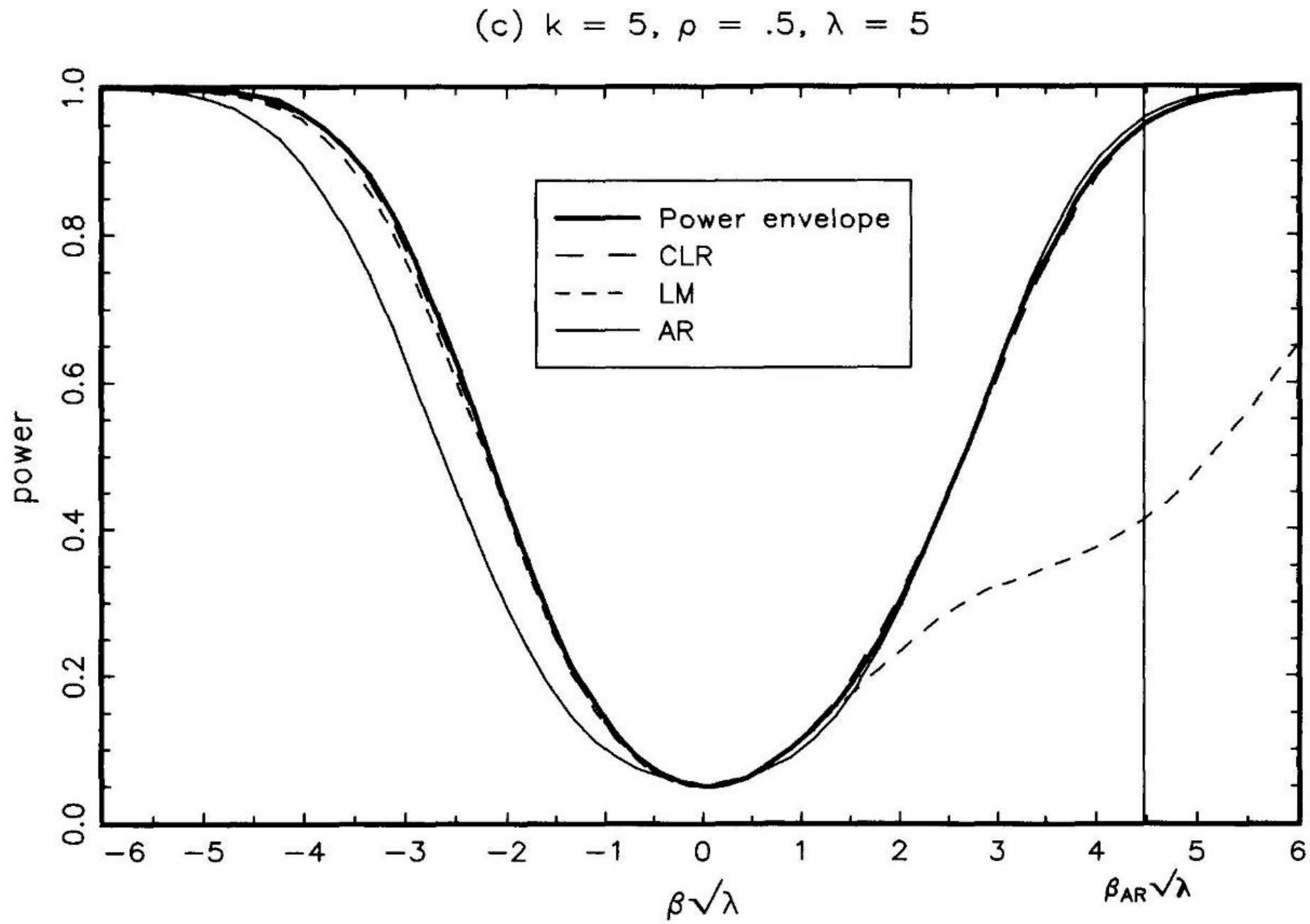


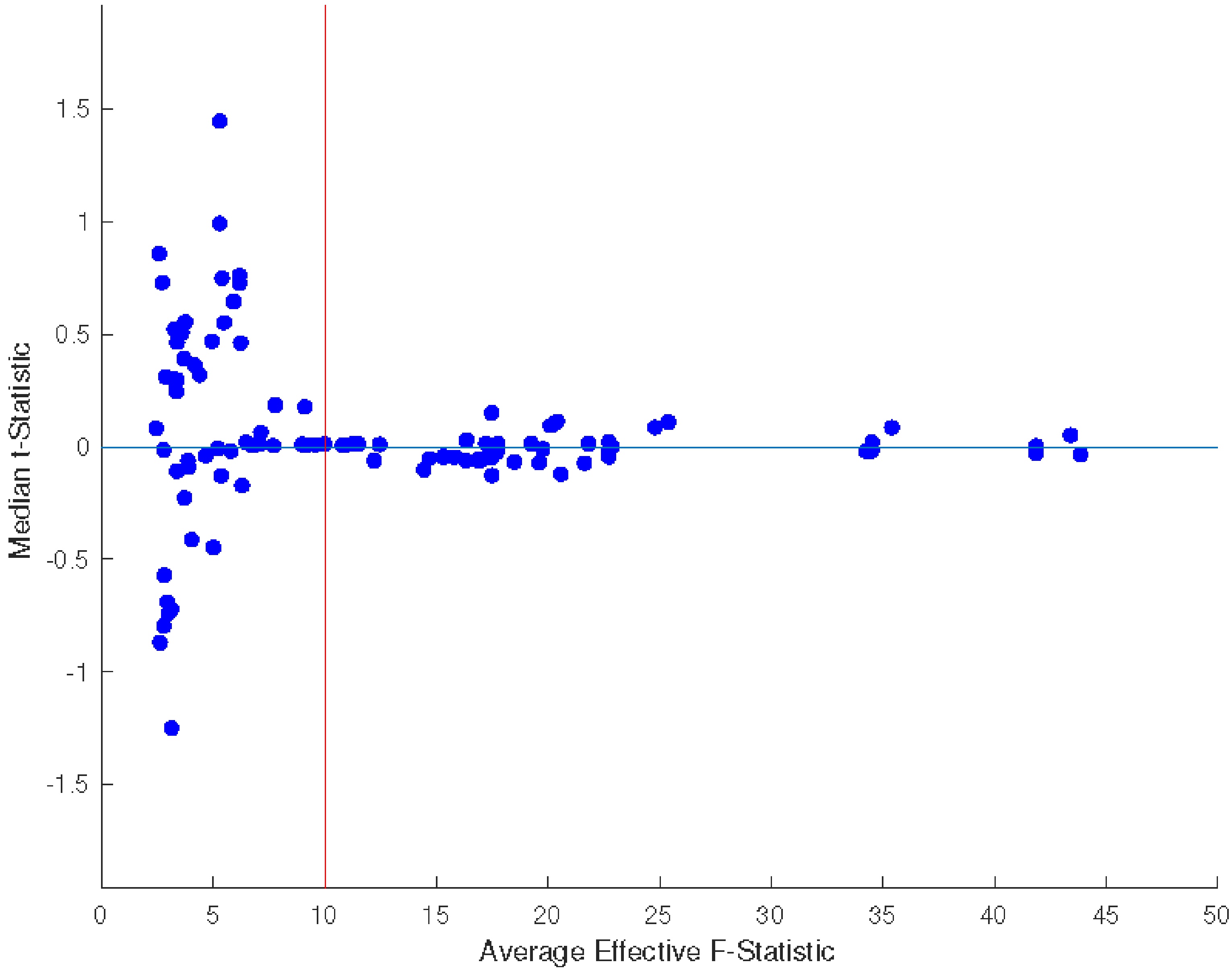
Figure: Power of AR, K, and C LR tests in homoskedastic case (from D. Andrews, Moreira, and Stock (2006))

Simulations from a sample of AER papers

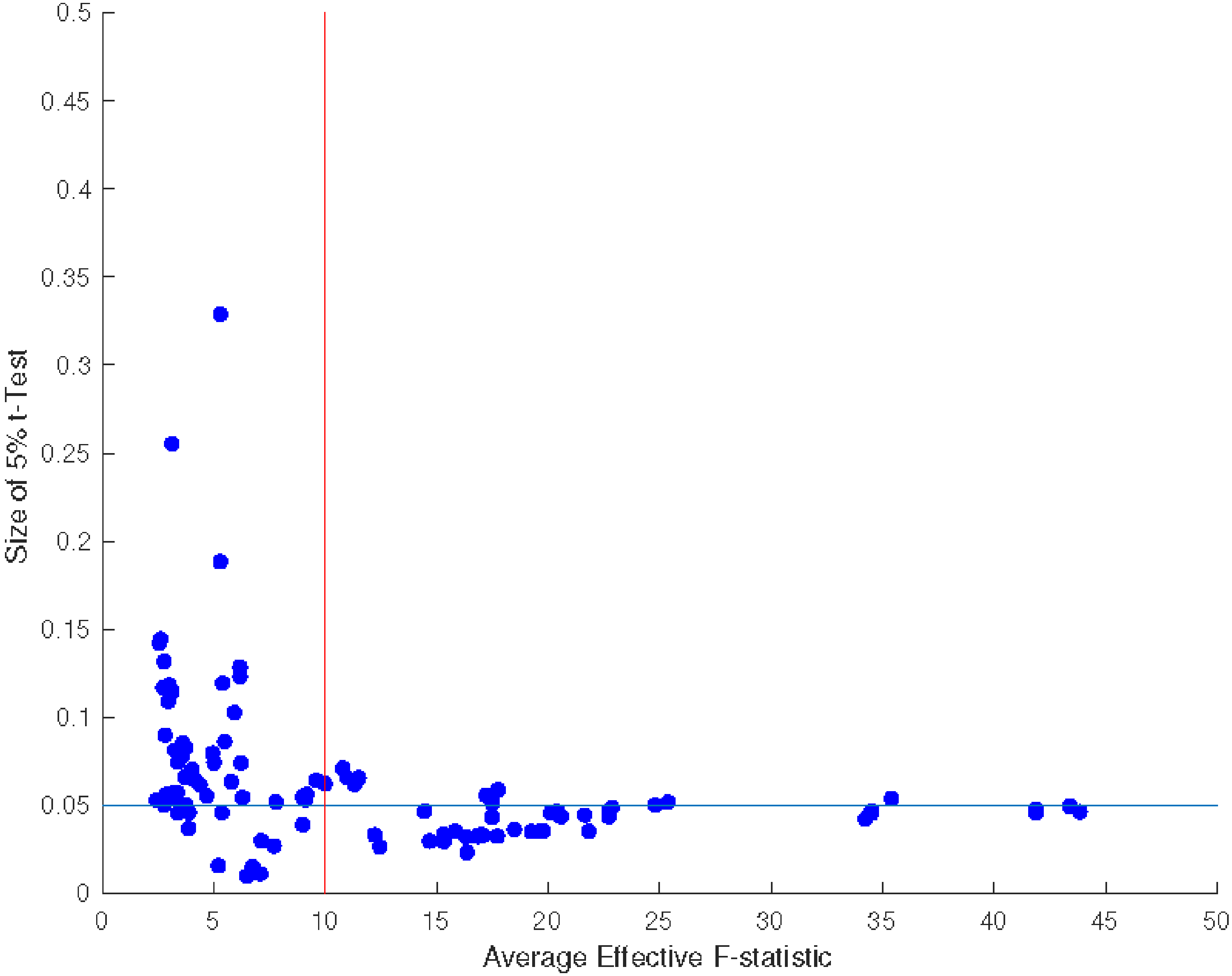
- Sample: All *AER* IV excluding P&P 2014-2018 that provide enough information to estimate the variance matrix of $(\hat{\delta}, \hat{\pi})$ (i.e. papers with replication data, + 1)
- 124 specifications from 18 papers.
- All specifications we examine here have a single endogenous regressor

The following slides provide a taste of the MC results

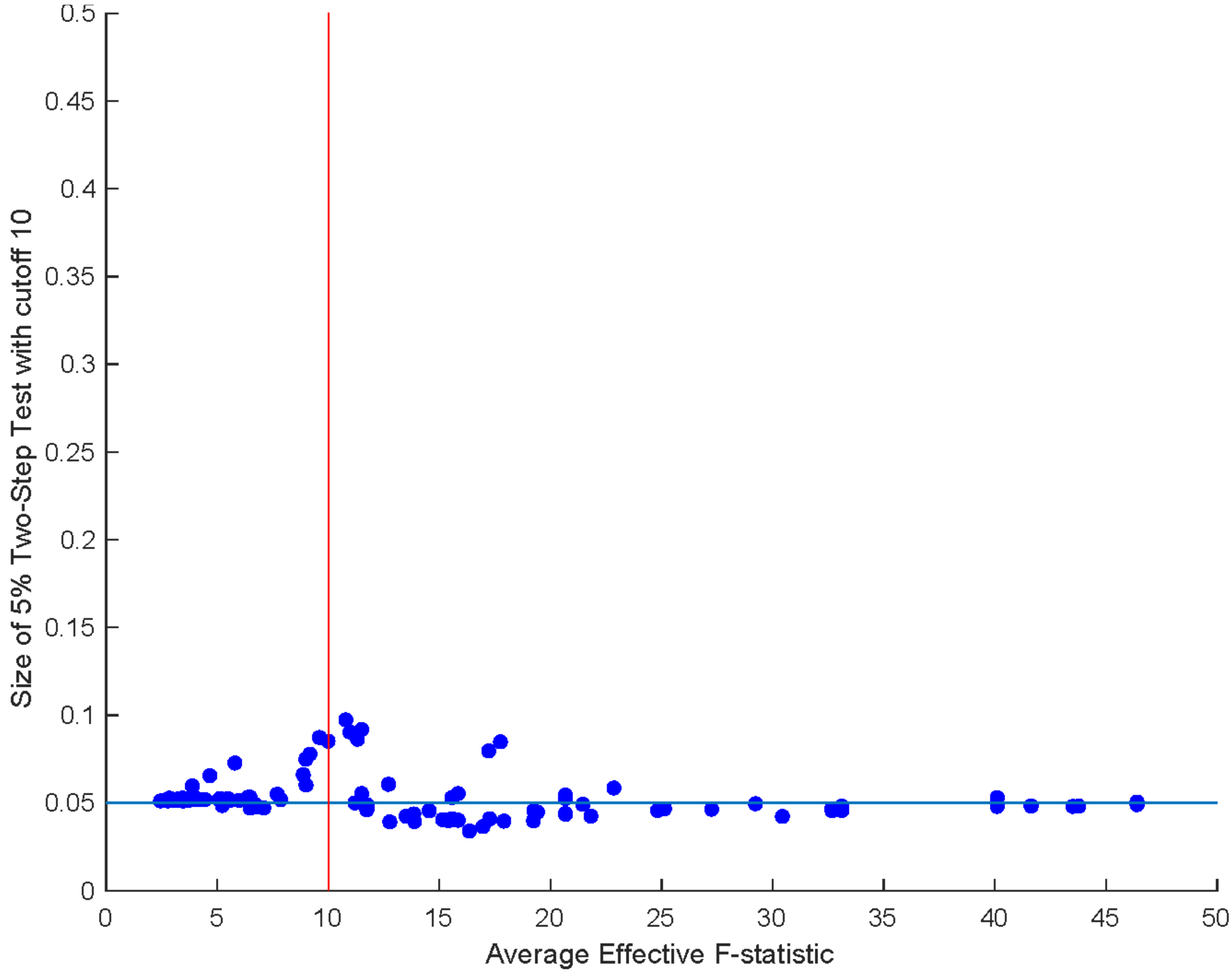
Median of distribution of *t*-statistic



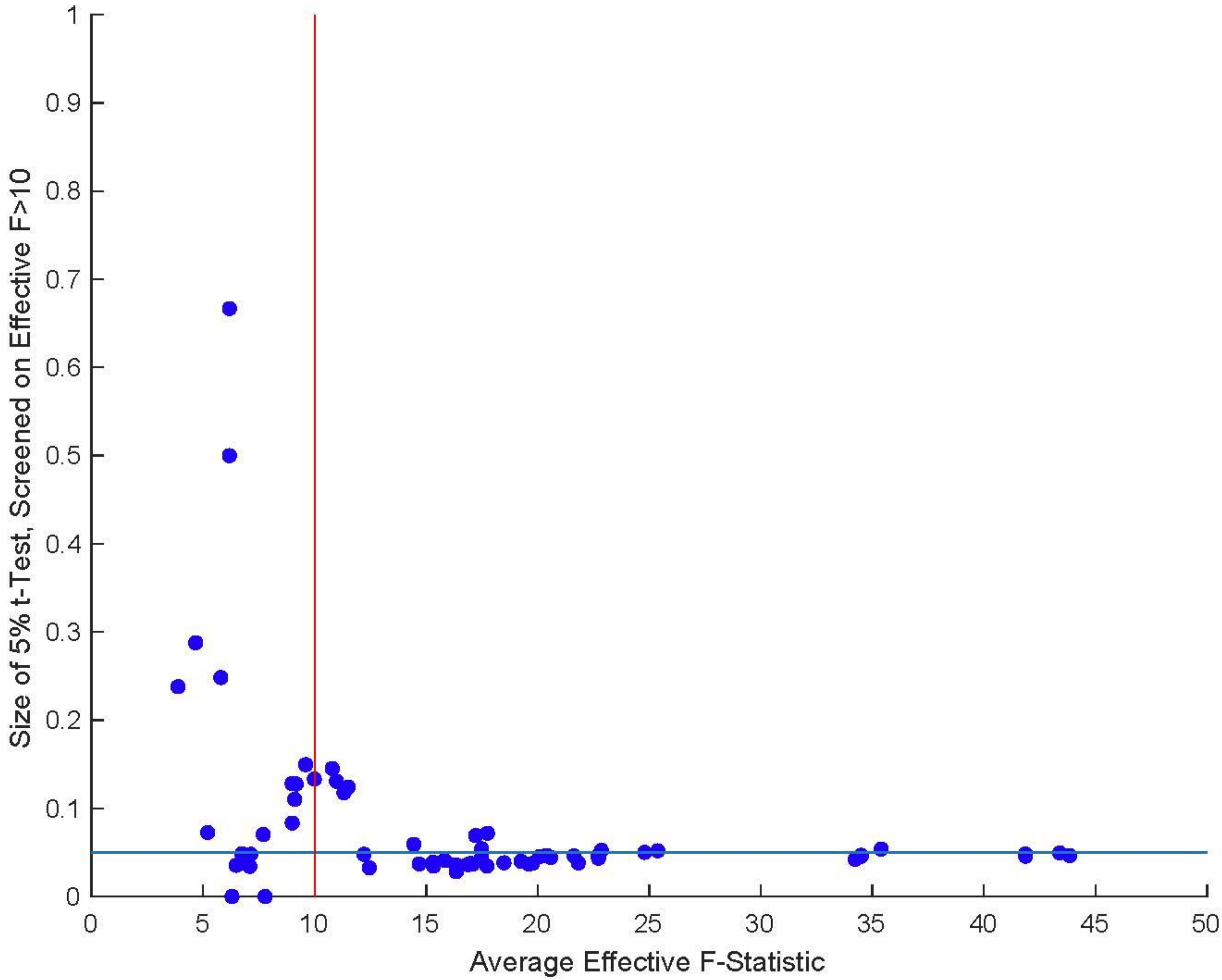
Size of *t*-statistic (fraction of $|t| > 1.96$ under the null)



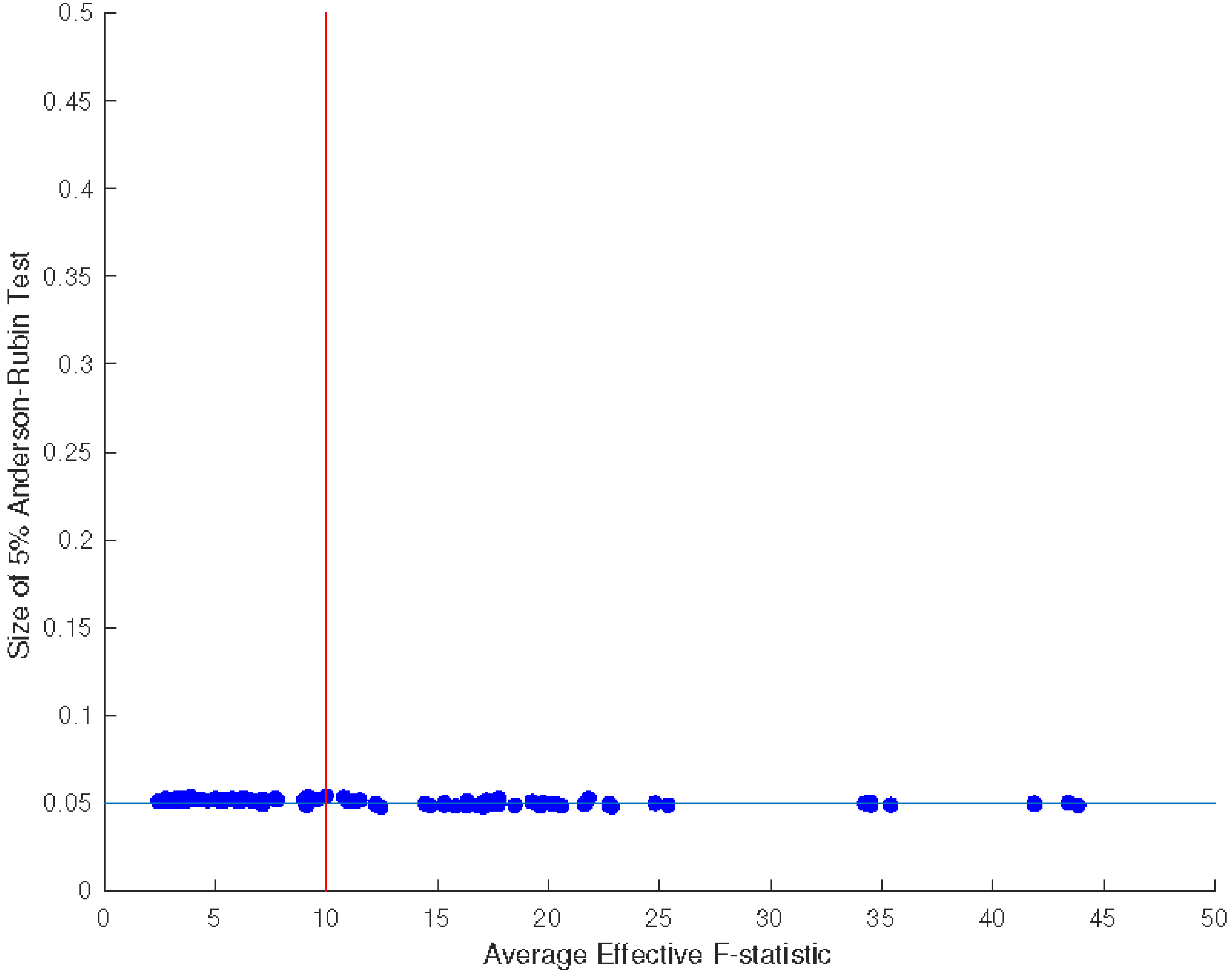
Size of 2-step method (robust if $F^{Eff} < 10$, conventional otherwise)



Size of screening method based on first stage F (reject if $F^{Eff} < 10$)



Size of AR statistic



5. Summary

1. Conventional t-test based confidence intervals can under-cover true parameter value, and be centered in the wrong place, when instruments are weak.

2. The MOP Effective First-stage F provides a guide to bias – but screening applications (rejecting papers) using the first-stage F can induced size distortions.

3. In the exactly identified case, use HR/HAR AR (strong optimality)

4. In the overidentified case, use pretest based on MOP F^{Eff} , with either AR or I. Andrews (2016) confidence intervals if $F^{Eff} < 10$ (or, use a MOP critical value).

References

Andrews I. (2016). "Conditional linear combination tests for weakly identified models." *Econometrica* 84:2155-2182.

Andrews, I., J. Stock, and L. Sun (2018). "Weak Instruments in IV Regression: Theory and Practice," manuscript, Harvard University.

Andrews, D.W.K. and X. Cheng (2012). "Estimation and Inference with Weak, Semi-Strong, and Strong Identification," *Econometrica* 80, 2153-2211.

Andrews, I. and T. Armstrong (2017). "Unbiased Instrumental Variables Estimation under Known First-Stage Sign," *Quantiative Economics* 8, 479-503.

Hahn, J., J. Hausman, and G. Kuersteiner (2004), "Estimation with weak instruments: Accuracy of higher order bias and MSE approximations," *Econometrics Journal*, 7, 272–306.

Kleibergen, F., and R. Paap (2006). "Generalized Reduced Rank Tests using the Singular Value Decomposition." *Journal of Econometrics* 133: 97–126.

Montiel Olea J.L. (2017). "Admissible, similar tests: A characterization." Unpublished Manuscript.

Montiel Olea, J.L. and C.E. Pflueger (2013). "A Robust Test for Weak Instruments," *Journal of Business and Economic Statistics* 31, 358-369.

Moreira M. (2003). "A conditional likelihood ratio test for structural models." *Econometrica* 71:1027-1048.

Moreira, H. and M. J. Moreira (2015). "Optimal Two-Sided Tests for Instrumental Variables Regression with Heteroskedastic and Autocorrelated Errors." arXiv:1505.06644.

Moreira M. and G. Ridder (2017). "Optimal invariant tests in an instrumental variables regression with heteroskedastic and autocorrelated errors." Unpublished Manuscript.

Moreira M. and G. Ridder (2017). "Efficiency Loss of Asymptotically Efficient Tests in an Instrumental Variables Regression," Unpublished Manuscript.

Nagar, A. L. (1959). “The bias and moment matrix of the general k-class estimators of the parameters in simultaneous equations,” *Econometrica* 27: 575–595.

Nelson, C. R., and R. Startz (1990). “Some further results on the exact small sample properties of the instrumental variable estimator.” *Econometrica* 58, 967–976.

Nelson, C. R., and R. Startz (2006). “The zero-information-limit condition and spurious inference in weakly identified models,” *Journal of Econometrics* 138, 47-62.

Pflueger, C.E. and S. Wang (2015). “A Robust Test for Weak Instruments in Stata,” *Stata Journal* 15, 216-225.

Sanderson, E. and F. Windmeijer (2016). “A Weak Instrument *F*-test in Linear IV Models with Multiple Endogenous Variables,” *Journal of Econometrics* 190, 212-221.

Staiger, D., and J. H. Stock. 1997. Instrumental variables regression with weak instruments. *Econometrica* 65: 557–86.

Stock, J.H. and M. Yogo (2005). “Testing for Weak Instruments in Linear IV Regression,” Ch. 5 in J.H. Stock and D.W.K. Andrews (eds), *Identification and Inference for Econometric Models: Essays in Honor of Thomas J. Rothenberg*, Cambridge University Press, 80-108.

AEA Continuing Education Course

Time Series Econometrics

Lecture 5: Dynamic factor models and prediction with large datasets

Mark W. Watson
January 8, 2019

Part 1: Dynamic factor models

… reference …

Stock, James H. and Mark W. Watson (2016)
Handbook of Macroeconomics, Vol 2. chapter

Historical Evolution of DFMs

I. Factor Analysis

- Spearman (1904)
- Lawley (1940), Joreskög (1967) … Lawley and Maxwell (1971)

Spearman's problem:

Data: X_{ij}, $i = 1, \ldots, N$ (individuals)

and $j = 1, \ldots n$ (measurements for each individual)

$$X_i = \begin{pmatrix} X_{i1} \\ X_{i2} \\ \vdots \\ X_{in} \end{pmatrix} \text{ and } \Sigma_{XX} = \text{cov}(X_i)$$

How can we measure 'intelligence'?

"GENERAL INTELLIGENCE," OBJECTIVELY DETERMINED AND MEASURED.

By C. SPEARMAN.

TABLE OF CONTENTS.

Experimental Series IV.

High Class Preparatory School for Boys.

A. Original Data.

Age		Pitch	Place in School (*before modification to eliminate Age*).												Music
			Classics			French			English			Mathem.			
Years	Months	Discrim. Thres. in ⅓ v. d., October, 1902	Xmas, 1902	Easter, 1903	July, 1903	Xmas, 1902	Easter, 1903	July, 1903	Xmas, 1902	Easter, 1903	July, 1903	Xmas, 1902	Easter, 1903	July, 1903	Ranked by Music Master
12	6	2	8	7	4	5	3	3	4	3	3	4	2	3	8
12	4	3	11	12	10	13	13	10	13	13	11	12	13	11	9
9	8	3	19	18	15	21	19	16	23	21	18	21	19	17	6
13	7	4	2	2	1	2	2	1	2	2	1	7	7	7	3
10	4	4	21		19	22		23	22		20	21		24	16
10	7	4	23	23	22	26	23	22	28	25	23	29	25	23	1
13	6	5	3			3			3			3			21
11	10	5	6	4	3	7	6	5	6	6	2	9	8	6	
10	1	5	29	26	24	23	25	21	27	26	22	25	23	19	7
11	1	6	20	20	18	20	21	18	21	20	19	17	16	15	14
13	4	7	1	1		1	1		1	1		1	1		5
10	6	7	26	24	21	27	16	13	26	19	17	22	18	16	11
12	3	7	18	17	16	17	20	19	25	23	21	19	17	14	20
13	1	8	5	5	5	4	4	2	5	8	5	5	4	1	4
11	1	10	22	19	17	19	18	17	20	17	15	23	21	21	18
9	9	10	33	29	27	33	29	27	33	27	27	32	29	27	17
10	4	11	28	25	23	30	27	24	18	18	13	30	27	22	
13	0	11	4	3	2	6	5	4	7	4	4	2	3	4	
10	2	11	7	6	6	12	7	6	8	5	8	11	9	8	
13	0	11	12	11	11	11	11	12	15	16	16	6	5	2	12
12	0	11	17	16		16	15		24	22		24	24		15
12	11	12	9	8	7	8	8	7	9	7	7	14	12	12	
13	1	14	10	9	8	10	9	8	11	10	9	10	10	9	13
10	4	14	27	21	14	24	22	15	17	11	10	26	20	18	2
10	1	15	24	22	20	18	17	14	29	24	24	18	15	13	
12	6	15	14	13	12	15	14	11	10	9	6	8	6	5	10
10	8	15	30	27		29	26		30	29		28	26		
12	8	18	16	15	13	25	24	20	14	14	12	20	21	20	19
9	5	20	32		25	31		25	32		26	33		26	
11	2	24	15	14	9	14	12	9	16	15	14	13	11	10	
10	9	50	25			28			19			15			
10	11	> 60	31	28	26	32	28	26	31	28	25	31	28	25	22
13	7	> 60	13	10		9	10		12	12		16	14		

Factor Model

$$X_{ij} = \lambda_j f_i + e_{ij} \text{ or } j = 1, \ldots, n$$

$$X_i = \lambda f_i + e_i$$

(All measurements for individual i)

$$\Sigma_{XX} = \sigma_f^2 \lambda\lambda' + \Sigma_{ee} \text{ with } \Sigma_{ee} \text{ diagonal}$$

$$X_i = \lambda f_i + e_i$$

$$\Sigma_{XX} = \sigma_f^2\ \lambda\lambda' + \Sigma_{ee} \text{ with } \Sigma_{ee} \text{ diagonal}$$

Issues:

(1) Estimation of parameters $(\sigma_f^2, \lambda, \sigma_{e_i}^2)$ (Lawley: Gaussian MLE)

(2) Estimation of $f_i \mid X_i, (\sigma_f^2, \lambda, \sigma_{e_i}^2)$: 'reverse regression'

$$(X_i \mid f_i) \sim \mathrm{N}(\lambda f_i, \Sigma_{ee}) \text{ and } f_i \sim \mathrm{N}(0, \sigma_f^2)$$

$$\Rightarrow f_i \mid X_i \sim \mathrm{N}(\beta' X_i\,, \sigma_{f|X}^2)$$

$$\text{with } \beta = \Sigma_{XX}^{-1}\Sigma_{Xf} = \left(\sigma_f^2\lambda\lambda' + \Sigma_{ee}\right)^{-1}\lambda\sigma_f^2$$

$$\sigma_{f|X}^2 = \sigma_f^2 - \sigma_f^2\lambda'\left(\sigma_f^2\lambda\lambda' + \Sigma_{ee}\right)^{-1}\lambda\sigma_f^2$$

Historical Evolution of DFMs:

2a: Replace covariance matrices with spectral density matrices. (Geweke (1977), Sargent and Sims (1977), Brillinger (1975)).

$$X_i = \lambda f_i + e_i$$

$$\Sigma_{XX} = \sigma_f^2\ \lambda\lambda' + \Sigma_{ee} \text{ with } \Sigma_{ee} \text{ diagonal}$$

becomes

$$X_t = \lambda(\mathrm{L})f_t + e_t$$

$$S_{XX}(\omega) = s_f^2(\omega)\ \lambda(\mathrm{e}^{-\mathrm{i}\omega})\lambda(\mathrm{e}^{\mathrm{i}\omega})' + S_{ee}(\omega) \text{ with } S_{ee}(\omega) \text{ diagonal}$$

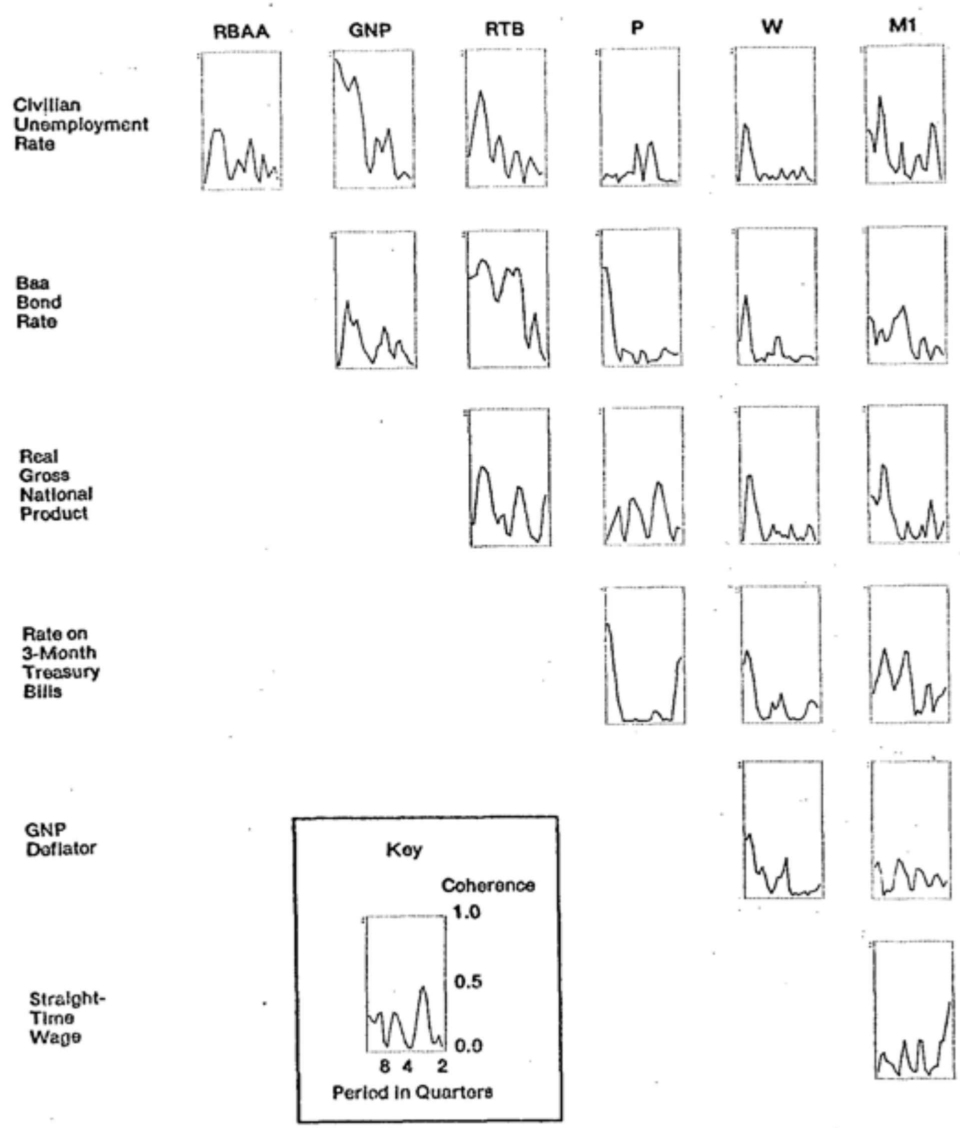
Business Cycle Modeling Without Pretending
to Have Too Much A Priori Economic Theory

Thomas J. Sargent
Christopher A. Sims

Revised, January 1977

Paper prepared for seminar on New Methods in Business Cycle Research, Federal Reserve Bank of Minneapolis, November 13-14, 1975. The views expressed herein are solely those of the authors and do not necessarily represent the views of the Federal Reserve Bank of Minneapolis or the Federal Reserve System. John Geweke adapted the maximum likelihood factor analysis algorithm for application to the frequency domain factory model and wrote a computer program for estimating and testing the one-index model. Paul Anderson extended that program to handle k noises and performed all the frequency domain calculations in this paper. Salih Neftci carried out the calculations for the observable index model. John Geweke's contribution in developing the factor analysis algorithm and in formulating the unobservable index model were enough for him to qualify as a coauthor of this paper.

Table 1 — GRAPHS OF COHERENCE OF ECONOMIC VARIABLES



Sargent and Sims used various subsets of 14 variables: long rate, short rate, GNP, prices, wages, money supply, government purchases, government deficit, unemployment rate, residential construction, inventories, plant and equip investment, consumption, corporate profits.

$$X_t = \lambda(\mathrm{L})f_t + e_t$$

$$S_{XX}(\omega) = s_f^2(\omega)\ \lambda(\mathrm{e}^{-\mathrm{i}\omega})(\mathrm{e}^{\mathrm{i}\omega})\lambda' + S_{ee}(\omega) \text{ with } S_{ee}(\omega) \text{ diagonal}$$

Issues:

(1) Estimation of parameters ($s_f^2(\omega)$, $\lambda(\mathrm{e}^{-\mathrm{i}\omega})$, $S_{ee}(\omega)$) (Local Gaussian MLE, frequency by frequency)

(2) Estimation of $f(\omega) \mid X(\omega)$: can use 'reverse regression'

New issues: Converting frequency domain back to time domain. Leads/lags. Constraints across frequencies.

2b: Use linear state-space models: (e.g., Engle and Watson (1981))

$$X_t = \lambda(\mathrm{L})f_t + e_t \text{ and } \phi(\mathrm{L})f_t = \eta_t$$

$$X_t = \begin{pmatrix} \lambda_0 & \lambda_1 & \cdots & \lambda_k \end{pmatrix} \begin{pmatrix} f_t \\ f_{t-1} \\ \vdots \\ f_{t-k} \end{pmatrix} + e_t$$

$$\begin{pmatrix} f_t \\ f_{t-1} \\ \vdots \\ f_{t-k} \end{pmatrix} = \begin{bmatrix} \phi_1 & \phi_2 & \cdots & \phi_{k+1} \\ 1 & 0 & \cdots & 0 \\ & \ddots & \ddots & \\ & & 1 & 0 \end{bmatrix} \begin{pmatrix} f_{t-1} \\ f_{t-2} \\ \vdots \\ f_{t-k-1} \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \eta_t$$

or

$$X_t = \Lambda F_t + e_t$$
$$F_t = \Phi F_{t-1} + \mathrm{G}\eta_t$$

(More generally F equation can be VAR(p): $\Phi(\mathrm{L})F_t = \mathrm{G}\eta_t$)

Issues:

(1) Estimation of parameters (Λ, σ_η^2, Φ, Σ_{ee}) (Gaussian MLE using prediction-error decomposition from Kalman filter)

(2) Estimation of $f_t \mid \left\{X_j\right\}_{j=1}^{T}$: 'reverse regression' computed using Kalman smoother.

New issues:

(a) State-space modeling afforded lots of flexibility.

(b) MLE hard when X_t is high dimensional. (Quah and Sargent (1993))

Example: “Improving GDP Measurement: A Measurement-Error Perspective” Aruoba, Diebold, Nalewaik, Schorfheide, Song (2016)

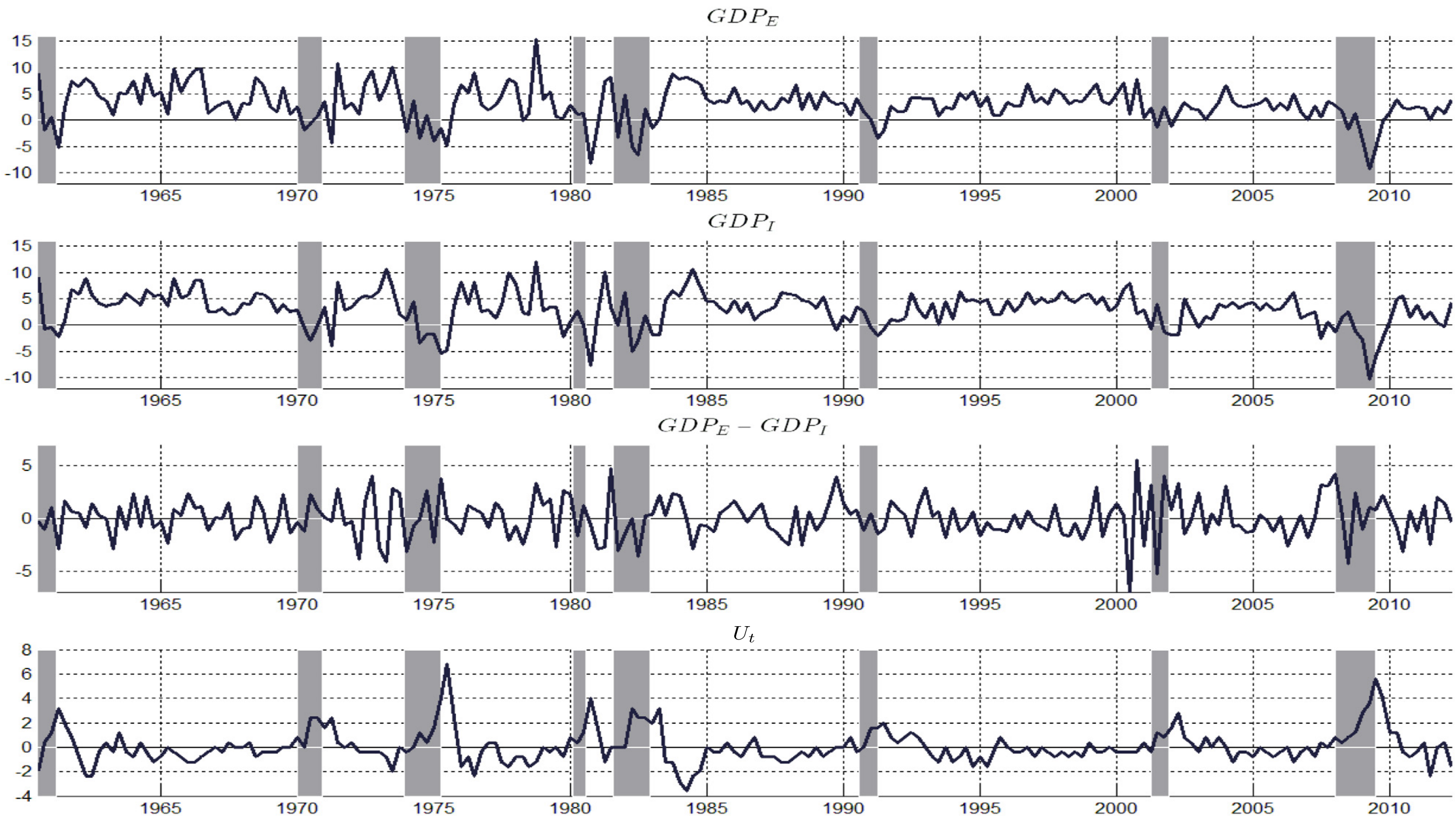


Fig. 1. *GDP* and unemployment data. GDP_E and GDP_I are in growth rates and U_t is in changes. All are measured in annualized percent.

$$\begin{bmatrix} GDP_{Et} \\ GDP_{It} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} GDP_t + \begin{bmatrix} \varepsilon_{Et} \\ \varepsilon_{It} \end{bmatrix}$$

$$GDP_t = \alpha + \rho GDP_{t-1} + \varepsilon_{Gt}$$

$$\mathrm{var}\begin{bmatrix} \varepsilon_g \\ \varepsilon_E \\ \varepsilon_I \end{bmatrix} = \Sigma = \begin{bmatrix} \sigma_{GG} & 0 & 0 \\ & \sigma_{EE} & \sigma_{EI} \\ & & \sigma_{II} \end{bmatrix} \text{ (identification issues)}$$

Results:

For the 2-equation model with Σ block-diagonal, we have

$$GDP_t = \underset{[2.77,3.34]}{3.06}(1-0.62) + \underset{[0.57,0.68]}{0.62} GDP_{t-1} + \epsilon_{Gt}, \tag{12}$$

$$\Sigma = \begin{bmatrix} \underset{[4.39,5.95]}{5.17} & 0 & 0 \\ 0 & \underset{[3.34,4.48]}{3.86} & \underset{[0.96,1.95]}{1.43} \\ 0 & \underset{[0.96,1.95]}{1.43} & \underset{[2.25,3.22]}{2.70} \end{bmatrix}. \tag{13}$$

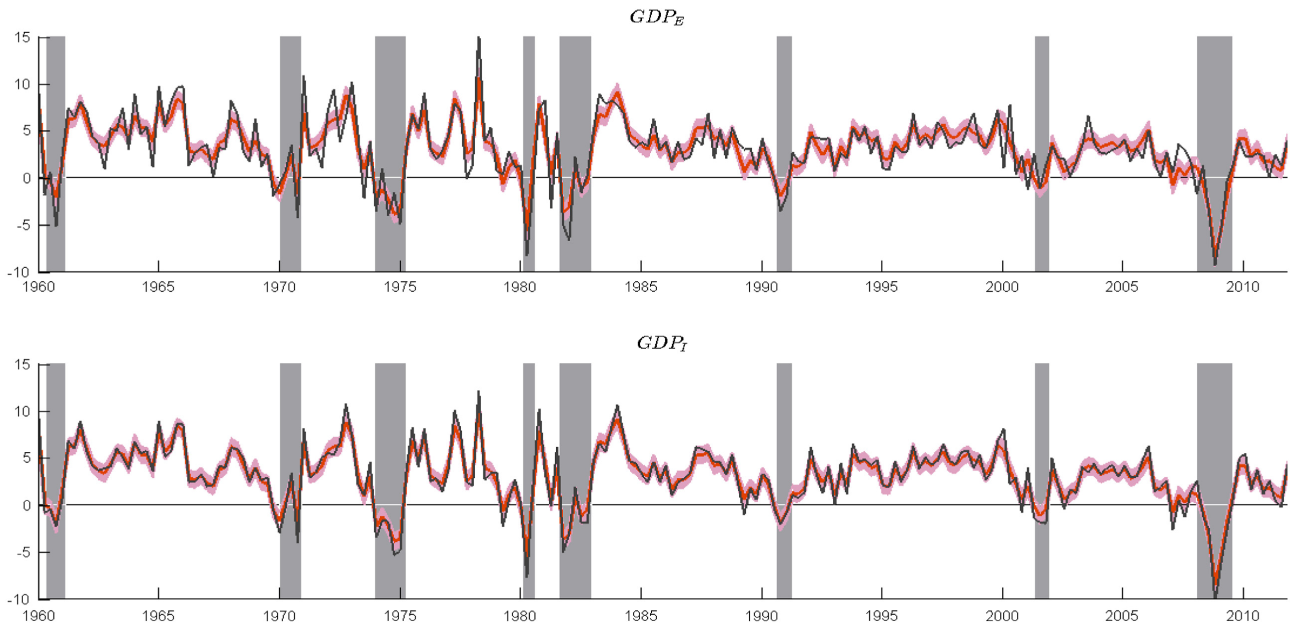
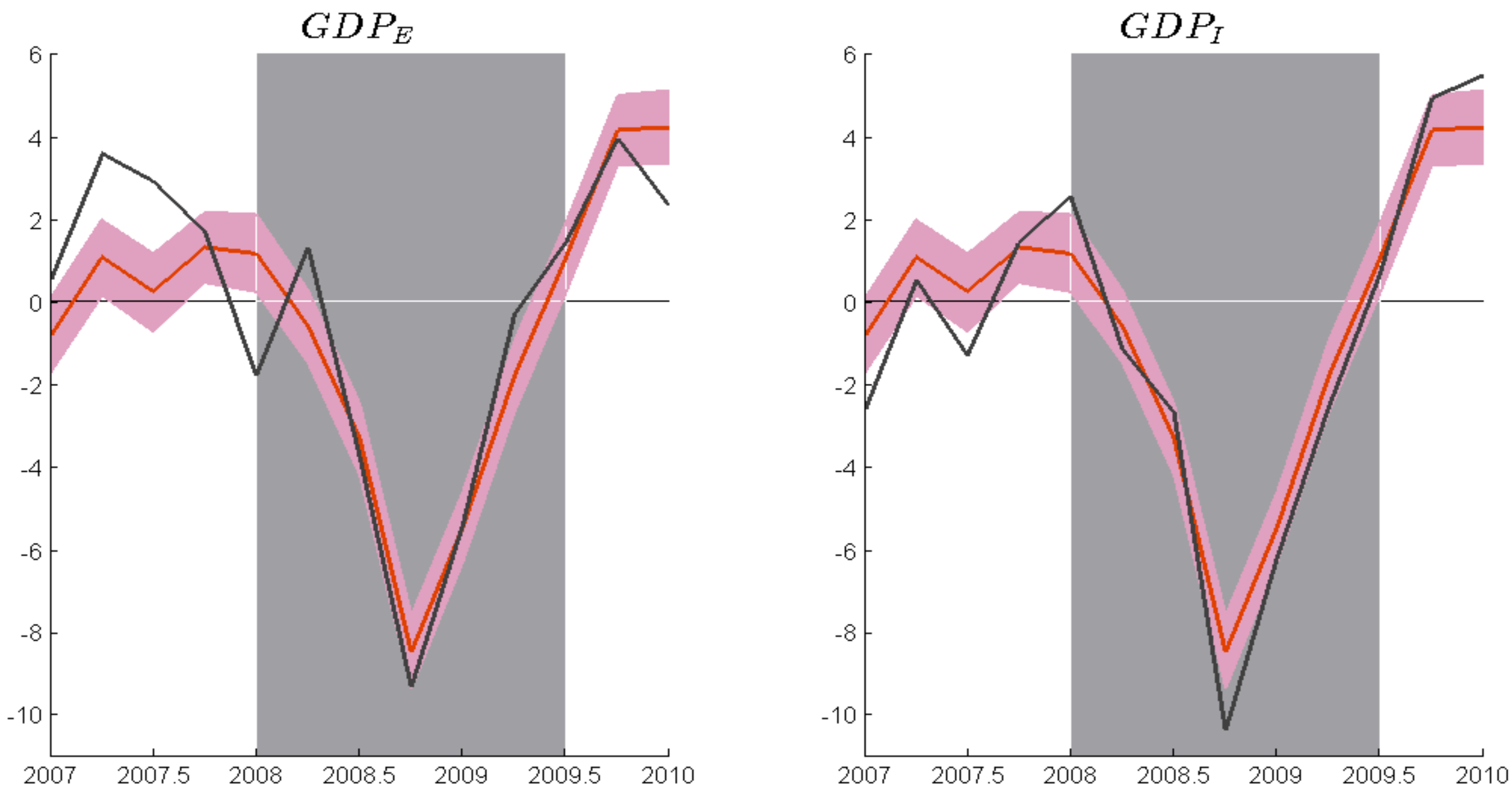


Fig. 3. *GDP* sample paths, 1960Q1–2011Q4. In each panel we show the sample path of GDP_M (light color) together with posterior interquartile range with shading and we show one of the competitor series (dark color). For GDP_M we use our benchmark estimate from the 2-equation model with $\zeta = 0.80$.

Figure 4: GDP Sample Paths, 2007Q1-2009Q4



Historical Evolution of DFMs:

3. Large-n approximations. Connor and Korajczyk (1986), Chamberlain and Rothschild (1983), Forni and Reichlin (1998), Stock and Watson (2002), …

Large n ... from curse to blessing: *An example following Forni and Reichlin (1998)*. Suppose f_t is scalar and $\lambda(L) = \lambda$ ("no lags in the factor loadings"), so

$$X_{it} = \lambda_i f_t + e_{it} \quad \text{for } i = 1, \dots n$$

Then: $$\frac{1}{n}\sum_{i=1}^{n} X_{it} = \frac{1}{n}\sum_{i=1}^{n}(\lambda_i f_t + e_{it}) = \left(\frac{1}{n}\sum_{i=1}^{n}\lambda_i\right) f_t + \frac{1}{n}\sum_{i=1}^{n} e_{it}$$

If the errors e_{it} have limited dependence across series, then as n gets large,

$$\frac{1}{n}\sum_{i=1}^{n} X_{it} \xrightarrow{p} \bar{\lambda} f_t$$

Large n lets us recover f_t up to a scale factor.

A “least squares” reason to use the sample mean.

Consider

$$\min_{\{f_t\},\{\lambda_i\}} \sum_{i,t} (X_{it} - \lambda_i f_t)^2 \quad \text{subject to } \bar{\lambda} = 1$$

Yields: $\hat{f}_t = \frac{1}{n}\sum_{i=1}^{n} X_{it}$

(Other normalizations: $T^{-1}\sum_{t=1}^{T} f_t^2 = 1$)

Multivariate Problem: $X_{it} = \lambda_i' F_t + e_{it}$, where λ_i' is i^{th} row of Λ.

$$\min_{\{f_t\},\{\lambda_i\}} \sum_{i,t} (X_{it} - \lambda_i' F_t)^2 \quad \text{subject } T^{-1}\sum_{t=1}^{T} F_t F_t' = \Gamma \text{ (diagonal, with } \gamma_i \geq \gamma_{i+1}\text{)}$$

Yields: $\hat{F}_t$ as the principal components (PC) of X_t, (i.e., the linear combinations of X_t with the largest variance).

Odds and ends:

Missing data

Weighted least squares

…

More generally

$X_t = \lambda(\mathrm{L})f_t + e_t$ and $\phi(\mathrm{L})f_t = \eta_t \Rightarrow X_t = \Lambda F_t + e_t$ and $\Phi(\mathrm{L})F_t = \mathrm{G}\eta_t$

So Principal Components (PC) can be used to estimate F in DFM.

A simple 2-step estimation problem:

(1) Estimate F_t by PC

(2) Estimate λ_i and var(e_{it}) from regression of X_{it} onto $\hat{F}_t$.

(3) Estimate dynamic equation for F using VAR with $\hat{F}_t$ replacing F.

Some results about these simple 2-step estimators when n and T are large:

Results for the exact static factor model:

Connor and Korajczyk (1986): consistency in the exact static FM with T fixed, $n \to \infty$.

Selected results for the approximate DFM: $X_t = \Lambda F_t + e_t$

Typical conditions (Stock-Watson (2002), Bai-Ng (2002, 2006)):

(a) $\frac{1}{T}\sum_{i=1}^{T} F_t F_t' \xrightarrow{p} \Sigma_F$ (stationary factors)

(b) $\Lambda'\Lambda/n \to$ (or $\xrightarrow{p}$) Σ_Λ Full rank factor loadings

(c) e_{it} are weakly dependent over time and across series

(d) F, e are uncorrelated at all leads and lags

Selected results for the approximate DFM, ctd.

Stock and Watson (2002)

- consistency in the approximate DFM, n, $T \to \infty$.
- justify using $\hat{F}_t$ as a regressor (no errors-in-variable bias. etc.)
- oracle property for forecasts

Bai and Ng (2006)

- $N^2/T \to \infty$
- asymptotic normality of PC estimator of the common component at rate min($n^{1/2}$, $T^{1/2}$) in approximate DFM. These can be used to compute confidence sets for $F_{t.}$
- Similar results are rates for the two estimators of Λ, Φ, Σ_{ee} and $\Sigma_{\eta\eta}$.

Historical Evolution of DFMs:

An issue in PC estimates of DFMs: F_t is estimated using averages of X_t. This ignores information in leads and lags of X that would be utilized using optimal estimator (Kalman smoother).

4. Hybrid estimators: Use PCs to get first-round estimates of Λ, Φ, Σ_{ee} and $\Sigma_{\eta\eta}$, then use Kalman smoother to get estimates of F, or do MLE using these as initial guesses of parameters. (Doz, Giannone, Reichlin (2011, 2012).)

Example: Nowcasting (Good reference: Banbura, Giannoni, Modugno, and Reichlin (2013).)

- Problem: y_t is a variable of interest (e.g., GDP growth rate in quarter t). It is available with a lag (say in t+1 or t+2). X_t is a vector of variables that are measured *during* period t (and perhaps earlier). How do you guess the value of y_t given the X data that has been revealed.

- 'Solution': Suppose X_{t_1} denotes the information known at time t_1. Then best guess of y_t is $E(y_t| X_{t_1})$.
  - But how do you compute $E(y_t| X_{t_1})$?
  - How do you update the estimate as another element of X_t is revealed?

Giannone, Reichlin, *et al* modeling approach:

$$\begin{bmatrix} y_t \\ X_{1t} \\ \vdots \\ X_{nt} \end{bmatrix} = \begin{bmatrix} \lambda_y \\ \lambda_1 \\ \vdots \\ \lambda_n \end{bmatrix} F_t + \begin{bmatrix} e_{yt} \\ e_{1t} \\ \vdots \\ e_{nt} \end{bmatrix}$$

$$\Phi(\mathrm{L})F_t = \eta_t$$

- $\mathrm{E}(y_t | X_{t_1}) = \lambda_y \times \mathrm{E}(F_t \mid X_{t_1})$
- $\mathrm{E}(F_t \mid X_{t_1})$ computed by Kalman filter

(Lots of details left out)

FEDERAL RESERVE BANK *of* NEW YORK *Serving the Second District and the Nation*

search

About the New York Fed | Markets & Policy Implementation | Economic Research | Financial Institution Supervision | Financial Services & Infrastructure | Outreach & Educati on

home > economic research >

Nowcasting Report

OVERVIEW | NOWCAST | METHODOLOGY | FAQS

Dec 21, 2018: New York Fed Staff Nowcast

- The New York Fed Staff Nowcast stands at 2.5% for 2018:Q4 and 2.1% for 2019:Q1.

MORE

2019:Q1 | **2018:Q4** | 2018:Q3 | 2018:Q2

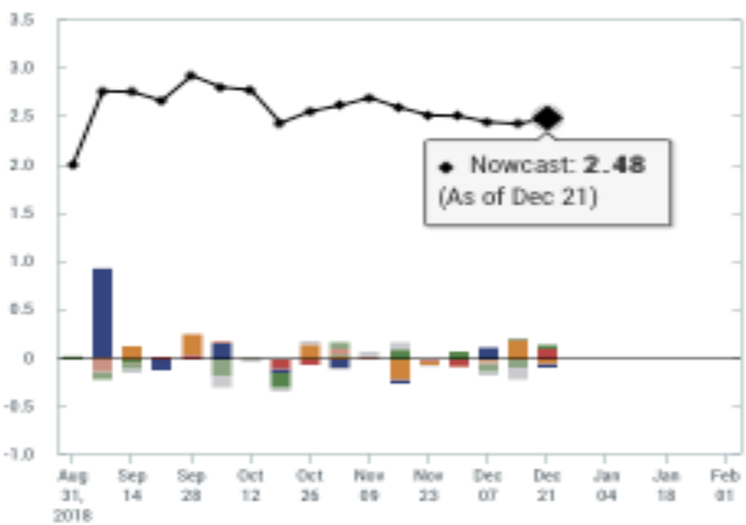
Last Release 11:15am EST Dec 21, 2018

ARCHIV LAYOUT

◆ The New York Fed Staff Nowcast ○ Advance GDP estimate □ Latest GDP estimate

■ Housing and construction ■ Manufacturing ■ Surveys ■ Retail and consumption ■ Income ■ Labor ■ International trade ■ Others

Percent (annual rate)



Expand

Data Flow (Dec 21, 2018)

Model Update	Release Date	Data Series	Actual	Impact	Nowcast GDP Growth
Dec 21					**2.48**
	10:00AM Dec 21	■ Real personal consumption expenditures	0.33	0.04	
	10:00AM Dec 21	■ PCE: Chain price index	0.06	-0.01	
	10:00AM Dec 21	■ PCE less food and energy: Chain price index	0.15	0.00	
	10:00AM Dec 21	■ Real disposable personal income	0.18	-0.00	
	8:30AM Dec 21	■ Manufacturers' inventories: Durable	0.26	0.00	

1.1 | Nowcast Detail

Housing and construction · Manufacturing · Surveys · Retail and consumption · Income · Labor · International trade · Others

Update	Release Date	Data Series	Reference Period	Units	Forecast [a]	Actual [b]	Weight [c]	Impact [$c(b-a)$]	Nowcast GDP Growth
Nov 23									**2.51**
	8:30 AM Nov 28	Merchant wholesalers: Inventories: Total	Oct	MoM % chg.	0.541	0.678	-0.156	-0.021	
	8:30 AM Nov 28	Real gross domestic income	Q3	QoQ % chg. AR	2.26	3.97	0.013	0.023	
	10:00 AM Nov 28	New single family houses sold	Oct	MoM % chg.	-0.140	-8.88	0.007	-0.065	
	8:30 AM Nov 29	Real disposable personal income	Oct	MoM % chg.	0.213	0.342	0.025	0.003	
	8:30 AM Nov 29	PCE less food and energy: Chain price index	Oct	MoM % chg.	0.153	0.102	0.254	-0.013	
	8:30 AM Nov 29	PCE: Chain price index	Oct	MoM % chg.	0.160	0.181	0.144	0.003	
	8:40 AM Nov 29	Real personal consumption expenditures	Oct	MoM % chg.	0.214	0.436	0.307	0.068	
		Data revisions						-0.011	
Nov 30									**2.50**
	10:00 AM Dec 03	ISM mfg.: PMI composite index	Nov	Index	57.0	59.3	0.049	0.111	
	10:00 AM Dec 03	ISM mfg.: Prices index	Nov	Index	69.6	60.7	0.008	-0.069	
	10:00 AM Dec 03	Value of construction put in place	Oct	MoM % chg.	0.203	-0.149	0.024	-0.009	
	10:00 AM Dec 03	ISM mfg.: Employment index	Nov	Index	55.7	58.4	0.027	0.073	
	8:05 AM Dec 06	ADP nonfarm private payroll employment	Nov	Level chg. (thousands)	205.7	179.0	0.543*	-0.014	
	8:30 AM Dec 06	Exports: Goods and services	Oct	MoM % chg.	0.743	-0.148	0.073	-0.065	
	8:30 AM Dec 06	Imports: Goods and services	Oct	MoM % chg.	0.618	0.234	0.056	-0.022	
	10:00 AM Dec 06	ISM nonmanufacturing: NMI composite index	Nov	Index	59.8	60.7	0.007	0.007	
	10:00 AM Dec 06	Inventories: Total business	Oct	MoM % chg.	0.467	0.558	-0.099	-0.009	
	8:30 AM Dec 07	All employees: Total nonfarm	Nov	Level chg. (thousands)	199.2	155.0	0.316*	-0.014	
	8:30 AM Dec 07	Civilian unemployment rate	Nov	Ppt. chg.	-0.071	0.000	-0.106	-0.007	
		Data revisions						-0.044	
Dec 07									**2.44**
	10:00 AM Dec 10	JOLTS: Job openings: Total	Oct	Level chg. (thousands)	203.4	119.0	-0.061*	0.005	
	8:30 AM Dec 11	PPI: Final demand	Nov	MoM % chg.	0.241	0.085	0.105	-0.016	
	8:40 AM Dec 12	CPI-U: All items	Nov	MoM % chg.	0.263	0.019	0.084	-0.021	
	8:40 AM Dec 12	CPI-U: All items less food and energy	Nov	MoM % chg.	0.173	0.209	0.099	0.004	
	8:30 AM Dec 13	Export price index	Nov	MoM % chg.	0.334	-0.860	0.044	-0.052	
	8:30 AM Dec 13	Import price index	Nov	MoM % chg.	0.433	-1.56	0.022	-0.045	
	8:30 AM Dec 14	Retail sales and food services	Nov	MoM % chg.	0.168	0.229	0.166	0.010	
	9:10 AM Dec 14	Industrial production index	Nov	MoM % chg.	0.177	0.606	0.263	0.113	
	9:10 AM Dec 14	Capacity utilization	Nov	Ppt. chg.	0.101	0.329	0.342	0.078	
		Data revisions						-0.092	
Dec 14									**2.42**
	8:30 AM Dec 17	Empire State Mfg. Survey: General business conditions	Dec	Index	22.1	10.9	0.003	-0.033	
	8:30 AM Dec 18	Housing starts	Nov	MoM % chg.	1.47	3.20	0.012	0.021	
	8:30 AM Dec 18	Building permits	Nov	Level chg. (thousands)	-0.014	63.0	0.001	0.091	
	8:30 AM Dec 20	Phila. Fed Mfg. business outlook: Current activity	Dec	Index	17.4	9.40	0.001	-0.009	
	8:30 AM Dec 21	Manufacturers' new orders: Durable goods	Nov	MoM % chg.	3.08	0.758	0.014	-0.033	
	8:30 AM Dec 21	Manufacturers' shipments: Durable goods	Nov	MoM % chg.	0.935	0.707	0.088	-0.020	
	8:30 AM Dec 21	Mfrs.' unfilled orders: All manufacturing industries	Nov	MoM % chg.	0.274	-0.144	-0.007	0.003	
	8:30 AM Dec 21	Manufacturers' inventories: Durable goods	Nov	MoM % chg.	0.256	0.256	-0.134	0.000	
	10:00 AM Dec 21	Real disposable personal income	Nov	MoM % chg.	0.188	0.184	0.017	-0.000	
	10:00 AM Dec 21	PCE less food and energy: Chain price index	Nov	MoM % chg.	0.148	0.148	0.163	0.000	
	10:00 AM Dec 21	PCE: Chain price index	Nov	MoM % chg.	0.181	0.056	0.082	-0.010	
	10:00 AM Dec 21	Real personal consumption expenditures	Nov	MoM % chg.	0.158	0.326	0.215	0.036	
		Data revisions						0.008	
Dec 21									**2.48**

Historical Evolution of DFMs:

Issue: Many parameters in DFM. Shrinkage might be useful.

5. Bayes estimators (Kim and Nelson (1998), Otrok and Whiteman (1998), …)

$$X_t = \Lambda F_t + e_t \text{ and } \Phi(\mathrm{L})F_t = \mathrm{G}\eta_t$$

Model is particularly amenable to MCMC methods:

(i) $(\Lambda, \Sigma_{ee}, \Phi, \Sigma_{\eta\eta} \mid \{X_t, F_t\})$: Linear regression problem

(ii) $(\{F_t\} \mid \{X_t\}, \Lambda, \Sigma_{ee}, \Phi, \Sigma_{\eta\eta})$: Linear signal extraction problem

$$X_t = \Lambda F_t + e_t \quad \text{and } \Phi(\mathrm{L})F_t = \mathrm{G}\,\eta_t$$

Generalizations (see SW HOM paper for references):

(1) Serial correlation in e

(2) Additional regressors in either equation

(3) Constraints on Λ ('sparsity')

(4) (Limited) cross-correlation between elements of e.

(5) Non-linearities and non-Gaussian evolution.
Examples: Markov Switching in F, etc.

(6) Robustness to various types of instability.

… many more.

A 207-Variable Macro Dataset for the U.S.

Table 1 Quarterly time series in the full dataset

	Category	Number of series	Number of series used for factor estimation
(1)	NIPA	20	12
(2)	Industrial production	11	7
(3)	Employment and unemployment	45	30
(4)	Orders, inventories, and sales	10	9
(5)	Housing starts and permits	8	6
(6)	Prices	37	24
(7)	Productivity and labor earnings	10	5
(8)	Interest rates	18	10
(9)	Money and credit	12	6
(10)	International	9	9
(11)	Asset prices, wealth, and household balance sheets	15	10
(12)	Other	2	2
(13)	Oil market variables	10	9
	Total	207	139

Notes: The real activity dataset consists of the variables in the categories 1–4.

Table A.1: Data Series

	Name	Description	Sample Period	T	O	F
	(1) NIPA					
1	GDP	Real Gross Domestic Product 3 Decimal	1959:Q1-2014:Q4	5	0	0
2	Consumption	Real Personal Consumption Expenditures	1959:Q1-2014:Q4	5	0	0
3	Cons:Dur	Real Personal Consumption Expenditures: Durable Goods Quantity Index	1959:Q1-2014:Q4	5	0	1
4	Cons:Svc	Real Personal Consumption Expenditures: Services Quantity Index	1959:Q1-2014:Q4	5	0	1
5	Cons:NonDur	Real Personal Consumption Expenditures: Nondurable Goods Quantity Index	1959:Q1-2014:Q4	5	0	1
6	Investment	Real Gross Private Domestic Investment 3 Decimal	1959:Q1-2014:Q4	5	0	0
7	FixedInv	Real Private Fixed Investment Quantity Index	1959:Q1-2014:Q4	5	0	0
8	Inv:Equip	Real Nonresidential Investment: Equipment Quantity Idenx	1959:Q1-2014:Q4	5	0	1
9	FixInv:NonRes	Real Private Nonresidential Fixed Investment Quantity Index	1959:Q1-2014:Q4	5	0	1
10	FixedInv:Res	Real Private Residential Fixed Investment Quantity Index	1959:Q1-2014:Q4	5	0	1
11	Ch. Inv/GDP	Change in Inventories /GDP	1959:Q1-2014:Q4	1	0	1
12	Gov.Spending	Real Government Consumption Expenditures & Gross Investment 3 Decimal	1959:Q1-2014:Q4	5	0	0
13	Gov:Fed	Real Federal Consumption Expenditures Quantity Index	1959:Q1-2014:Q4	5	0	1
14	Real_Gov Receipts	Government Current Receipts (Nominal) Defl by GDP Deflator	1959:Q1-2014:Q3	5	0	1
15	Gov:State&Local	Real State & Local Consumption Expenditures Quantity Index	1959:Q1-2014:Q4	5	0	1
16	Exports	Real Exports of Goods & Services 3 Decimal	1959:Q1-2014:Q4	5	0	1
17	Imports	Real Imports of Goods & Services 3 Decimal	1959:Q1-2014:Q4	5	0	1
18	Disp-Income	Real Disposable Personal Income	1959:Q1-2014:Q4	5	0	0
19	Ouput:NFB	Nonfarm Business Sector: Output	1959:Q1-2014:Q4	5	0	0
20	Output:Bus	Business Sector: Output	1959:Q1-2014:Q4	5	0	0
	(2) Industrial Production					
21	IP: Total index	IP: Total index	1959:Q1-2014:Q4	5	0	0
22	IP: Final products	Industrial Production: Final Products (Market Group)	1959:Q1-2014:Q4	5	0	0
23	IP: Consumer goods	IP: Consumer goods	1959:Q1-2014:Q4	5	0	0
24	IP: Materials	Industrial Production: Materials	1959:Q1-2014:Q4	5	0	0
25	IP: Dur gds materials	Industrial Production: Durable Materials	1959:Q1-2014:Q4	5	0	1
26	IP: Nondur gds materials	Industrial Production: nondurable Materials	1959:Q1-2014:Q4	5	0	1
27	IP: Dur Cons. Goods	Industrial Production: Durable Consumer Goods	1959:Q1-2014:Q4	5	0	1
28	IP: Auto	IP: Automotive products	1959:Q1-2014:Q4	5	0	1
29	IP:NonDur Cons God	Industrial Production: Nondurable Consumer Goods	1959:Q1-2014:Q4	5	0	1
30	IP: Bus Equip	Industrial Production: Business Equipment	1959:Q1-2014:Q4	5	0	1
31	Capu Tot	Capacity Utilization: Total Industry	1967:Q1-2014:Q4	1	0	1
	(3) Employment and Unemployment					
32	Emp:Nonfarm	Total Nonfarm Payrolls: All Employees	1959:Q1-2014:Q4	5	0	0
33	Emp: Private	All Employees: Total Private Industries	1959:Q1-2014:Q4	5	0	0
34	Emp: mfg	All Employees: Manufacturing	1959:Q1-2014:Q4	5	0	0
35	Emp:Services	All Employees: Service-Providing Industries	1959:Q1-2014:Q4	5	0	0

36	Emp:Goods	All Employees: Goods-Producing Industries	1959:Q1-2014:Q4	5	0	0
37	Emp: DurGoods	All Employees: Durable Goods Manufacturing	1959:Q1-2014:Q4	5	0	1
38	Emp: Nondur Goods	All Employees: Nondurable Goods Manufacturing	1959:Q1-2014:Q4	5	0	0
39	Emp: Const	All Employees: Construction	1959:Q1-2014:Q4	5	0	1
40	Emp: Edu&Health	All Employees: Education & Health Services	1959:Q1-2014:Q4	5	0	1
41	Emp: Finance	All Employees: Financial Activities	1959:Q1-2014:Q4	5	0	1
42	Emp: Infor	All Employees: Information Services	1959:Q1-2014:Q4	5	1	1
43	Emp: Bus Serv	All Employees: Professional & Business Services	1959:Q1-2014:Q4	5	0	1
44	Emp:Leisure	All Employees: Leisure & Hospitality	1959:Q1-2014:Q4	5	0	1
45	Emp:OtherSvcs	All Employees: Other Services	1959:Q1-2014:Q4	5	0	1
46	Emp: Mining/NatRes	All Employees: Natural Resources & Mining	1959:Q1-2014:Q4	5	1	1
47	Emp:Trade&Trans	All Employees: Trade Transportation & Utilities	1959:Q1-2014:Q4	5	0	1
48	Emp: Gov	All Employees: Government	1959:Q1-2014:Q4	5	0	0
49	Emp:Retail	All Employees: Retail Trade	1959:Q1-2014:Q4	5	0	1
50	Emp:Wholesal	All Employees: Wholesale Trade	1959:Q1-2014:Q4	5	0	1
51	Emp: Gov(Fed)	Employment Federal Government	1959:Q1-2014:Q4	5	2	1
52	Emp: Gov (State)	Employment State government	1959:Q1-2014:Q4	5	0	1
53	Emp: Gov (Local)	Employment Local government	1959:Q1-2014:Q4	5	0	1
54	Emp: Total (HHSurve)	Emp Total (Household Survey)	1959:Q1-2014:Q4	5	0	0
55	LF Part Rate	LaborForce Participation Rate (16 Over) SA	1959:Q1-2014:Q4	2	0	0
56	Unemp Rate	Urate	1959:Q1-2014:Q4	2	0	0
57	Urate_ST	Urate Short Term (< 27 weeks)	1959:Q1-2014:Q4	2	0	0
58	Urate_LT	Urate Long Term (>= 27 weeks)	1959:Q1-2014:Q4	2	0	0
59	Urate: Age16-19	Unemployment Rate - 16-19 yrs	1959:Q1-2014:Q4	2	0	1
60	Urate:Age>20 Men	Unemployment Rate - 20 yrs. & over Men	1959:Q1-2014:Q4	2	0	1
61	Urate: Age>20 Women	Unemployment Rate - 20 yrs. & over Women	1959:Q1-2014:Q4	2	0	1
62	U: Dur<5wks	Number Unemployed for Less than 5 Weeks	1959:Q1-2014:Q4	5	0	1
63	U:Dur5-14wks	Number Unemployed for 5-14 Weeks	1959:Q1-2014:Q4	5	0	1
64	U:dur>15-26wks	Civilians Unemployed for 15-26 Weeks	1959:Q1-2014:Q4	5	0	1
65	U: Dur>27wks	Number Unemployed for 27 Weeks & over	1959:Q1-2014:Q4	5	0	1
66	U: Job losers	Unemployment Level - Job Losers	1967:Q1-2014:Q4	5	0	1
67	U: LF Reenty	Unemployment Level - Reentrants to Labor Force	1967:Q1-2014:Q4	5	1	1
68	U: Job Leavers	Unemployment Level - Job Leavers	1967:Q1-2014:Q4	5	0	1
69	U: New Entrants	Unemployment Level - New Entrants	1967:Q1-2014:Q4	5	1	1
70	Emp:SlackWk	Employment Level - Part-Time for Economic Reasons All Industries	1959:Q1-2014:Q4	5	1	1
71	EmpHrs:Bus Sec	Business Sector: Hours of All Persons	1959:Q1-2014:Q4	5	0	0
72	EmpHrs:nfb	Nonfarm Business Sector: Hours of All Persons	1959:Q1-2014:Q4	5	0	0
73	AWH Man	Average Weekly Hours: Manufacturing	1959:Q1-2014:Q4	1	0	1
74	AWH Privat	Average Weekly Hours: Total Private Industry	1964:Q1-2014:Q4	2	0	1
75	AWH Overtime	Average Weekly Hours: Overtime: Manufacturing	1959:Q1-2014:Q4	2	0	1
76	HelpWnted	Index of Help-Wanted Advertising in Newspapers (Data truncated in 2000)	1959:Q1-1999:Q4	1	0	0
	(4) Orders, Inventories, and Sales					
77	MT Sales	Manufacturing and trade sales (mil. Chain 2005 $)	1959:Q1-2014:Q3	5	0	0
78	Ret. Sale	Sales of retail stores (mil. Chain 2000 $)	1959:Q1-2014:Q3	5	0	1

79	Orders (DurMfg)	Mfrs' new orders durable goods industries (bil. chain 2000 $)	1959:Q1-2014:Q4	5	0	1
80	Orders (Cons. Gds & Mat.)	Mfrs' new orders consumer goods and materials (mil. 1982 $)	1959:Q1-2014:Q4	5	0	1
81	UnfOrders(DurGds)	Mfrs' unfilled orders durable goods indus. (bil. chain 2000 $)	1959:Q1-2014:Q4	5	0	1
82	Orders(NonDefCap)	Mfrs' new orders nondefense capital goods (mil. 1982 $)	1959:Q1-2014:Q4	5	0	1
83	VendPerf	ISM Manufacturing: Supplier Deliveries Index©	1959:Q1-2014:Q4	1	0	1
84	NAPM:INV	ISM Manufacturing: Inventories Index©	1959:Q1-2014:Q4	1	0	1
85	NAPM:ORD	ISM Manufacturing: New Orders Index©; Index;	1959:Q1-2014:Q4	1	0	1
86	MT Invent	Manufacturing and trade inventories (bil. Chain 2005 $)	1959:Q1-2014:Q3	5	0	1
	(5) Housing Starts and Permits					
87	Hstarts	Housing Starts: Total: New Privately Owned Housing Units Started	1959:Q1-2014:Q3	5	0	0
88	Hstarts >5units	Privately Owned Housing Starts: 5-Unit Structures or More	1959:Q1-2014:Q3	5	0	0
89	Hpermits	New Private Housing Units Authorized by Building Permit	1960:Q1-2014:Q4	5	0	1
90	Hstarts:MW	Housing Starts in Midwest Census Region	1959:Q1-2014:Q3	5	0	1
91	Hstarts:NE	Housing Starts in Northeast Census Region	1959:Q1-2014:Q3	5	0	1
92	Hstarts:S	Housing Starts in South Census Region	1959:Q1-2014:Q3	5	0	1
93	Hstarts:W	Housing Starts in West Census Region	1959:Q1-2014:Q3	5	0	1
94	Constr. Contracts	Construction contracts (mil. sq. ft.) (Copyright McGraw-Hill)	1963:Q1-2014:Q4	4	0	1
	(6) Prices					
95	PCED	Personal Consumption Expenditures: Chain-type Price Index	1959:Q1-2014:Q4	6	0	0
96	PCED_LFE	Personal Consumption Expenditures: Chain-type Price Index Less Food and Energy	1959:Q1-2014:Q4	6	0	0
97	GDP Defl	Gross Domestic Product: Chain-type Price Index	1959:Q1-2014:Q4	6	0	0
98	GPDI Defl	Gross Private Domestic Investment: Chain-type Price Index	1959:Q1-2014:Q4	6	0	1
99	BusSec Defl	Business Sector: Implicit Price Deflator	1959:Q1-2014:Q4	6	0	1
100	PCED_Goods	Goods	1959:Q1-2014:Q4	6	0	0
101	PCED_DurGoods	Durable goods	1959:Q1-2014:Q4	6	0	0
102	PCED_NDurGoods	Nondurable goods	1959:Q1-2014:Q4	6	0	0
103	PCED_Serv	Services	1959:Q1-2014:Q4	6	0	0
104	PCED_HouseholdServices	Household consumption expenditures (for services)	1959:Q1-2014:Q4	6	0	0
105	PCED_MotorVec	Motor vehicles and parts	1959:Q1-2014:Q4	6	0	1
106	PCED_DurHousehold	Furnishings and durable household equipment	1959:Q1-2014:Q4	6	0	1
107	PCED_Recreation	Recreational goods and vehicles	1959:Q1-2014:Q4	6	0	1
108	PCED_OthDurGds	Other durable goods	1959:Q1-2014:Q4	6	0	1
109	PCED_Food_Bev	Food and beverages purchased for off-premises consumption	1959:Q1-2014:Q4	6	0	1
110	PCED_Clothing	Clothing and footwear	1959:Q1-2014:Q4	6	0	1
111	PCED_Gas_Enrgy	Gasoline and other energy goods	1959:Q1-2014:Q4	6	0	1
112	PCED_OthNDurGds	Other nondurable goods	1959:Q1-2014:Q4	6	0	1
113	PCED_Housing-Utilities	Housing and utilities	1959:Q1-2014:Q4	6	0	1
114	PCED_HealthCare	Health care	1959:Q1-2014:Q4	6	0	1
115	PCED_TransSvg	Transportation services	1959:Q1-2014:Q4	6	0	1
116	PCED_RecServices	Recreation services	1959:Q1-2014:Q4	6	0	1
117	PCED_FoodServ_Acc.	Food services and accommodations	1959:Q1-2014:Q4	6	0	1

118	PCED_FIRE	Financial services and insurance	1959:Q1-2014:Q4	6	0	1
119	PCED_OtherServices	Other services	1959:Q1-2014:Q4	6	0	1
120	CPI	Consumer Price Index For All Urban Consumers: All Items	1959:Q1-2014:Q4	6	0	0
121	CPI_LFE	Consumer Price Index for All Urban Consumers: All Items Less Food & Energy	1959:Q1-2014:Q4	6	0	0
122	PPI:FinGds	Producer Price Index: Finished Goods	1959:Q1-2014:Q4	6	0	0
123	PPI	Producer Price Index: All Commodities	1959:Q1-2014:Q3	6	0	0
124	PPI:FinConsGds	Producer Price Index: Finished Consumer Goods	1959:Q1-2014:Q4	6	0	1
125	PPI:FinConsGds (Food)	Producer Price Index: Finished Consumer Foods	1959:Q1-2014:Q4	6	0	1
126	PPI:IndCom	Producer Price Index: Industrial Commodities	1959:Q1-2014:Q4	6	0	1
127	PPI:IntMat	Producer Price Index: Intermediate Materials: Supplies & Components	1959:Q1-2014:Q4	6	0	1
128	Real_P:SensMat	Index of Sensitive Matrerials Prices (Discontinued) Defl by PCE(LFE) Def	1959:Q1-2004:Q1	5	0	1
129	Real_Commod: spot price	Spot market price index:BLS & CRB: all commodities(1967=100) Defl by PCE(LFE)	1959:Q1-2009:Q1	5	0	0
130	NAPM com price	ISM Manufacturing: Prices Paid Index©	1959:Q1-2014:Q4	1	0	1
131	Real_Price:NatGas	PPI: Natural Gas Defl by PCE(LFE)	1967:Q1-2014:Q4	5	0	1
	(7) Productivity and Earnings					
132	Real_AHE:PrivInd	Average Hourly Earnings: Total Private Industries Defl by PCE(LFE)	1964:Q1-2014:Q4	5	0	0
133	Real_AHE:Const	Average Hourly Earnings: Construction Defl by PCE(LFE)	1959:Q1-2014:Q4	5	0	0
134	Real_AHE:MFG	Average Hourly Earnings: Manufacturing Defl by PCE(LFE)	1959:Q1-2014:Q4	5	0	0
135	CPH:NFB	Nonfarm Business Sector: Real Compensation Per Hour	1959:Q1-2014:Q4	5	0	1
136	CPH:Bus	Business Sector: Real Compensation Per Hour	1959:Q1-2014:Q4	5	0	1
137	OPH:nfb	Nonfarm Business Sector: Output Per Hour of All Persons	1959:Q1-2014:Q4	5	0	1
138	OPH:Bus	Business Sector: Output Per Hour of All Persons	1959:Q1-2014:Q4	5	0	0
139	ULC:Bus	Business Sector: Unit Labor Cost	1959:Q1-2014:Q4	5	0	0
140	ULC:NFB	Nonfarm Business Sector: Unit Labor Cost	1959:Q1-2014:Q4	5	0	1
141	UNLPay:nfb	Nonfarm Business Sector: Unit Nonlabor Payments	1959:Q1-2014:Q4	5	0	1
	(8) Interest Rates					
142	FedFunds	Effective Federal Funds Rate	1959:Q1-2014:Q4	2	0	1
143	TB-3Mth	3-Month Treasury Bill: Secondary Market Rate	1959:Q1-2014:Q4	2	0	1
144	TM-6MTH	6-Month Treasury Bill: Secondary Market Rate	1959:Q1-2014:Q4	2	0	0
145	EuroDol3M	3-Month Eurodollar Deposit Rate (London)	1971:Q1-2014:Q4	2	0	0
146	TB-1YR	1-Year Treasury Constant Maturity Rate	1959:Q1-2014:Q4	2	0	0
147	TB-10YR	10-Year Treasury Constant Maturity Rate	1959:Q1-2014:Q4	2	0	0
148	Mort-30Yr	30-Year Conventional Mortgage Rate	1971:Q2-2014:Q4	2	0	0
149	AAA Bond	Moody's Seasoned Aaa Corporate Bond Yield	1959:Q1-2014:Q4	2	0	0
150	BAA Bond	Moody's Seasoned Baa Corporate Bond Yield	1959:Q1-2014:Q4	2	0	0
151	BAA_GS10	BAA-GS10 Spread	1959:Q1-2014:Q4	1	0	1
152	MRTG_GS10	Mortg-GS10 Spread	1971:Q2-2014:Q4	1	0	1
153	tb6m_tb3m	tb6m-tb3m	1959:Q1-2014:Q4	1	0	1
154	GS1_tb3m	GS1_Tb3m	1959:Q1-2014:Q4	1	0	1
155	GS10_tb3m	GS10_Tb3m	1959:Q1-2014:Q4	1	0	1
156	CP_Tbill Spread	CP3FM-TB3MS	1959:Q1-2014:Q4	1	0	1
157	Ted_spr	MED3-TB3MS (Version of TED Spread)	1971:Q1-2014:Q4	1	0	1

158	gz_spread	Gilchrist-Zakrajsek Spread (Unadjusted)	1973:Q1-2012:Q4	1	0	0
159	gz_ebp	Gilchrist-Zakrajsek Excess Bond Premium	1973:Q1-2012:Q4	1	0	1
	(9) Money and Credit					
160	Real_mbase	St. Louis Adjusted Monetary Base; Bil. of $; M; SA; Defl by PCE(LFE)	1959:Q1-2014:Q4	5	0	0
161	Real_InsMMF	Institutional Money Funds Defl by PCE(LFE)	1980:Q1-2014:Q4	5	0	0
162	Real_m1	M1 Money Stock Defl by PCE(LFE)	1959:Q1-2014:Q4	5	0	0
163	Real_m2	M2SL Defl by PCE(LFE)	1959:Q1-2014:Q4	5	0	0
164	Real_mzm	MZM Money Stock Defl by PCE(LFE)	1959:Q1-2014:Q4	5	0	0
165	Real_C&Lloand	Commercial and Industrial Loans at All Commercial Banks Defl by PCE(LFE)	1959:Q1-2014:Q4	5	0	1
166	Real_ConsLoans	Consumer (Individual) Loans at All Commercial Banks/ Outlier Code because of change in data in April 2010. See FRB H8 Release Defl by PCE(LFE)	1959:Q1-2014:Q4	5	1	1
167	Real_NonRevCredit	Total Nonrevolving Credit Outstanding Defl by PCE(LFE)	1959:Q1-2014:Q4	5	0	1
168	Real_LoansRealEst	Real Estate Loans at All Commercial Banks Defl by PCE(LFE)	1959:Q1-2014:Q4	5	0	1
169	Real_RevolvCredit	Total Revolving Credit Outstanding Defl by PCE(LFE)	1968:Q1-2014:Q4	5	1	1
170	Real_ConsuCred	Total Consumer Credit Outstanding Defl by PCE(LFE)	1959:Q1-2014:Q4	5	0	0
171	FRBSLO_Consumers	FRB Senior Loans Officer Opions. Net Percentage of Domestic Respondents Reporting Increased Willingness to Make Consumer Installment Loans (Fred from 1982:Q2 on Earlier is DB series)	1970:Q1-2014:Q4	1	0	1
	(10) International Variables					
172	Ex rate: major	FRB Nominal Major Currencies Dollar Index (Linked to EXRUS in 1973:1)	1959:Q1-2014:Q4	5	0	1
173	Ex rate: Euro	U.S. / Euro Foreign Exchange Rate	1999:Q1-2014:Q4	5	0	1
174	Ex rate: Switz	Foreign exchange rate: Switzerland (Swiss franc per U.S.$) Fred 1971. EXRSW previous	1971:Q1-2014:Q4	5	0	1
175	Ex rate: Japan	Foreign exchange rate: Japan (yen per U.S.$) Fred 1971- EXRJAN previous	1971:Q1-2014:Q4	5	0	1
176	Ex rate: UK	Foreign exchange rate: United Kingdom (cents per pound) Fred 1971-> EXRUK Previous	1971:Q1-2014:Q4	5	0	1
177	EX rate: Canada	Foreign exchange rate: Canada (Canadian $ per U.S.$) Fred 1971 -> EXRCAN previous	1971:Q1-2014:Q4	5	0	1
178	OECD GDP	OECD: Gross Domestic Product by Expenditure in Constant Prices: Total Gross; Growth Rate (Quartely); Fred Series NAEXKP01O1Q657S	1961:Q2-2013:Q4	1	0	1
179	IP Europe	OECD: Total Ind. Prod (excl Construction) Europe Growth Rate (Quarterly); Fred Series PRINTO01OEQ657S	1960:Q2-2013:Q4	1	0	1
180	Global Ec Activity	Kilian's estimate of glaobal economic activity in industrial commodity markets (Kilian website)	1968:Q1-2014:Q4	1	0	1
	(11) **Asset Prices, Wealth, and Household Balance Sheets**					
181	S&P 500	S&P's Common Stock Price Index: Composite (1941-43=10)	1959:Q1-2014:Q4	5	0	1
182	Real_HHW:TA	Households and nonprofit organizations; total assets (FoF) Seasonally Adjusted (RATS X11) Defl by PCE(LFE)	1959:Q1-2014:Q3	5	0	0
183	Real_HHW:TL	Households and nonprofit organizations; total liabilities Seasonally Adjusted (RATS X11) Defl by PCE(LFE)	1959:Q1-2014:Q3	5	0	1
184	liab_PDI	Liabilities Relative to Person Disp Income	1959:Q1-2014:Q3	5	0	0
185	Real_HHW:W	Households and nonprofit organizations; net worth (FoF) Seasonally Adjusted (RATS X11) Defl by PCE(LFE)	1959:Q1-2014:Q3	5	0	1
186	W_PDI	Networth Relative to Personal Disp Income	1959:Q1-2014:Q3	1	0	0
187	Real_HHW:TFA	Households and nonprofit organizations; total financial assets Seasonally Adjusted (RATS X11) Defl by PCE(LFE)	1959:Q1-2014:Q3	5	0	0
188	Real_HHW:TA_RE	TotalAssets minus Real Estate Assets Defl by PCE(LFE)	1959:Q1-2014:Q3	5	0	1

189	Real_HHW:TNFA	Households and nonprofit organizations; total nonfinancial assets (FoF) Seasonally Adjusted (RATS X11) Defl by PCE(LFE)	1959:Q1-2014:Q3	5	0	0
190	Real_HHW:RE	Households and nonprofit organizations; real estate at market value Seasonally Adjusted (RATS X11) Defl by PCE(LFE)	1959:Q1-2014:Q3	5	0	1
191	DJIA	Common Stock Prices: Dow Jones Industrial Average	1959:Q1-2014:Q4	5	0	1
192	VXO	VXO (Linked by N. Bloom) .. Average daily VIX from 2009 ->	1962:Q3-2014:Q4	1	0	1
193	Real_Hprice:OFHEO	House Price Index for the United States Defl by PCE(LFE)	1975:Q1-2014:Q4	5	0	1
194	Real_CS_10	Case-Shiller 10 City Average Defl by PCE(LFE)	1987:Q1-2014:Q4	5	0	1
195	Real_CS_20	Case-Shiller 20 City Average Defl by PCE(LFE)	2000:Q1-2014:Q4	5	0	1
	(12) Other					
196	Cons. Expectations	Consumer expectations NSA (Copyright University of Michigan)	1959:Q1-2014:Q4	1	0	1
197	PoilcyUncertainty	Baker Bloom Davis Policy Uncertainty Index	1985:Q1-2014:Q4	2	0	1
	(13) Oil Market Variables					
198	World Oil Production	World Oil Production.1994:Q1 on from EIA (Crude Oil including Lease Condensate); Data prior to 1994 from From Baumeister and Peerlman (2013)	1959:Q1-2014:Q3	5	0	0
199	World Oil Production	World Oil Production.1994:Q1 on from EIA (Crude Oil including Lease Condensate); Data prior to 1994 from From Baumeister and Peerlman (2013); Seasonally adjusted using RATS X11 (note seasonality before 1970)	1959:Q1-2014:Q3	5	0	1
200	IP: Energy Prds	IP: Consumer Energy Products	1959:Q1-2014:Q4	5	0	1
201	Petroleum Stocks	U.S. Ending Stocks excluding SPR of Crude Oil and Petroleum Products (Thousand Barrels); SA using X11 in RATS	1959:Q1-2014:Q4	5	0	1
202	Real_Price:Oil	PPI: Crude Petroleum Defl by PCE(LFE)	1959:Q1-2014:Q4	5	0	1
203	Real_Crudeoil Price	Crude Oil: West Texas Intermediate (WTI) - Cushing Oklahoma Defl by PCE(LFE)	1986:Q1-2014:Q4	5	0	1
204	Real_CrudeOil	Crude Oil Prices: Brent - Europe Defl by PCE(LFE) Def	1987:Q3-2014:Q4	5	0	1
205	Real_Price Gasoline	Conventional Gasoline Prices: New York Harbor Regular Defl by PCE(LFE)	1986:Q3-2014:Q4	5	0	1
206	Real_Refiners Acq. Cost (Imports)	U.S. Crude Oil Imported Acquisition Cost by Refiners (Dollars per Barrel) Defl by PCE(LFE)	1974:Q1-2014:Q4	5	0	1
207	Real_CPI Gasoline	CPI Gasoline (NSA) BLS: CUUR0000SETB01 Defl by PCE(LFE)	1959:Q1-2014:Q4	5	0	1

Dealing with large datasets

(1) Outliers

(2) Non-stationarities and 'trends'

Usual transformations (logs, differences, spreads, etc.)

Low-frequency 'demeaning'

(3) Aggregates (139 vs. 207)

(4) Estimate factors using standarized data ('weights' in weighted least squares). [$\min_{\{F_t\},\{\lambda_i\}} \sum_{i,t} (X_{it} - \lambda_i ' F_t)^2$]

Low-frequency 'demeaning' weights and sprectral gain

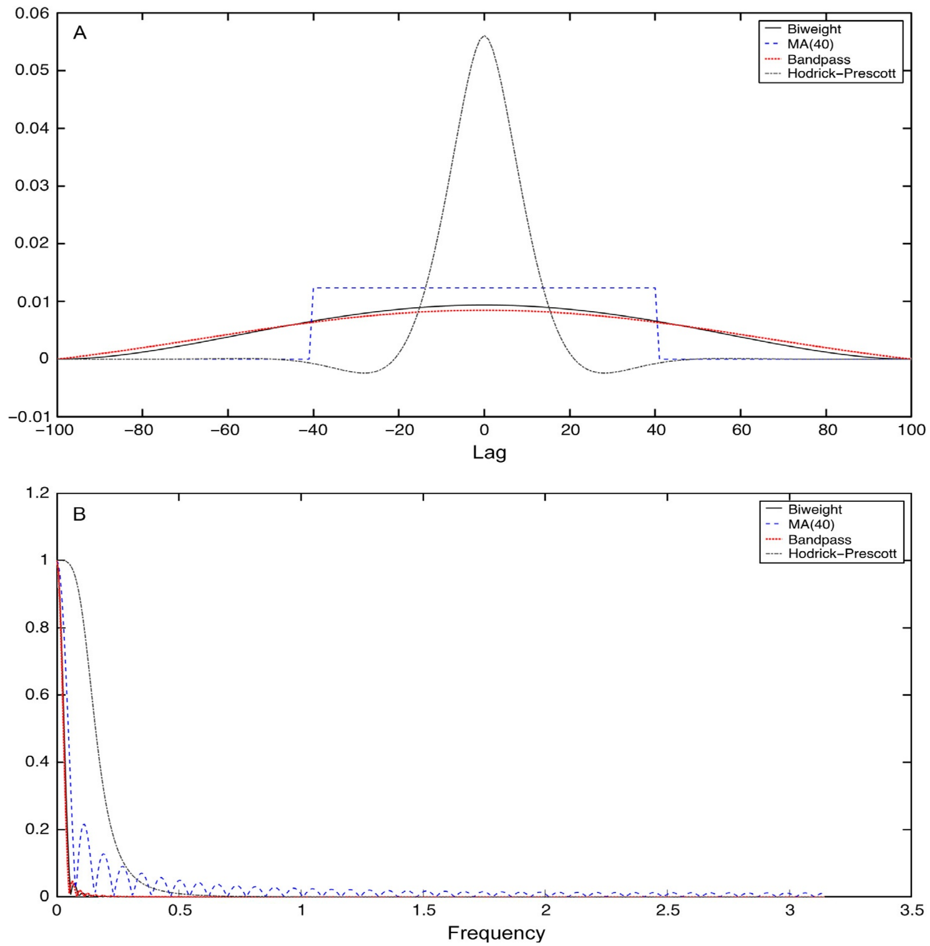


Fig. 2 Lag weights and spectral gain of trend filters. *Notes:* The biweight filter uses a bandwidth (truncation parameter) of 100 quarters. The bandpass filter is a 200-quarter low-pass filter truncated after 100 leads and lags (Baxter and King, 1999). The moving average is equal-weighted with 40 leads and lags. The Hodrick and Prescott (1997) filter uses 1600 as its tuning parameter.

How Many Factors?

(1) Scree plot

(2) Information criteria

(3) Others

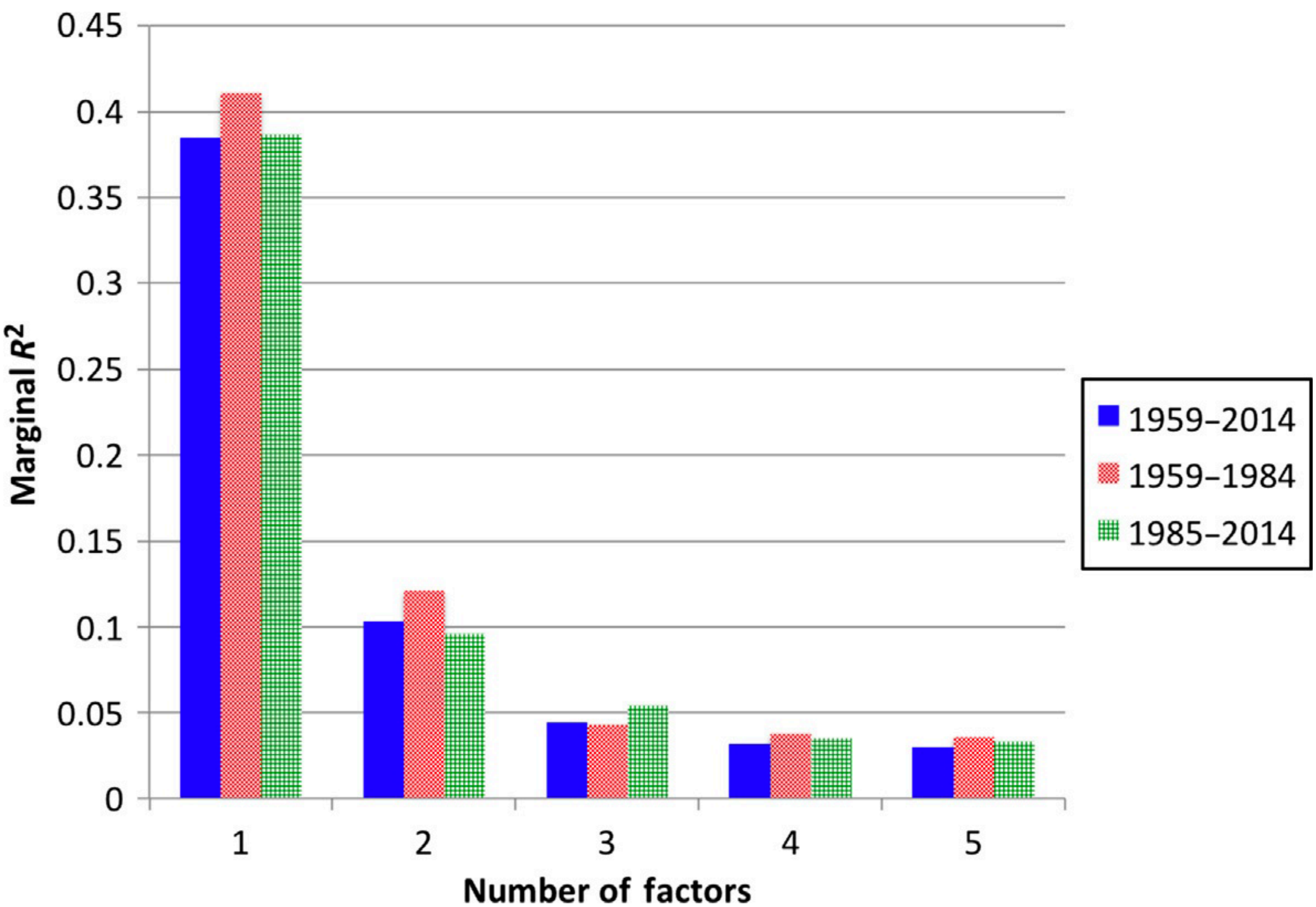
Least squares objective function for r factors:

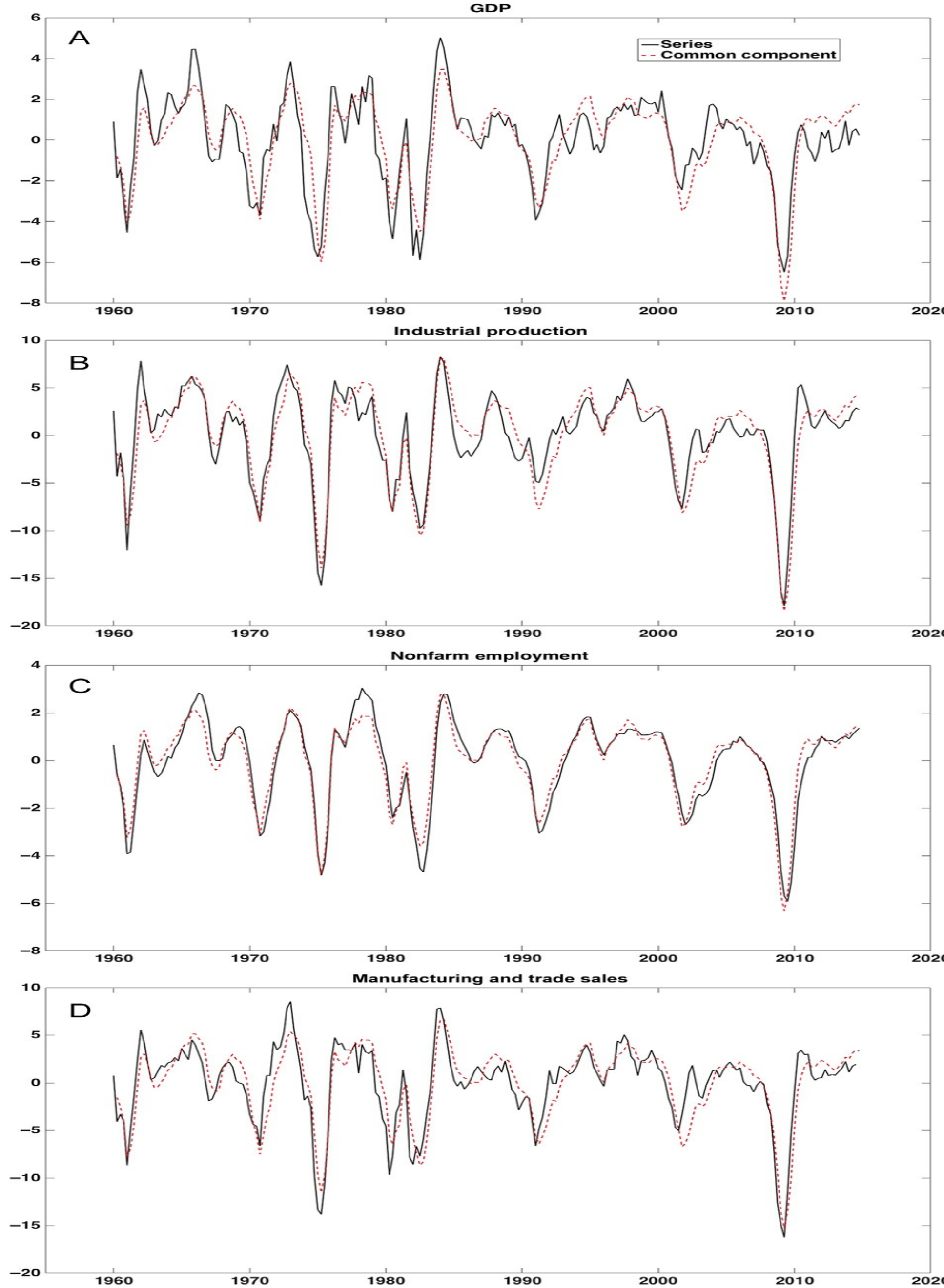
$$SSR(r) = \min_{\{F_t\},\{\lambda_i\}} \sum_{i,t} (X_{it} - \lambda_i' F_t)^2$$

where F_t and λ_i are $r \times 1$ vectors.

Scree plot: Marginal (trace) R^2 for factor k:

Scree plot for 58 real variables





etrended four-quarter growth rates of US GDP, industrial production, nonfarm

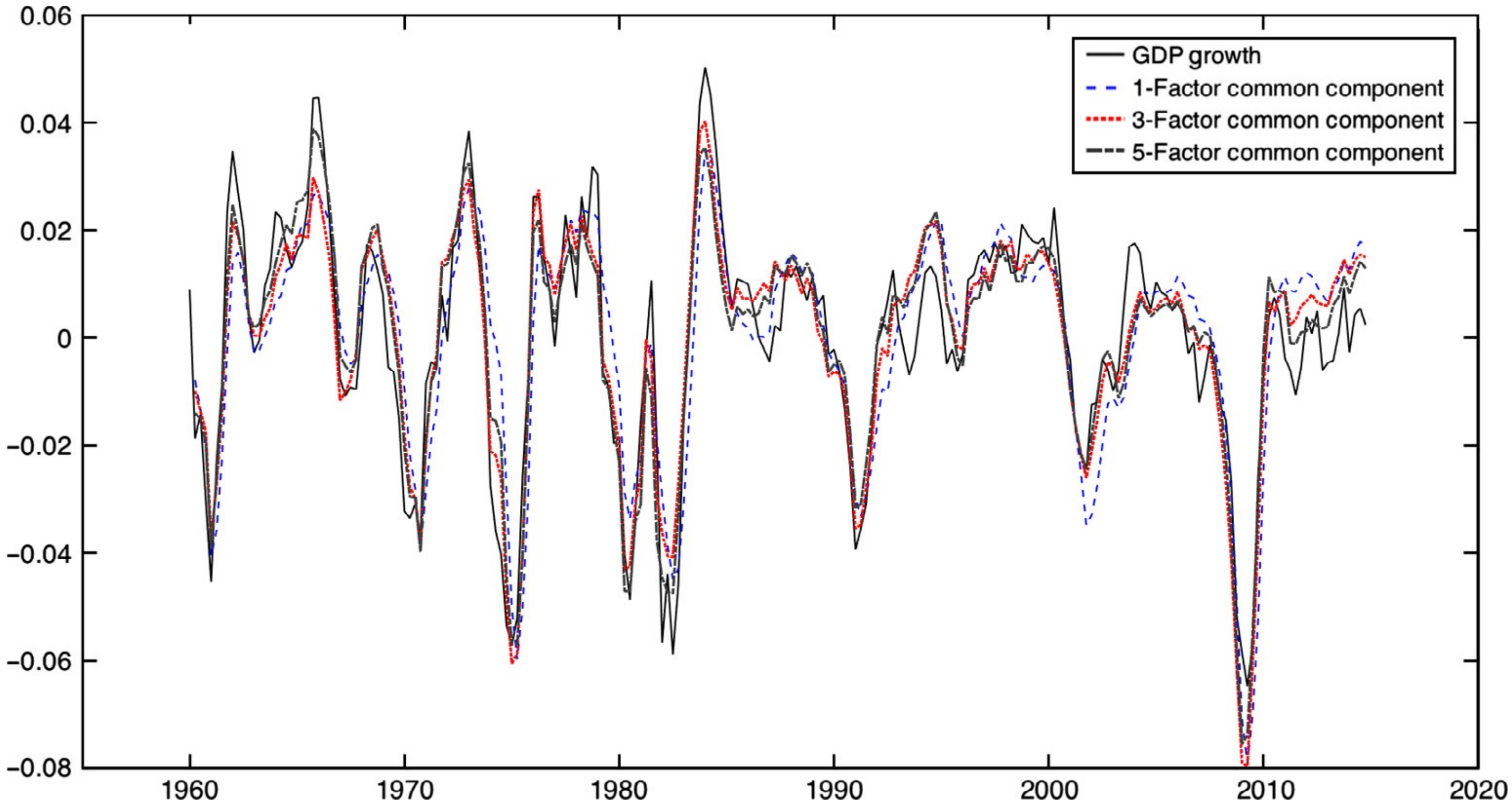
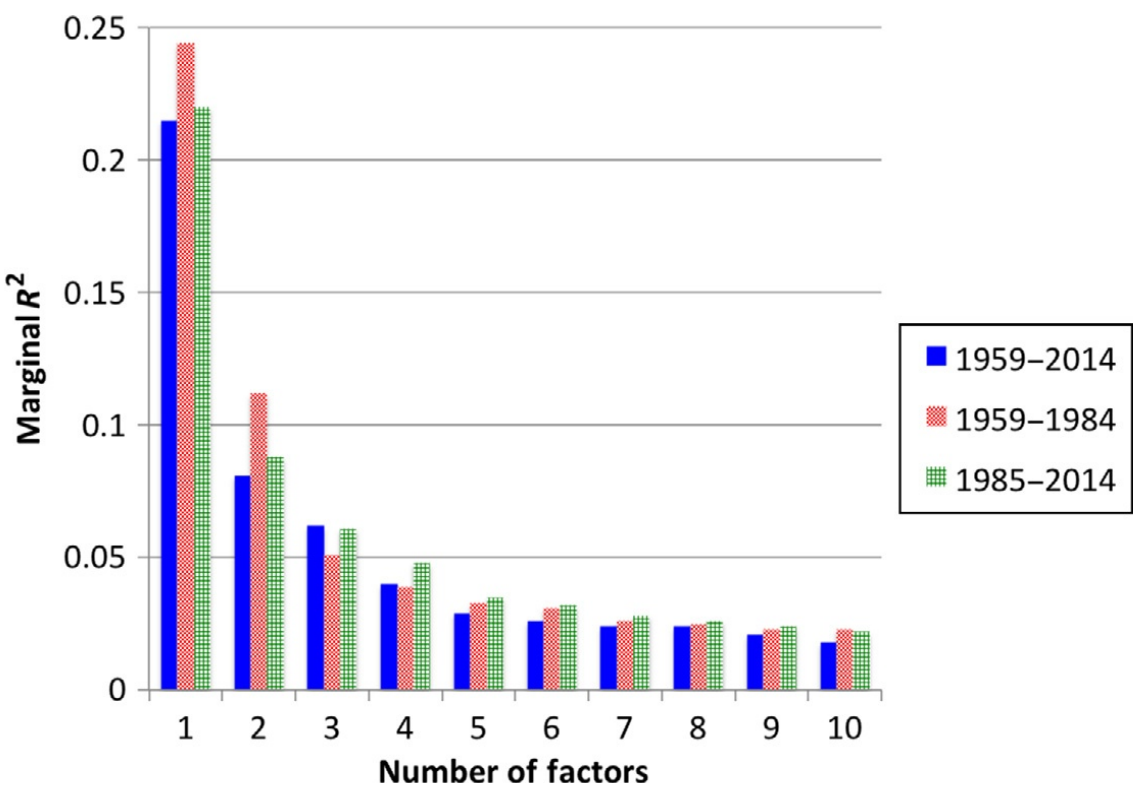


Fig. 4 Four-quarter GDP growth (*black*) and its common component based on 1, 3, and 5 static factors: real activity dataset.

Scree plot – Full data set (139 variables)

Factor Models and Structural Vector Autoregressio



Information criteria: Bai and Ng

IC(r) = ln(SSR(r)) + r × ***g*(sample size)**

Sample size: n and T

$$\text{BNIC}(r) = \ln\left(\text{SSR}(r)\right) + r \times \left(\frac{n+T}{nT}\right)\ln\left(\min(n,T)\right)$$

Note: when $n = T$ this is BNIC(r) = ln(SSR(r)) + r×**2ln(*T*)/*T***.

Table 2 Statistics for estimating the number of static factors

(A) Real activity dataset ($N = 58$ disaggregates used for estimating factors)

Number of static factors	Trace R^2	Marginal trace R^2	BN-IC_{p2}	AH-ER
1	0.385	0.385	−0.398	**3.739**
2	0.489	0.103	−0.493	2.338
3	0.533	0.044	**−0.494**	1.384
4	0.565	0.032	−0.475	1.059
5	0.595	0.030	−0.458	1.082

(B) Full dataset ($N = 139$ disaggregates used for estimating factors)

Number of static factors	Trace R^2	Marginal trace R^2	BN-IC_{p2}	AH-ER
1	0.215	0.215	−0.183	**2.662**
2	0.296	0.081	−0.233	1.313
3	0.358	0.062	−0.266	1.540
4	0.398	0.040	**−0.271**	1.368
5	0.427	0.029	−0.262	1.127
6	0.453	0.026	−0.249	1.064
7	0.478	0.024	−0.235	1.035
8	0.501	0.024	−0.223	1.151
9	0.522	0.021	−0.205	1.123
10	0.540	0.018	−0.185	1.057

What about many more factors?
(Full 138-variable dataset)

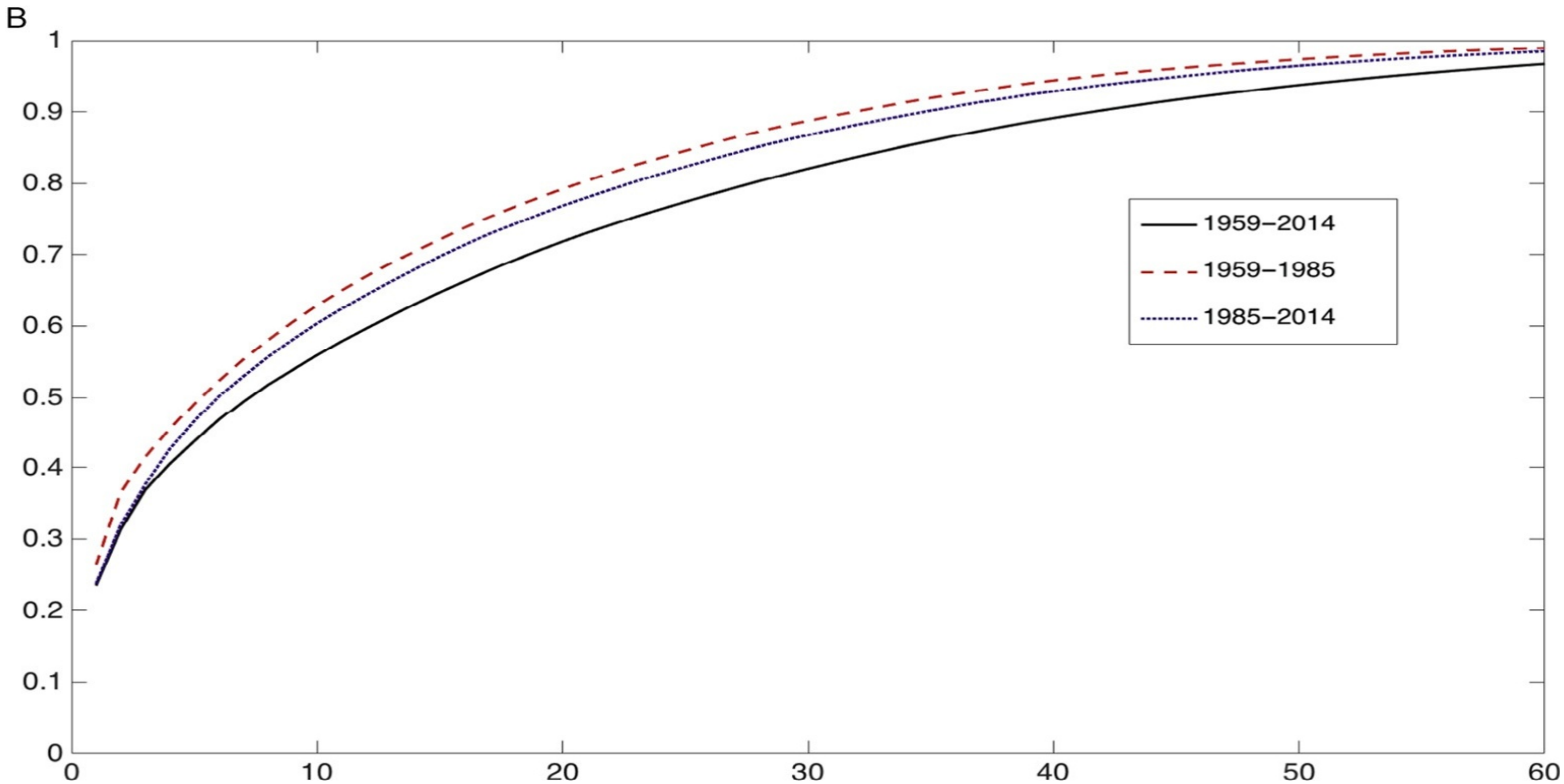


Fig. 6 (A) Scree plot for full dataset: full sample, pre-1984, and post-1984. (B) Cumulative R^2 as a

Is there useful information in additional factors? (For forecasting, maybe)

Structural **DFM or SFDM:** SVAR analysis, but now using DFM

SVAR problems that the DFM might solve:

(a) Many variables, thus invertibility is more plausible.

(b) Errors-in-variables, several indicators for same theoretical concept ('aggregate prices','oil prices', etc.)

(c) Framework for computing IRFs from structural shocks to many variables.

Can't I just do a VAR? .. No

Table 5 Approximating the eight-factor DFM by a eight-variable VAR

	Canonical correlation							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
(A) Innovations								
VAR-A	0.76	0.64	0.6	0.49				
VAR-B	0.83	0.67	0.59	0.56	0.37	0.33	0.18	0.01
VAR-C	0.86	0.81	0.78	0.76	0.73	0.58	0.43	0.35
VAR-O	0.83	0.80	0.69	0.56	0.50	0.26	0.16	0.02
(B) Variables and factors								
VAR-A	0.97	0.85	0.79	0.57				
VAR-B	0.97	0.95	0.89	0.83	0.61	0.43	0.26	0.10
VAR-C	0.98	0.93	0.90	0.87	0.79	0.78	0.57	0.41
VAR-O	0.98	0.96	0.88	0.84	0.72	0.39	0.18	0.02

Notes: All VARs contain four lags of all variables. The canonical correlations in panel A are between the VAR residuals and the residuals of a VAR estimated for the eight static factors.
VAR-A was chosen to be typical of four-variable VARs seen in empirical applications. Variables: GDP, total employment, PCE inflation, and Fed funds rate.
VAR-B was chosen to be typical of eight-variable VARs seen in empirical applications. Variables: GDP, total employment, PCE inflation, Fed funds, ISM manufacturing index, real oil prices (PPI-oil), corporate paper-90-day treasury spread, and 10 year–3 month treasury spread.
VAR-C variables were chosen by stepwise maximization of the canonical correlations between the VAR innovations and the static factor innovations. Variables: industrial commodities PPI, stock returns (SP500), unit labor cost (NFB), exchange rates, industrial production, Fed funds, labor compensation per hour (business), and total employment (private).
VAR-O variables: real oil prices (PPI-oil), global oil production, global commodity shipment index, GDP, total employment (private), PCE inflation, Fed funds rate, and trade-weighted US exchange rate index.
Entries are canonical correlations between (A) factor innovations and VAR residuals and (B) factors and observable variables.

The SDFM:

$$\underset{t}{\overset{n\times 1}{X}} = \overset{n\times r}{\Lambda} \underset{t}{\overset{r\times 1}{F}} + \underset{t}{\overset{n\times 1}{e}}$$

$$\overset{r\times r}{\Phi(L)} \underset{t}{\overset{r\times 1}{F}} = \overset{r\times q}{G} \underset{t}{\overset{q\times 1}{\eta}}$$

where $\Phi(L) = I - \Phi_1 L - \ldots - \Phi_p L^p$,

$$\underset{t}{\overset{q\times 1}{\eta}} = \overset{q\times q}{H} \underset{t}{\overset{q\times 1}{\varepsilon}}$$

$$X_t = \Lambda\Phi(L)^{-1} GH\varepsilon_t + e_t$$

IRFs: $\Lambda\Phi(L)^{-1}GH$

IRF from ε_{1t}: $\Lambda\Phi(L)^{-1}GH_1$

Three Normalizations

1. $\Lambda F_t = \Lambda PP^{-1}F_t$ for any matrix P. Set P rows of Λ equal to rows of identity matrix. Rearranging the order of the *Xs* this yields

$$\begin{pmatrix} X_{1:r} \\ X_{r+1:n} \end{pmatrix}_t = \begin{pmatrix} I_r \\ \Lambda_{r+1:n} \end{pmatrix} F_t + e_t$$

This 'names' the first factor as the X_1 factor, the second factor as the X_2 factor and so forth. Example: $X_{1,t}$ is the logarithm of oil prices, then $F_{1,t}$ is called the oil price factor.

2. G = I (if $q = r$) or $G_{1:q} = I_q$ if $q < r$. Recall

$$X_t = \lambda(L)f_t + e_t \text{ and } \phi(L)f_t = \eta_t$$

$$X_t = \begin{pmatrix} \lambda_0 & \lambda_1 & \cdots & \lambda_k \end{pmatrix} \begin{pmatrix} f_t \\ f_{t-1} \\ \vdots \\ f_{t-k} \end{pmatrix} + e_t$$

$$\begin{pmatrix} f_t \\ f_{t-1} \\ \vdots \\ f_{t-k} \end{pmatrix} = \begin{bmatrix} \phi_1 & \phi_2 & \cdots & \phi_{k+1} \\ 1 & 0 & \cdots & 0 \\ & \ddots & \ddots & \\ & & 1 & 0 \end{bmatrix} \begin{pmatrix} f_{t-1} \\ f_{t-2} \\ \vdots \\ f_{t-k-1} \end{pmatrix} + \begin{pmatrix} I \\ 0 \\ \vdots \\ 0 \end{pmatrix} \eta_t$$

where f_t and η_t are $q \times 1$.

3. The diagonal elements of H are unity. That is, ε_{1t} has a unit effect of $F_{1,t}$ and so forth. Same as in SVAR.

Putting these together:

$X_{1:q,t} = \mathrm{H}\varepsilon_t + \text{lags of } \varepsilon_t + e_t$

(Same 'unit-effect' normalization used in SVAR (JS), but only applied to the first q elements of X_t).

$F_{1:q,t} = \mathrm{H}\varepsilon_t + \text{lags of } \varepsilon_t$

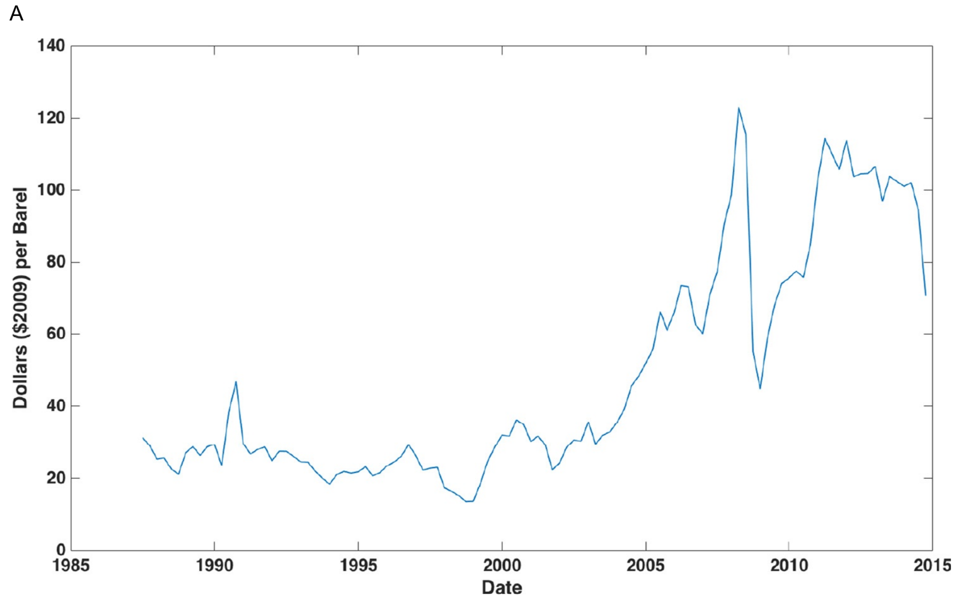
etc.

This means that everything in SVARs carry over here.

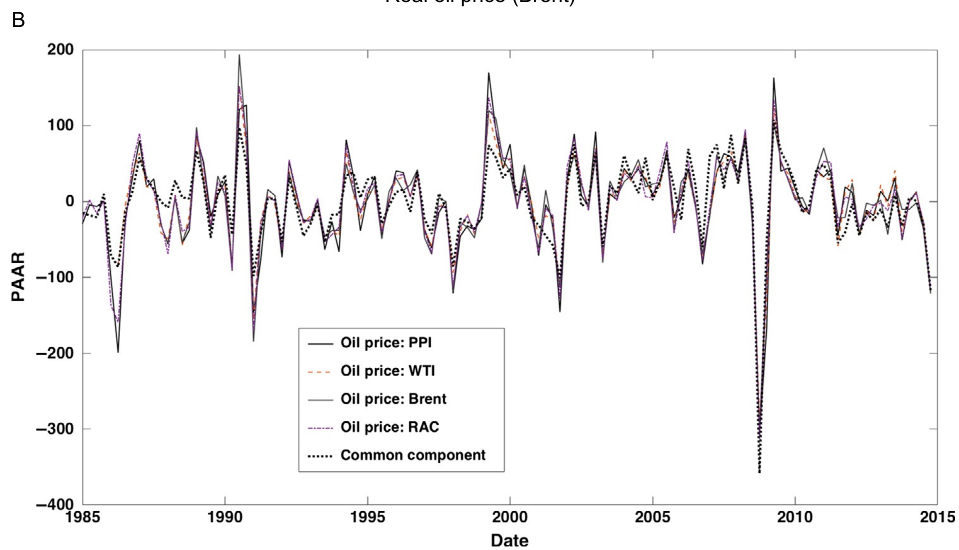
Additional flexibility in SDFM

(1) Measurement error allowed: With normalization, F follows SVAR, and $X = \Lambda F + e$.

(2) Multiple measurements: Example Oil prices



Real oil price (Brent)



Quarterly percent change in real oil price: four oil price series and the common component

Fig. 7 Real oil price (2009 dollars) and its quarterly percent change.

$$\begin{bmatrix} p_t^{PPI-Oil} \\ p_t^{Brent} \\ p_t^{WTI} \\ p_t^{RAC} \\ X_{5:n,t} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ & & \Lambda_{5:n} & \end{bmatrix} \begin{bmatrix} F_t^{oil} \\ F_{2:r,t} \end{bmatrix} + e_t$$

(3) "Factor Augmented" VAR) (FAVAR) (Bernanke, Boivin, Eliasz (2005))

Easily implemented in this framework:

$$\begin{pmatrix} Y_t \\ X_t \end{pmatrix} = \begin{pmatrix} 1 & 0_{1\times r} \\ \multicolumn{2}{c}{\Lambda} \end{pmatrix} \begin{pmatrix} \tilde{F}_t \\ F_t \end{pmatrix} + \begin{pmatrix} 0 \\ e_t \end{pmatrix}$$

$$F_t^+ = \Phi(L) F_{t-1}^+ + G\eta_t$$

where

$$F_t^+ = \begin{pmatrix} \tilde{F}_t \\ F_t \end{pmatrix},$$

$$\eta_t = \mathrm{H}\varepsilon_t.$$

Example: Macroeconomic Effects of Oil Supply Shocks

2 Identifications:

(1) Oil Price exogenous

$$\eta_t = \begin{pmatrix} 1 & 0 \\ H_{\bullet 1} & H_{\bullet\bullet} \end{pmatrix} \begin{pmatrix} \varepsilon_t^{oil} \\ \tilde{\eta}_{\bullet t} \end{pmatrix}$$

$$\begin{bmatrix} p_t^{PPI-Oil} \\ p_t^{Brent} \\ p_t^{WTI} \\ p_t^{RAC} \\ X_{5:n,t} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ \lambda_{21} & \lambda_{22} & \lambda_{23} & \cdots & \lambda_{28} \\ \lambda_{31} & & & \cdots & \lambda_{38} \\ \lambda_{41} & & & \cdots & \lambda_{48} \\ \Lambda_{5:n} & & & & \end{bmatrix} \begin{bmatrix} F_t^{oilprice} \\ F_{2,t} \\ F_{3,t} \\ \vdots \\ F_{8,t} \end{bmatrix} + \begin{bmatrix} e_t^{PPI-oil} \\ e_t^{Brent} \\ e_t^{WTI} \\ e_t^{RAC} \\ e_t^{X} \end{bmatrix}$$

SVAR, FAVAR and SDFM versions

(2) Killian (2009) Identification

$$\begin{bmatrix} OilSupply \\ GlobalActivity_t \\ p_t^{PPI-Oil} \\ p_t^{Brent} \\ p_t^{WTI} \\ p_t^{RAC} \\ X_{7:n,t} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ \Lambda_{8:n} & & & & & \end{bmatrix} \begin{bmatrix} F_t^{OilSupply} \\ F_t^{Globalactivity} \\ F_t^{oilprice} \\ F_{4:r,t} \end{bmatrix} + e_t$$

$$\Phi(L)F_t = \begin{pmatrix} 1 & 0 & 0 & 0 \\ H_{12} & 1 & 0 & 0 \\ H_{13} & H_{23} & 1 & 0 \\ H_{1\bullet} & H_{2\bullet} & H_{3\bullet} & H_{\bullet\bullet} \end{pmatrix} \begin{pmatrix} \varepsilon_t^{OS} \\ \varepsilon_t^{GD} \\ \varepsilon_t^{OD} \\ \tilde{\eta}_{\bullet t} \end{pmatrix}$$

Empirical Results … see paper

Lecture 5: Part 2

Prediction with large datasets

Population Problem: Y is a scalar, X is an $n \times 1$ vector. Predict Y given X.

Population Solution: Use $\mathbf{E}(Y \mid X)$

Sample Problem: Given T in-sample observations on Y and X, how should you predict an out-of-sample value for Y.

Simplification: Suppose $\mathbf{E}(Y \mid X)$ is linear in X. How should you estimate the linear regression coefficients for the purposes of prediction?

Sample Solution:

(a) n/T is small … use OLS.

(b) n/T not small … do NOT use OLS

Forecasting when n is large

Imposing more structure:

(1) X and Y are related through a 'few' common factors. (Use version of DFM).

(2) Regression coefficients are 'small'. (Use shrinkage)

(3) Many regressions coefficients are zero. (Impose 'sparsity')

(Notation will use 1-period-ahead forecasts).

DFM Forecasting

Forecasting setup: $Y_{t+1} = F_t{}'\alpha + \varepsilon_{t+1}$

$$X_t = \Lambda F_t + e_t$$

$$\Phi(\mathrm{L})F_t = G\eta_t$$

Implication: $\mathbf{E}(Y_{t+1} \mid X^t) = \mathbf{E}(F_t \mid X^t)'\alpha$

Use X to estimate F (say using $\hat{F}^{PC}$ or Kalman Filter).

When n is large, $\hat{F}^{PC}$ is very close to F. Thus, use $\hat{F}^{PC}$ as if they were true values of F.

Result (Stock-Watson (2002)): $\hat{y}_{T+1}\left(\hat{F}^{PC}\right) - \hat{y}_{T+1}\left(F\right) \xrightarrow{ms} 0$

(Addition references in DFM section above).

Regression coefficients are 'small'. (Use shrinkage)

Linear prediction problem: $Y_{t+1} = X_t'\beta + \varepsilon_{t+1}$

Simpler problem: Orthonormal regressors.

Transform regressors as $p_t = HX_t$ where H is chosen so that

$$T^{-1}\sum_{t=1}^{T} p_t p_t' = T^{-1}P'P = I_n.$$ (Note: This requires $n \le T$)

Regression equation: $Y_{t+1} = p_t'\alpha + \varepsilon_{t+1}$

OLS Estimator: $\hat{\alpha} = (P'P)^{-1}P'Y = T^{-1}P'Y$

so that $\hat{\alpha}_i = T^{-1}\sum_{t=1}^{T} p_{it}Y_{t+1}$

Note: Suppose p_t are strictly exogenous and $\varepsilon_t \sim iidN(0,\sigma^2)$. (This will motivate the estimators .. more discussion below).

In this simple setting:

(1) $\hat{\alpha}$ are sufficient for α.

(2) $\left(\hat{\alpha}-\alpha\right) \sim N\left(0, T^{-1}\sigma^2 I_n\right)$

(3) MSFE: $E\left(\sum_{i=1}^{n} p_{iT}(\alpha_i - \tilde{\alpha}_i)\right)^2 + \sigma^2 \approx \frac{n}{T} MSE(\tilde{\alpha}) + \sigma^2$

So we can think about analyzing n-independent normal random variables, $\hat{\alpha}_i$, to construct estimators $\tilde{\alpha}(\hat{\alpha}_i)$ that have small MSE – shrinkage can help achieve this.

Shrinkage: Basic idea

Consider two estimators: (1) $\hat{\alpha}_i \sim \mathrm{N}(\alpha_i\,, T^{-1}\sigma^2)$

(2) $\tilde{\alpha}_i = ½\, \hat{\alpha}_i$

$\mathrm{MSE}(\hat{\alpha}_i) = T^{-1}\sigma^2$

$\mathrm{MSE}(\hat{\alpha}_i) = 0.25 \times (T^{-1}\sigma^2 + \alpha_i^2\,)$

$$\mathrm{MSFE}(\hat{\alpha}) = \frac{n}{T}\sigma^2 + \sigma^2$$

$$\mathrm{MSFE}(\tilde{\alpha}) = 0.25 \times \left[\frac{n}{T}\sigma^2 + \sum_{i=1}^{n}\alpha_i^2\right] + \sigma^2$$

How big is $\sum_{i=1}^{n}\alpha_i^2$?

What is optimal amount (and form) of shrinkage?

It depends on distribution of $\{\alpha_i\}$

- Bayesian methods use priors for the distribution

- Empirical Bayes methods estimate the distribution

Examples: L_2 – Shrinkage

Bayes: Suppose $\alpha_i \sim \text{iidN}(0, T^{-1}\omega^2)$

Then, with $\hat{\alpha}_i | \alpha_i \sim \text{N}(\alpha_i, T^{-1}\sigma^2)$,

$$\begin{bmatrix} \alpha_i \\ \hat{\alpha}_i \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, T^{-1} \begin{bmatrix} \omega^2 & \omega^2 \\ \omega^2 & \sigma^2 + \omega^2 \end{bmatrix} \right)$$

so that $\alpha_i | \hat{\alpha}_i \sim N\left(\dfrac{\omega^2}{\sigma^2+\omega^2}\hat{\alpha}_i, T^{-1}\dfrac{\omega^2\sigma^2}{\sigma^2+\omega^2} \right)$

MSE minimizing estimator conditional mean: $\tilde{\alpha}_i = \dfrac{\omega^2}{\omega^2+\sigma^2}\hat{\alpha}_i$

Empirical Bayes: Requires estimates of σ^2 and ω^2

If $T-n$ is large, then σ^2 can be accurately estimated.

If n is large, then ω^2 can be accurately estimated:

$$E(\hat{\alpha}_i^2) = T^{-1}(\sigma^2 + \omega^2), \text{ so } \hat{\omega}^2 = \frac{T}{n}\sum_{i=1}^{n}\hat{\alpha}_i^2 - \hat{\sigma}^2$$

(Extensions to more general distributions, etc. in this prediction framework – see Zhang (2005), and Knox, Stock and Watson (2004) and references therein.)

Alternative Formulation:

Write joint density of data and α as

$$\text{constant} \times \exp\left\{-0.5\left[\frac{1}{\sigma^2}\sum_{t=1}^{T}(y_{t+1}-p_t'\alpha)^2+\frac{1}{\omega^2}\sum_{i=1}^{n}\alpha_i^2\right]\right\}$$

Which is proportional to posterior for α. Because posterior is normal, mean = mode, so $\tilde{\alpha}$ can be found by maximizing posterior. Equivalently by solving:

$$\min_{\tilde{\alpha}}\sum_{t=1}^{T}(y_{t+1}-p_t'\tilde{\alpha})^2+\lambda\sum_{i=1}^{n}\tilde{\alpha}_i^2 \quad \text{with } \lambda = \sigma^2/\omega^2$$

This is called “Ridge Regression”

In the original X – regressor model, the ridge estimator of

$$\tilde{\beta}^{Ridge} = \left(X'X + \lambda I_n\right)^{-1}(X'Y)$$

and λ can be determined by prior-knowledge, or estimated (empirical Bayes, cross-validation, etc.)

(Note this estimator allows $n > T$.)

Relationship between Ridge and Principal components:

Let $F_t = RX_t$ where R is $n \times n$, $RR' = R'R = I_n$ and

$$\sum_{t=1}^{T} F_t F_t' = diag(\gamma_i^2) \text{ with } \gamma_1^2 \geq \gamma_2^2 \ldots \geq \gamma_n^2 \,.$$

Then $y_t = x_t'\beta + \varepsilon_t = F_t'\phi + \varepsilon_t$ with $\phi = R\beta$.

Let $\tilde{\phi}^{Ridge} = R\tilde{\beta}^{Ridge}$. Then algebra shows $\tilde{\phi}_i^{Ridge} = \hat{\phi}_i^{OLS}\left(\dfrac{\gamma_i^2}{\gamma_i^2 + \lambda}\right)$.

and $\tilde{\phi}_i^{PC} = \hat{\phi}_i^{OLS} \times 1(i \leq$ number of included PCs)

But included PCs are those with large γ_i^2.

Other shrinkage methods (There are many, of course, that depend on the assumed distribution of the regressions coefficients).

Many regression coefficients are zero. 'Sparse' modeling

Sparse models: Many/most values of β_i or α_i are zero.

Can be interpreted as shrinkage with lots of point mass at zero:

Approaches:

- Bayesian Model Averaging … (but can be computationally challenging … 2^n models): Hoeting, Madigan, Raftery, and Volinsky (1999))

- Hard thresholds (AIC/BIC) or smoothed out using "Bagging": (Breiman (1996), Bühlmann and Yu (2002); Inoue and Kilian (2008))

- L_1 penalization: Lasso ("Least Absolute Shrinkage and Selection Operator"): Tibshirani (1996)

Lasso: (With orthonormal regressors)

Ridge: $\min_{\tilde{\alpha}} \sum_{t=1}^{T} (y_{t+1} - p_t{}'\tilde{\alpha})^2 + \lambda \sum_{i=1}^{n} \tilde{\alpha}_i^2$

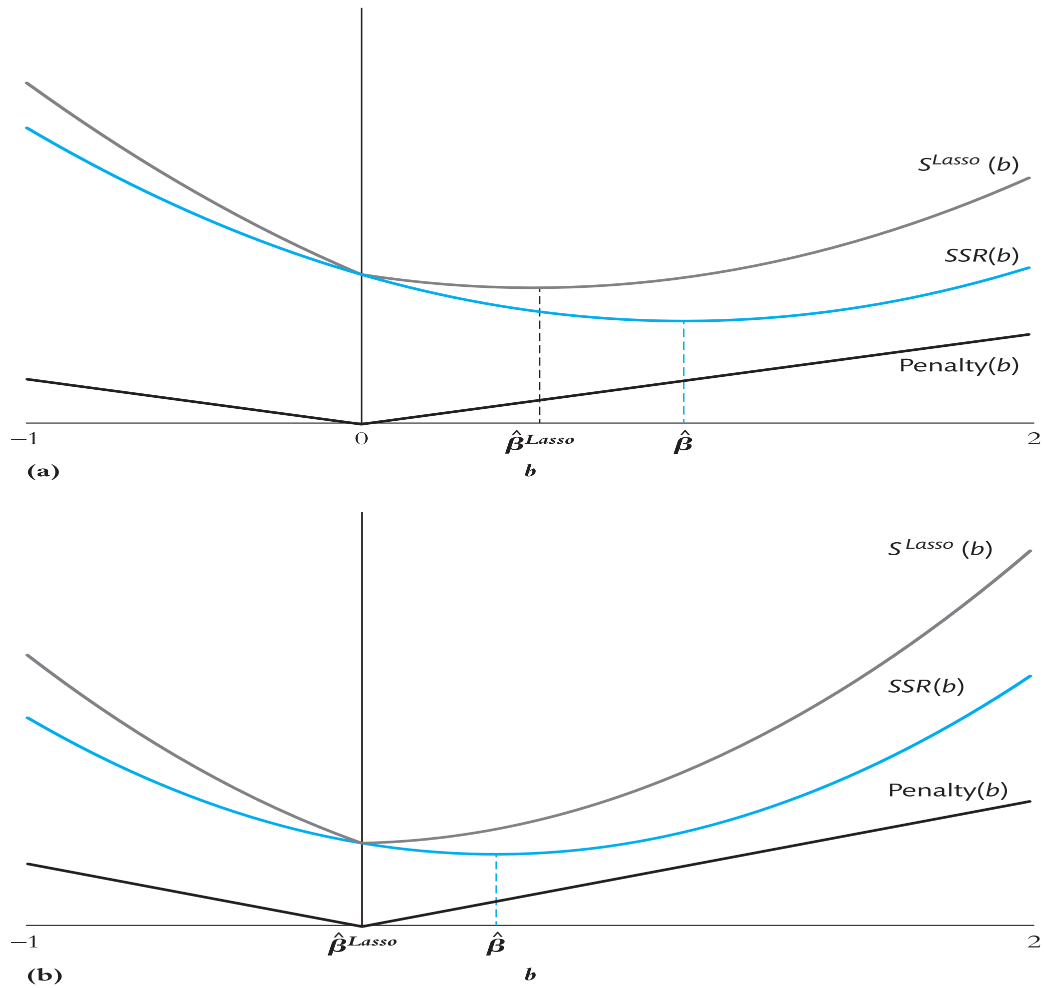
Lasso: $\min_{\tilde{\alpha}} \sum_{t=1}^{T} (y_{t+1} - p_t{}'\tilde{\alpha})^2 + \lambda \sum_{i=1}^{n} |\tilde{\alpha}_i|$

Equivalently: $\min_{\tilde{\alpha}} \sum_{i=1}^{n} (\hat{\alpha}_i - \tilde{\alpha}_i)^2 + \lambda \sum_{i=1}^{n} |\tilde{\alpha}_i|$

$$\min_{\tilde{\alpha}} \sum_{i=1}^{n} (\hat{\alpha}_i - \tilde{\alpha}_i)^2 + \lambda \sum_{i=1}^{n} |\tilde{\alpha}_i|$$

FIGURE 14.3 The Lasso Estimator Minimizes the Sum of Squared Residuals Plus a Penalty That Is Linear in the Absolute Value of *b*

For a single regressor, (a) when the OLS estimator is far from zero, the Lasso estimator shrinks it toward 0; (b) when the OLS estimator is close to 0, the Lasso estimator becomes exactly 0.



Notes:

- The solution yields $\text{sign}(\tilde{\alpha}_i) = \text{sign}(\hat{\alpha}_i)$

- Suppose $\hat{\alpha}_i > 0$. FOC … $2(\hat{\alpha}_i - \tilde{\alpha}_i) + \lambda = 0$
  so solution is

$$\tilde{\alpha}_i = \begin{cases} \hat{\alpha}_i - \lambda/2 \text{ if } (\hat{\alpha}_i - \lambda/2) > 0 \\ 0 \text{ otherwise} \end{cases}$$

- Similarly for $\hat{\alpha}_i < 0$.

Comments:

(1) No closed form expression for estimator with non-orthogonal X, but efficient computational procedures using LARS (Efron, Johnstone, Hastie, and Tibshirani (2002), Hastie, Tibshirani, Friedman (2009)).

(2) "Oracle" Results: Fan and Li (2001), Zhao and Yu (2006), Zou (2006), Leeb and Pötscher (2008), Bickel, Ritov, and Tsybakov (2009).

(3) Nice overview for economists and economic research: Belloni, Chernozhukov, and Hansen (2014); application to choosing "controls" Belloni, Chernozhukov, and Hansen (2014b), and instruments Belloni, Chen, Chernozhukov, and Hansen (2012).

(4) Bayes Interpretation: Park and Casella (2008)

Suppose $\alpha_i \sim$ iid with $f(\alpha_i)$ = constant × $\exp\left(-\gamma\left|\alpha_i\right|\right)$

Then posterior is

$$\text{constant} \times \exp\left\{-0.5\left[\frac{1}{\sigma^2}\sum_{t=1}^{T}(y_{t+1} - p_t{}'\alpha)^2 + 2\gamma\sum_{i=1}^{n}\left|\alpha_i\right|\right]\right\}$$

The lasso estimator (with $\lambda = 2\gamma\sigma^2$) yields the posterior mode.

But note mode $\neq$ mean for this distribution.

Some empirical results from Giannone, Lenza and Primiceri (2018)

Model: $y_t = u_t'\phi + x_t'\beta + \varepsilon_t$

Bayes estimation with diffuse prior for ϕ and σ^2 = var(ε)

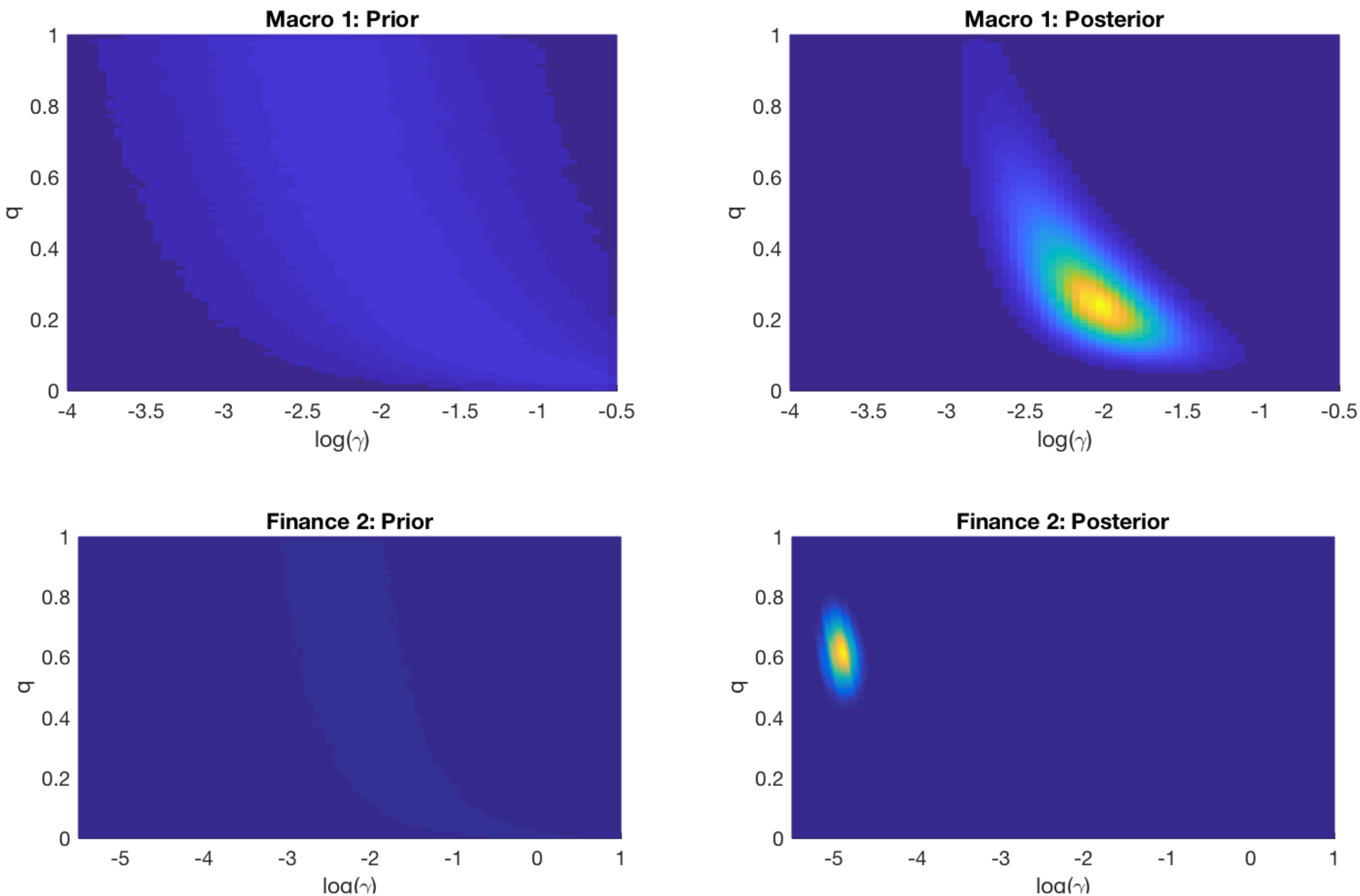
$$\beta_i \mid \sigma^2, \gamma^2, q \sim \begin{cases} N(0, \sigma^2\gamma^2) \text{ with probability } q \\ 0 \text{ with probability } (1-q) \end{cases}$$

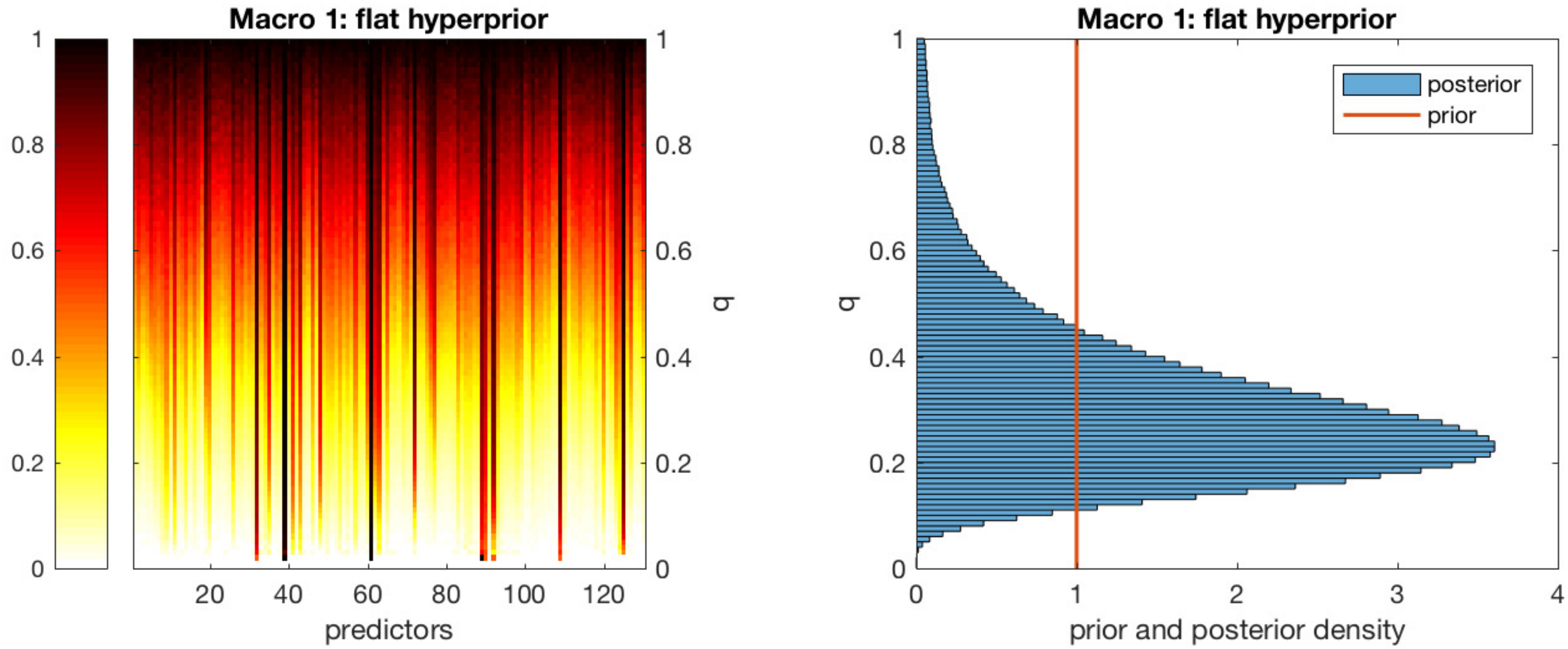
'shrinkage': γ^2 small and q large

'sparse': γ^2 large and q small

Table 1. Description of the datasets.

	Dependent variable	**Possible predictors**	**Sample**
Macro 1	Monthly growth rate of US industrial production	130 lagged macroeconomic indicators	659 monthly time-series observations, from February 1960 to December 2014
Macro 2	Average growth rate of GDP over the sample 1960-1985	60 socio-economic, institutional and geographical characteristics, measured at pre-60s value	90 cross-sectional country observations
Finance 1	US equity premium (S&P 500)	16 lagged financial and macroeconomic indicators	58 annual time-series observations, from 1948 to 2015
Finance 2	Stock returns of US firms	144 dummies classifying stock as very low, low, high or very high in terms of 36 lagged characteristics	1400k panel observations for an average of 2250 stocks over a span of 624 months, from July 1963 to June 2015
Micro 1	Per-capita crime (murder) rates	Effective abortion rate and 284 controls including possible covariate of crime and their transformations	576 panel observations for 48 US states over a span of 144 months, from January 1986 to December 1997
Micro 2	Number of pro-plaintiff eminent domain decisions in a specific circuit and in a specific year	Characteristics of judicial panels capturing aspects related to gender, race, religion, political affiliation, education and professional history of the judges, together with some interactions among the latter, for a total of 138 regressors	312 panel circuit/year observations, from 1975 to 2008


Macro 1: Prior
Macro 1: Posterior
Finance 2: Prior
Finance 2: Posterior
q
log(γ)
1
0.8
0.6
0.4
0.2
0
-4
-3.5
-3
-2.5
-2
-1.5
-1
-0.5
-5
0



Probability of variable inclusion as a function of q.

References for Lecture 5

Aruoba, S.B., Diebold, F.X., Nalewaik, J. Schorfheide, F. and Song, D. (2016), "Improving GDP Measurement: A Measurement-Error Perspective," *Journal of Econometrics,* 191, 384-397.

Bai, J., and S. Ng, (2002). “Determining the Number of Factors in Approximate Factor Models,” *Econometrica*, 70, 191-221.

Bai, J., and S. Ng, (2006). “Confidence Intervals for Diffusion Index Forecasts and Inference for Factor-Augmented Regressions,” *Econometrica*, 74, 1133-1150.

Bańbura, M., D. Giannone, M. Modugno, and L. Reichlin (2013). “Nowcasting and the Real-Time Data Flow,” ch. 4 in G. Elliott and A. Timmermann (eds.). *Handbook of Economic Forecasting*, vol. 2. Elsevier: North-Holland, 195-237.

Belloni, A. V. D. Chen, Chernozhukov, and C. Hansen (2012) “Sparse Models and Methods for Optimal Instruments with an Application to Eminent Domain,” *Econometrica*, 80:6, 657-681.

Belloni, A. V. Chernozhukov, and C. Hansen (2014a) “High-Dimensional Metho,ds and Inference on Structural and Treatment Effects,” *Journal of Economic Perspectives,* 28:2, 29-50.

Belloni, A. V. Chernozhukov, and C. Hansen (2014b) “Inference on Treatment Effects after Selection Amongst High-Dimensional Controls,” *Review of Economic Studies,* 81, 608-650.

Bickel, P.J., Y. Rotov, and A.B. Tsybakov (2009), “Simultaneous Analysis of Lasso and Dantzig Selector,” *The Annals of Statistics,* 37:4, 1705-1732.

Breiman, L. (1996), “Bagging Predictors,” *Machine Learning*, 36, 105-139.

Brillinger, D. R. (1975), *Time Series Data Analysis and Theory*. New York: Holt, Rinehart and Winston. [542]

Bühlmann, P., and Yu, B. (2002), “Analyzing Bagging,” *Annals of Statistics*, 30, 927-961.

Chamberlain, G., and M. Rothschild, (1983). “Arbitrage Factor Structure, and Mean-Variance Analysis of Large Asset Markets,” *Econometrica*, 51, 1281-1304

Connor, G., and R.A. Korajczyk, (1986). “Performance Measurement with the Arbitrage Pricing Theory,” *Journal of Financial Economics*, 15, 373-394.

Doz, C., D. Giannone, and L. Reichlin (2011). “A two-step estimator for large approximate dynamic factor models based on Kalman filtering,” *Journal of Econometrics*, 164(1). 188–205.

Doz, C., D. Giannone, and L. Reichlin, (2012). “A Quasi Maximum Likelihood Approach for Large Approximate Dynamic Factor Models,” *The Review of Economics and Statistics* 94, 1014-1024.

Efron, B., Johnstone, I., Hastie, T. and Tibshirani, R. (2004). “*Least angle regression,*” 32:2, 407-499.

Engle, R.F., and M.W. Watson, (1981). “A One-Factor Multivariate Time Series Model of Metropolitan Wage Rates,” *Journal of the American Statistical Association*, 76, 774-781.

Forni, M., and L. Reichlin, (1998). “Let’s Get Real: A Dynamic Factor Analytical Approach to Disaggregated Business Cycle,” *Review of Economic Studies*, 65, 453-474.

Geweke, J., (1977). “The Dynamic Factor Analysis of Economic Time Series,” in *Latent Variables in Socio-Economic Models*, ed. by D.J. Aigner and A.S. Goldberger, Amsterdam: North-Holland.

Giannoni, D., M. Lenza, and G.E, Primiceri (2018), "Economic Predictions with Big Data: The Illusion of Sparsity," manuscript, Northwestern University.

Hastie, T., R. Tibshirani, and J. Friedman (2009), *The Elements of Statistical Learning*, 2nd Edition, New York: Springer.

Joreskög, K.G. (1967). :A General Approach to Confirmatory Maximum Likelihood Factor Analysis," *Psychometrika*, vo. 34, No, 2, pp. 183-202.

Kilian, L. (2009). "Not All Oil Price Shocks Are Alike: Disentangling Demand and Supply Shocks in the Crude Oil Market," *American Economic Review* 99, 1053-1069.
Kim and Nelson (1998)."Business Cycle Turning Points, A New Coincident Index, and Tests of Duration Dependence Based on a Dynamic Factor Model with Regime Switching," *Review of Economics and Statistics*, 80(2), pp. 188-201.
Lawley, D.N. (1940). "The Estimation of Factor Loadings by the Method of Maximum Likelihood," Proc. Royal Society of Edinburgh, Vol. lX, pp. 64-83.
Lawley, D.N. and A.E. Maxwell (1971). *Factor Analysis as a Statistical Method*, 2nd edition, Butterworth: London.
Park, T., and Casella, G. (2008) The Bayesian Lasso, *Journal of the American Statistical Association*, 103:681-686.
Otrok, C. and C. H. Whiteman (1998). "Bayesian Leading Indicators: Measuring and Predicting Economic Conditions in Iowa," *International Economic Review*, Vol. 39, No 4., pp. 997-1014.
Quah, D., and T.J. Sargent (1993), "A Dynamic Index Model for Large Cross Sections" (with discussion), in *Business Cycles, Indicators, and Forecasting*, ed. by J.H. Stock and M.W. Watson, Chicago: University of Chicago Press for the NBER, 285-310.
Sargent, T.J., and C.A. Sims (1977). "Business Cycle Modeling Without Pretending to Have Too Much A-Priori Economic Theory," in *New Methods in Business Cycle Research*, ed. by C. Sims et al., Minneapolis: Federal Reserve Bank of Minneapolis.
Spearman, C. (1904). " 'General Intelligence,' Objectively Determined and Measured," The American Journal of Psychology, Vol. 15, No. 2, pp. 201-292.
Stock, J.H., and M.W. Watson (2002). "Forecasting Using Principal Components from a Large Number of Predictors," *Journal of the American Statistical Association*, 97, 1167-1179.
Stock, J. H. and M. W. Watson (2016), "Factor Models and Structural Vector Autoregressions in Macroeconomic," in *Handbook of Macroeconomics,* Vol2A, John B. Taylor and Harald Uhlig (eds), 2016, Chapter 8, pp 415-526.
Tibshirani, R. (1996). "Regression shrinkage and selection via the lasso," *J. Royal. Statist. Soc B*., Vol. 58, No. 1, pages 267-288).
Zhao, P. and B. Yu (2006), ′On Model Selection Consistency of Lasso," *Journal of Machine Learning*, 7, 2541-2563.
Zou, H. (2006), "The Adaptive Lasso and Its Oracle Properties," *Journal of the American Statistical Association,* 101, 1418-1429.

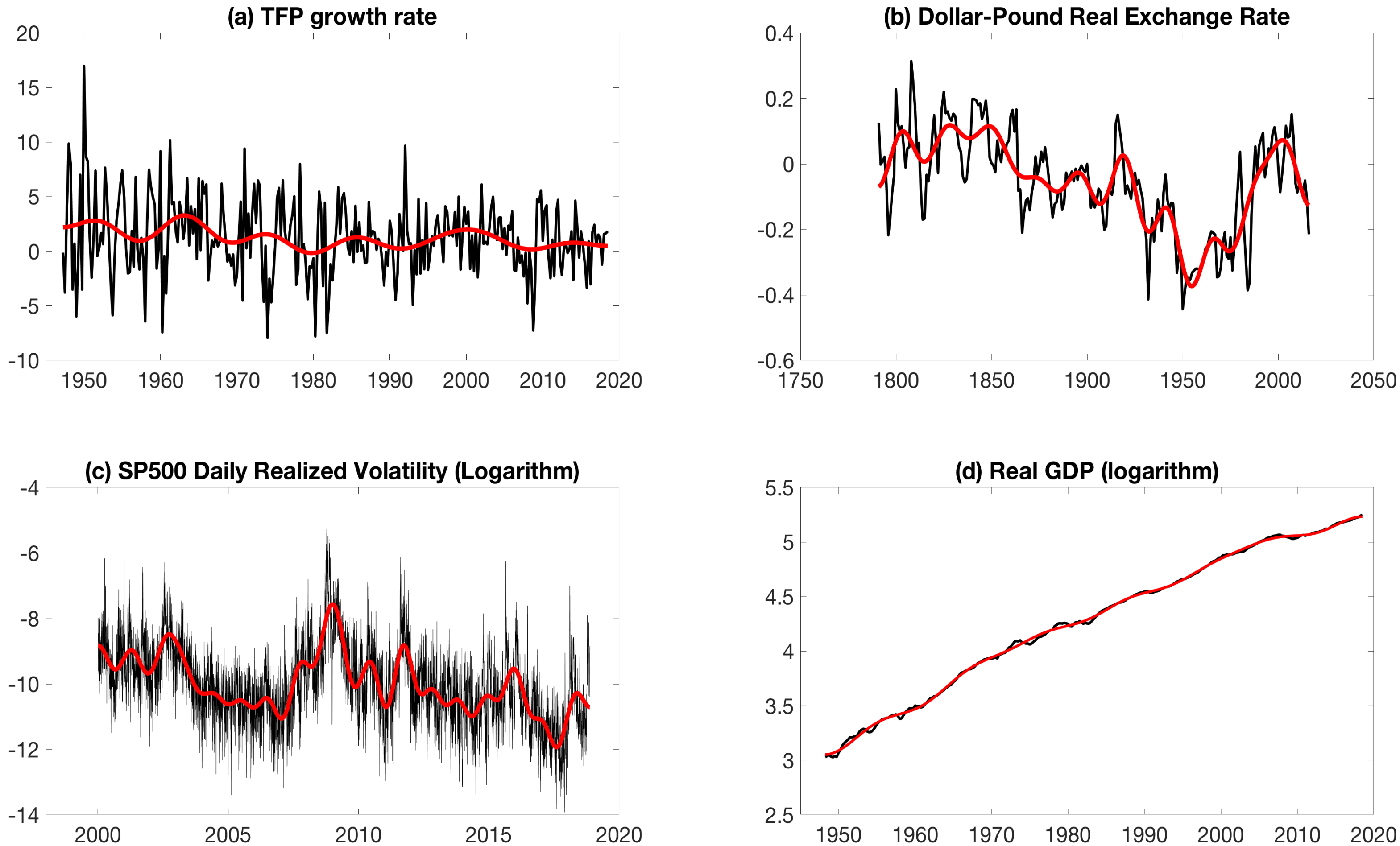
AEA Continuing Education Course

Time Series Econometrics

Lecture 6: Low-frequency analysis of economic time series

Mark W. Watson
January 8, 2019

Four Economic Time Series and 'Low-frequency' (aka 'long-run') components



Some Questions:

1. What is the long-run level ('mean') of a variable.

2. How persistent are deviations from its mean or the effect of shocks? ('halflife').

3. What is the long-run correlation between X and Y? (Or partial correlation given Z, regression coefficient, IV coefficient, …)

4. What can be said about the value of Y over the next 100 years? (What is the probability that $Y_{T+100years}$ will be between two values a and b?)

Some background (and selected references):

(1) Time trend regressions: Klein and Kosobud (1961), Grenander and Rosenblatt (1957)

(2) Spectral regression: Hannan (1963), Engle (1974)

(3) Spurious Regression: Yule (1926), Granger and Newbold (1974), Phillips(1986,1998)

(4) I(0) analysis: JS lecture for detailed references

(5) I(1) analysis: Dickey and Fuller (1979), Elliott-Rothenberg-Stock (1996), Engle and Granger (1987), Johansen (1988), …

(6) Others: I(d), local-to-unity, …: Robinson (2003), Chan and Wei (1987), Elliott (1998) …

This lecture: A return to spectral (-like) methods (extended and simplified)

Outline of Method

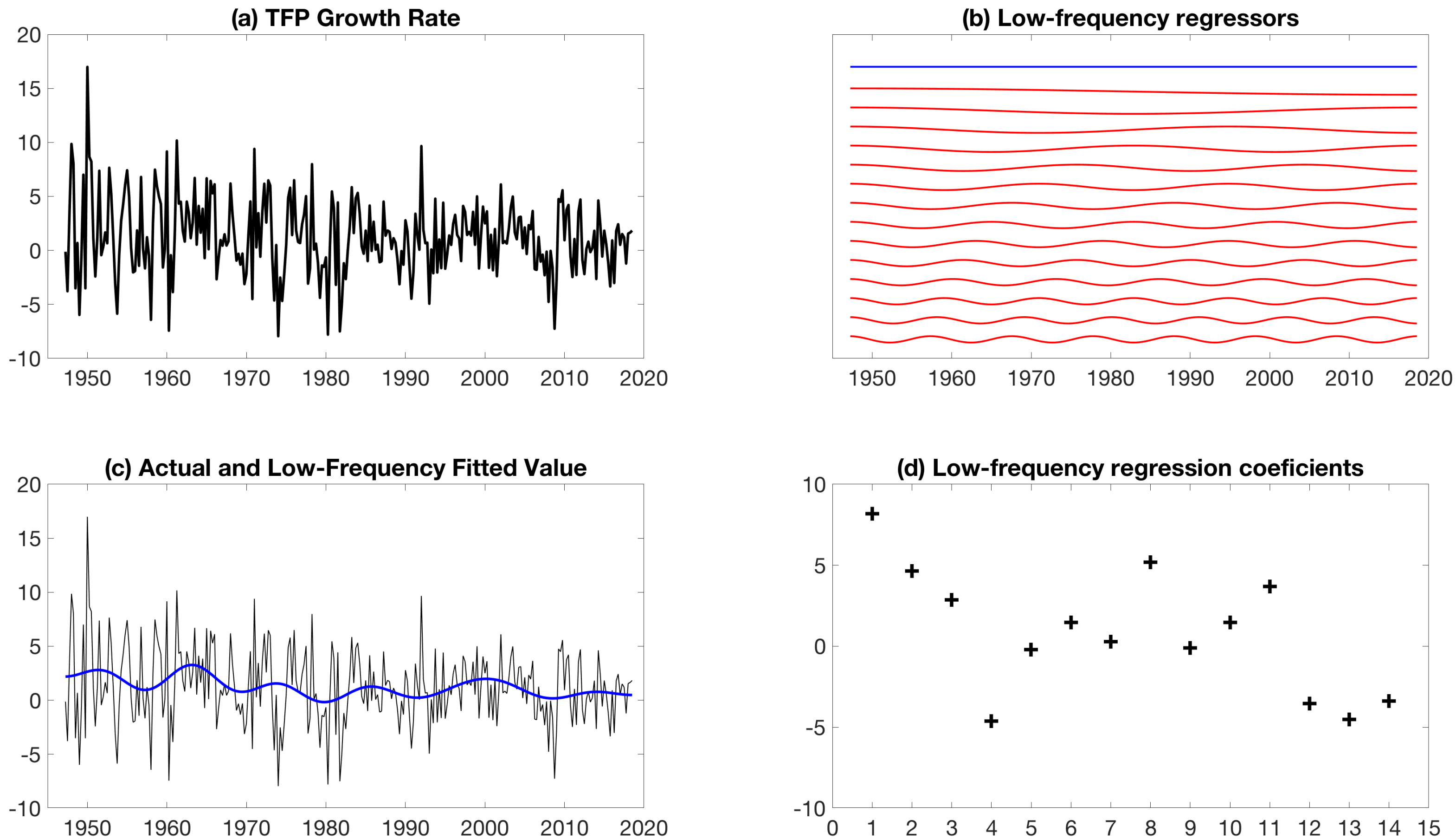
(1) Construct low-frequency components via regression onto deterministic low-frequency regressors (cosines, sines, etc.).

> (a) The estimated regression coefficients are low-frequency summaries of the sample data.
>
> (b) The estimated regression coefficients are (approximately) normally distributed.

(2) Translate low-frequency questions in questions about the normal distribution characterizing the estimated regression coefficients.

(3) Carry out 'normal' inference to answer questions.

(1) Construct low-frequency components via regression onto deterministic low-frequency regressors



(Odds and Ends: Time trend in list of low-frequency regressors)

Low-Frequency Projections

$$x_t = \bar{x}_{1:T} + \sum_{j=1}^{q} \underbrace{\sqrt{2}\cos\left(j\pi\left(\frac{t-0.5}{T}\right)\right)}_{\text{Regressor (period}=2T/j)} \underbrace{X_{jT}}_{\text{OLS Coefficient}} + \textit{residual}$$

Notation, etc.

$$x_{1:\mathrm{T}} = \bar{x}_{1:T} l_T + \Psi_T X_T + \textit{residual}$$

$$\hat{x}_{1:T} = \bar{x}_{1:T} l_T + \Psi_T X_T$$

Large-sample normality

$$\text{Under a set of conditions: } T^{1/2}X_T \Rightarrow X \sim \mathrm{N}(0, \Sigma)$$

- Conditions: … **(1–L)** $x_t = C_T(\mathrm{L})\varepsilon_t$ ('well-behaved', 'stationary', etc.)
  - (1-L) … allows x to be persistent and non-stationary.

- Σ depends on the persistence in series:
  - x is I(0), $\Sigma = \sigma^2 I$ (… JS lecture 2 on HAR)
  - x is I(1), $\Sigma = \sigma^2 D_{\mathrm{I(1)}}$ with $D_{\mathrm{I(1)},j} = 1/(j\pi)^2$
  - Generally Σ depends on spectrum of Δx_t near frequency 0

Odds and ends:

(1) Stationary processes: $$T^{1/2}\begin{bmatrix} \left(\bar{x}_{1:T}-\mu\right) \\ X_T \end{bmatrix} \Rightarrow X \sim N(0,\Sigma)$$

(2) Forecasting: $$T^{1/2}\begin{bmatrix} \left(\bar{x}_{T+1:T+h}-\bar{x}_{1:T}\right) \\ X_T \end{bmatrix} \Rightarrow X \sim N(0,\Sigma)$$

Example: HAR inference about mean in I(0) model,

Time series model: $x_t = \mu + u_t$ where $u_t \sim$ I(0)

$$T^{1/2}\begin{bmatrix} \left(\bar{x}_{1:T} - \mu\right) \\ X_T \end{bmatrix} \Rightarrow Y \sim N(0,\Sigma) \text{ with } \Sigma = \sigma^2 I$$

yields the approximation:

$$\begin{bmatrix} \bar{x}_{1:T} \\ X_T \end{bmatrix} \sim N\left(\begin{bmatrix} \mu \\ 0 \end{bmatrix}, T^{-1}\sigma^2 I \right)$$

$$\begin{bmatrix} \bar{x}_{1:T} \\ X_T \end{bmatrix} \sim N\left(\begin{bmatrix} \mu \\ 0 \end{bmatrix}, T^{-1}\sigma^2 I \right)$$

with $\quad s^2 = \frac{T}{q}\sum_{j=1}^{q} X_{jT}^2 \qquad$ then $\qquad \frac{\sqrt{T}(\bar{x}_{1:T} - \mu)}{s} \sim t_q$

<u>Frequentist CI for μ (JS Lecture 2):</u> $\quad \bar{x}_{1:T} \pm t_{q,1-\alpha/2}\sqrt{s^2/T}$ ('HAR', Müller (2004), …, multivariate extensions, etc.) (Notation: In JS lecture, this lecture's q was denoted by B.)

<u>Bayes (diffuse-prior) CS for μ:</u> $\quad \bar{x}_{1:T} \pm t_{q,1-\alpha/2}\sqrt{s^2/T}$

What is q ?

$$x_t = \bar{x}_{1:T} + \sum_{j=1}^{q} \underbrace{\sqrt{2}\cos\left(j\pi\left(\frac{t-0.5}{T}\right)\right)}_{\text{Regressor (period}=2T/j)} \underbrace{X_{jT}}_{\text{OLS Coefficient}} + residual$$

Shortest period: $2T/q$ …. Choosing q:

(1) HAR I(0) inference: how persistent are your data? ($q \approx 10$?)
(JS lecture .. used B for this lecture's q. Lots of discussion on choice of B.)

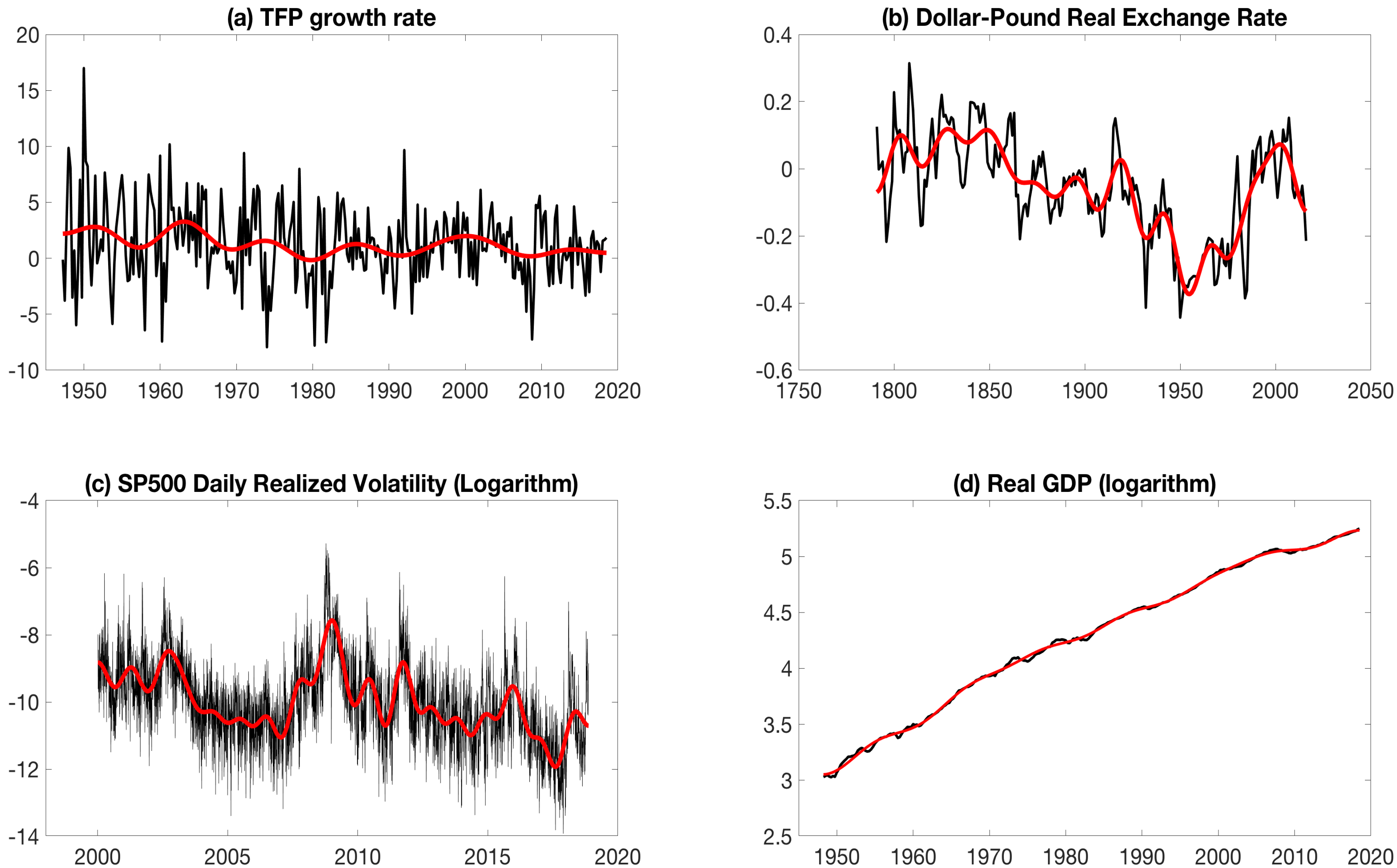
(2) Defines 'long-run' question of interest:

(a) Macro questions … periods longer than 10 years … sample size 70 years .. $q = 14$.

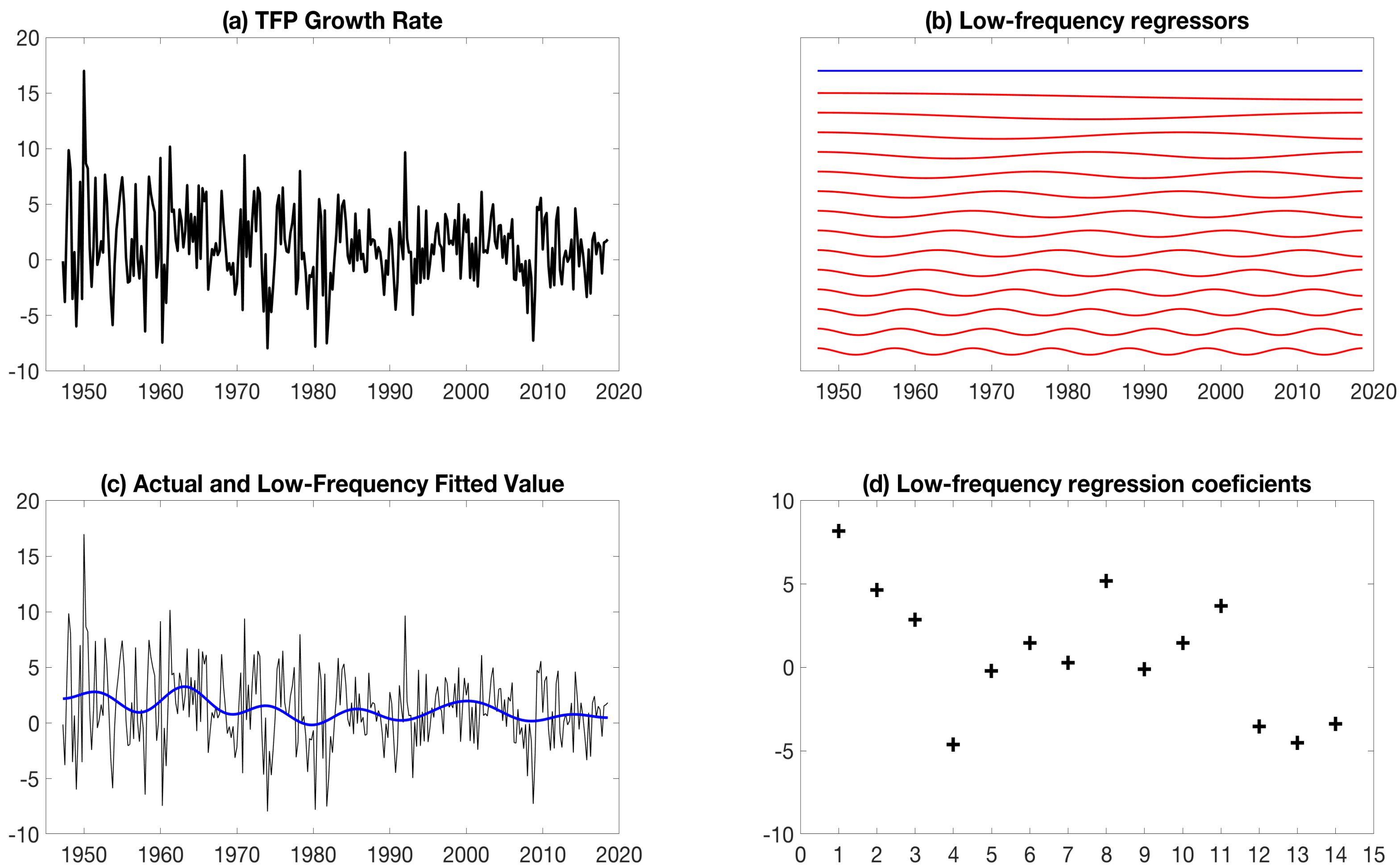
(b) PPP: … periods longer than 20 years (?) .. sample size 220 years, $q = 22$.

Limited information inference: fixed q

Four Examples



Example .. TFP .. inference about the mean



$\bar{x}_{1:T} = 1.24$

$s^2 = 14.8$

HAR-SE = $s/T^{1/2} = 0.23$

q (= df = 'B' in JS lecture) = 14

$$\bar{x}_{1:T} \pm t_{q,1-\alpha/2}\sqrt{s^2 / T}$$

	Posterior percentiles				
Parameter	0.05	1/6	0.50	5/6	0.95
μ	0.84	1.01	1.24	1.47	1.64

In this example, Bayes and Frequentist inference coincide. More generally, in these 'small-sample' (q) problems they will differ.

$X \sim \mathrm{N}(\mu, \Sigma)$: Bayes and Frequentist methods: $\mu = \mu(\theta)$, $\Sigma = \Sigma(\theta)$

Likelihood: $f(X\,|\,\theta) \propto \left|\Sigma(\theta)\right|^{-1/2} e^{-\frac{1}{2}(X-\mu(\theta))'\Sigma(\theta)^{-1}(X-\mu(\theta))}$

Bayes: $\theta \sim f_{prior}$, then invert to find posterior $f(\theta\,|\,X)$. (extensions to predictive distributions, etc.)

Frequentist: $\theta = (\theta_1, \theta_2)$ H_0: $\theta_1 = \theta_{1,0}$ and H_1: $\theta_1 \neq \theta_{1,0}$

A bit more complicated because of θ_2, but well-studied problem and many standard ways to handle.

Questions:

1. Long-run level ('mean')

2. Long-run persistence ('half-life')

3. Long-run correlation (correlation, linear regression, IV, etc.)

4. Long-run predictions (point forecasts and uncertainty bands)

Bayes inference examples follow

Inference about persistence parameters:

$T^{1/2}X_T \Rightarrow X \sim \mathrm{N}(0, \Sigma)$, $\Sigma = \sigma^2\Omega(\theta)$, θ is persistence parameter.

($\bar{x}_{1:T}$ not used. Diffuse prior for μ is stationary models, no well-defined limit in non-stationary models).

Generic procedure:

Gaussian likelihood: $f(X_T \mid \theta, \sigma^2) \propto \left|\sigma^2\Omega(\theta)\right|^{-1/2} e^{-\frac{T}{2\sigma^2}X_T\Omega(\theta)^{-1}X_T}$

Specify prior for σ^2 and θ

Turn Bayes crank. (Koop (2003), Geweke (2005))

Digression: Bayes analysis with discrete θ.

Suppose $\theta \in \{\theta_1, \theta_2, \ldots \theta_k)$ with $P(\theta = \theta_i) = p_i$.

$$P(\theta = \theta_i \mid X) = \frac{f(X_T \mid \theta_i) p_i}{\sum_{j=1}^{k} f(X_T \mid \theta_j) p_j}$$

Example: Local-level-persistence

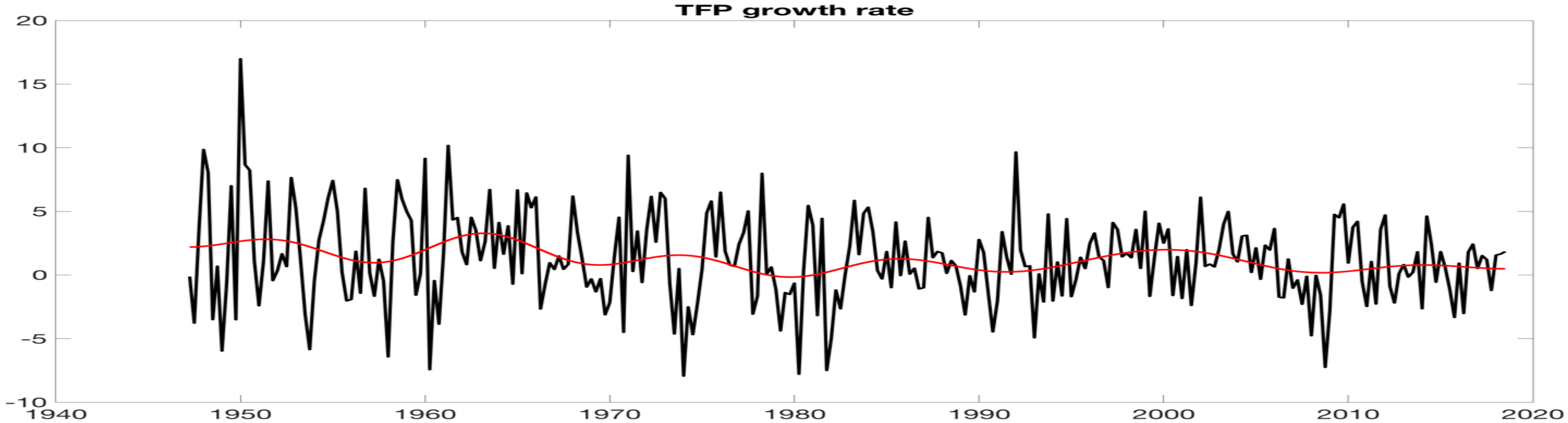
- Time series model: $x_t = a_t + b_t$ where a_t is I(0) and b_t is I(1)

- Parameters: long-run standard deviations σ_a and $\sigma_{\Delta b}$

- Standard Parameterization: $b_t = b_0 + (\theta/T)\sum_{i=1}^{t} e_i$, where (a,e) are mutually uncorrelated I(0) processeses with LRV σ^2.

- In this case $\Sigma = \sigma^2\Omega(\theta)$ with $\Omega(\theta) = I + \theta D_{I(1)}$

Example: TFP (Shortest period = 10 years, q = 14)



$$x_t = a_t + b_t,\ \ b_t = b_0 + (\theta/T)\sum_{i=1}^{t} e_i$$

Diffuse prior for σ, $\ln(\theta) \sim U[\ln(0.1), \ln(500)]$; ('equally-spaced grid').

	Posterior percentiles				
Parameter	**0.05**	**1/6**	**0.50**	**5/6**	**0.95**
θ	0.14	0.48	4.58	10.92	21.89
$\sigma = \sigma_a$	2.076	2.573	3.332	4.258	5.089
$\boldsymbol{\sigma\,\theta/T = \sigma_b}$	0.002	0.007	0.052	0.117	0.190
$\theta/T = \sigma_b/\sigma_a$	0.000	0.002	0.016	0.038	0.077

Standard deviation of change in b_t over 10 years = $\sigma_b\sqrt{40}$ ≈ 0.33

Example: AR-persistence

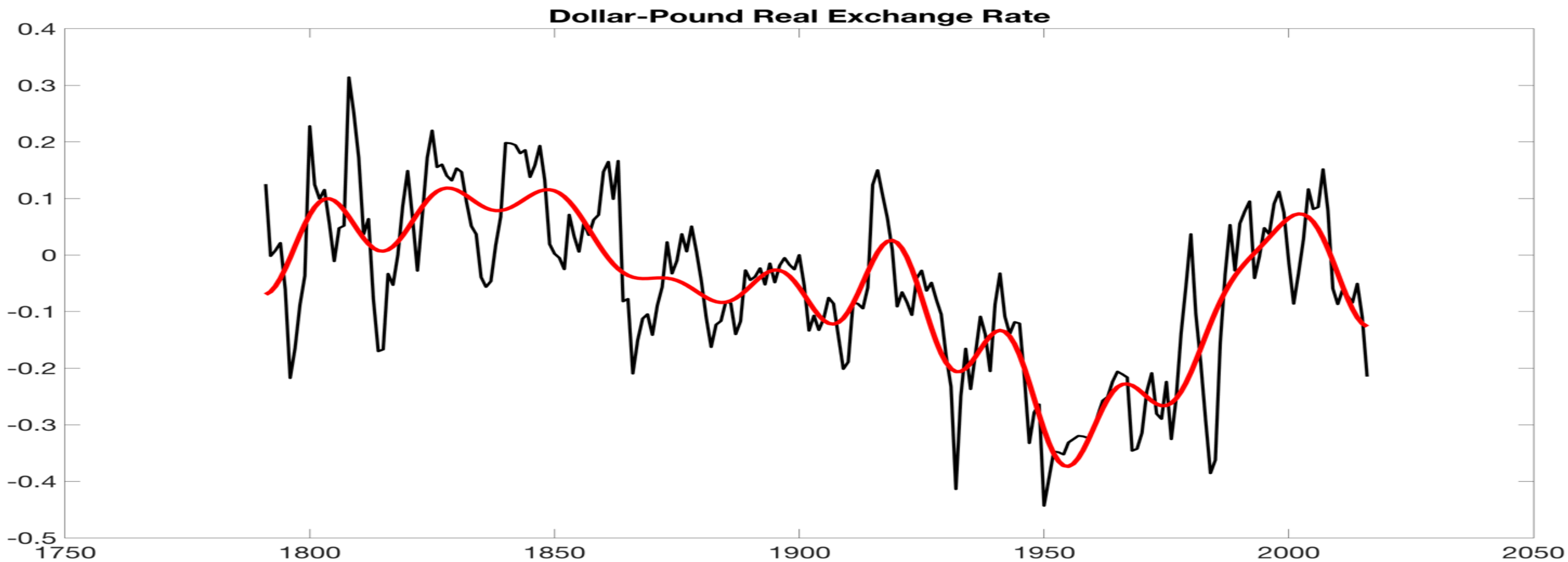
- Time series model: $x_t = \mu + u_t$ where $u_t = \rho u_{t-1} + a_t$, where $\rho \approx 1$ and $a_t \sim \mathrm{I}(0)$

- Parameters: long-run standard deviations ρ and σ (= LR SD of a)

- Standard Parameterization: $\rho = (1 - \theta/T)$ (θ is local-to-unity parameter)

- In this case $\Sigma = \sigma^2\Omega(\theta)$ with $\Omega(\theta)$ LTU variance.

Example: Real Exchange Rate (Shortest period = 20 years, q = 22)



$$x_t = \mu + u_t \text{ where } u_t = \rho u_{t-1} + a_t,\ \rho = (1 - \theta/T)$$

Diffuse prior for σ, $\ln(\theta) \sim U[\ln(0.1), \ln(500)]$; ('equally-spaced grid').

	Posterior percentiles				
Parameter	**0.05**	**1/6**	**0.50**	**5/6**	**0.95**
θ	0.17	0.68	5.45	12.99	21.89
$\rho = (1-\theta/T)$	0.90	0.94	0.98	1.00	1.00
$\rho^{half\text{-}life} = 1/2$	7.2	12.1	28.8	231.5	930.0

Example: I(d)-persistence

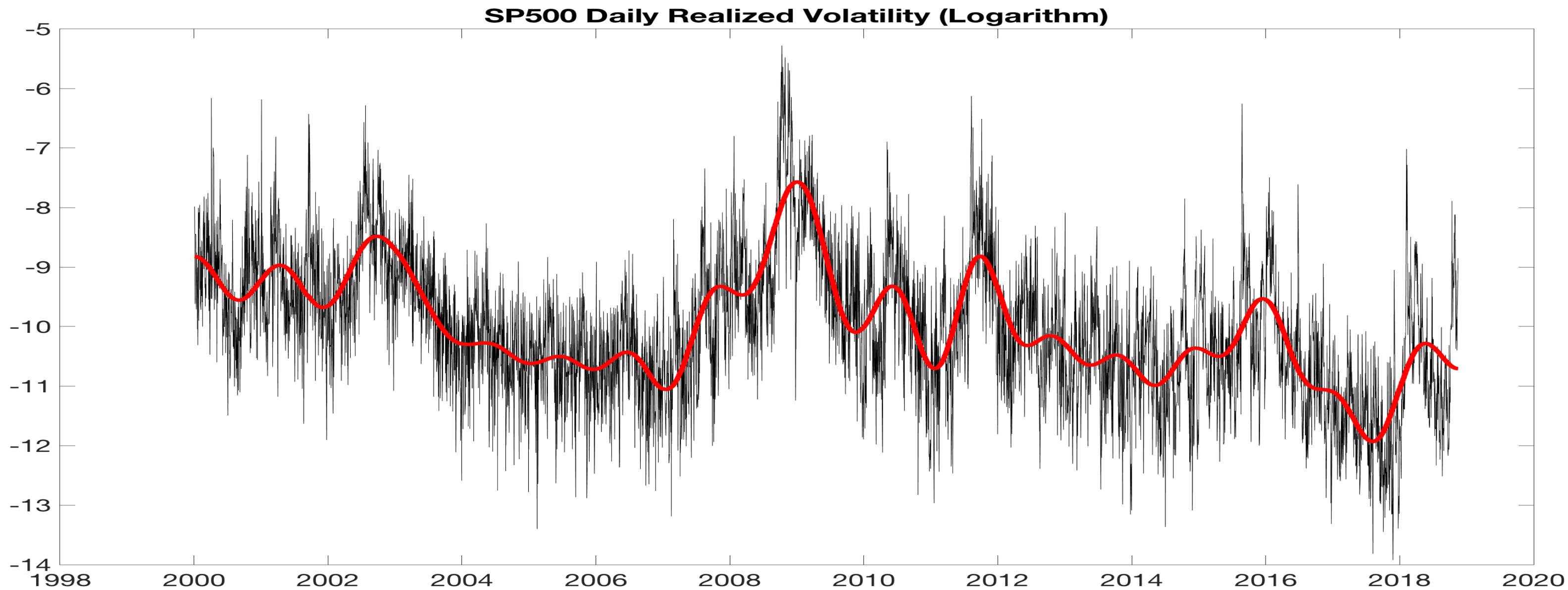
- Time series model: $x_t = \mu + u_t$ with $(1–L)^d u_t = e_t$ where $e_t \sim I(0)$

- Parameters: d and σ (= LR SD of e)

- Standard Parameterization: here $\theta = d$. (Limiting normality obtains for $–0.5 < d < 1.5$)

- In this case $\Sigma = \sigma^2\Omega(\theta)$ with $\Omega(\theta)$ fractional variance.

Example: Logarithm of daily SP500 realized volatility (Shortest period = 250 days, q = 37)



$$x_t = \mu + u_t \text{ with } (1\text{–L})^d\, u_t = e_t \text{ where } e_t \sim \text{I}(0)$$

Diffuse prior for σ, $d \sim$ U[-0.4, 1.4]; ('equally-spaced grid').

	Posterior percentiles				
Parameter	0.05	1/6	0.50	5/6	0.95
$\theta = d$	0.37	0.44	0.59	0.74	0.81

Return to inference about long-run level ('mean')

- $x_t = \mu + u_t$
  - $u_t \sim \text{I}(0)$ (done)
  - u_t persistent, but stationary (*Not covered by JS HAR lecture*)

- $x_t = \mu_t + u_t$

Example: Inference about mean when data are persistent

- Time series model: $x_t = \mu + u_t$ where $u_t = \rho u_{t\text{-}1} + e_t$, $e_t \sim$ I(0) and ρ is close to 1.

- Parameters: μ, ρ (= (1 – θ/T)), σ

- large-sample approximation

$$T^{1/2}\begin{bmatrix} \left(\bar{x}_{1:T} - \mu\right) \\ X_T \end{bmatrix} \sim N(0, \sigma^2\Omega(\theta))$$

Example: Unemployment rate (nsa), shortest period = 10 years, q = 14

$$x_t = \mu + u_t \text{ with } u_t = \rho u_{t-1} + e_t \text{ where } e_t \sim \mathrm{I}(0) \text{ and } \rho = (1-\theta/T)$$

Diffuse prior for μ and σ, $\ln(\theta) \sim U[\ln(0.1), \ln(500)]$; ('equally-spaced grid').

	Posterior percentiles				
Parameter	0.05	1/6	0.50	5/6	0.95
μ	4.43	5.19	5.72	6.18	6.61
$\rho = (1-\theta/T)$	0.51	0.71	0.93	0.98	1.00
μ, I(0)	5.16	5.39	5.77	6.14	6.38

Example: Inference about time varying 'level' in local-level model

- Time series model: $x_t = a_t + b_t$, $b_t = b_0 + (\theta/T)\sum_{i=1}^{t} e_i$

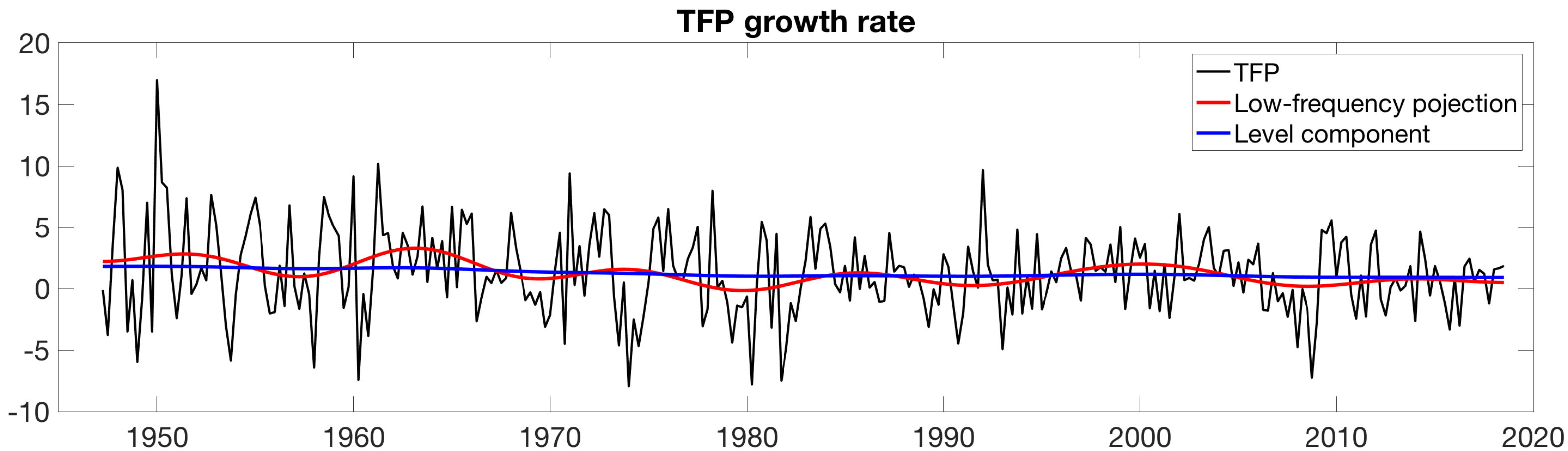
- Question: What the (low-frequency) value of b_t | low-frequency of x_t. (As in Kalman filter example)

  - $x_t \rightarrow X_T$ and $(a_t + b_t) \rightarrow A_T + B_T$

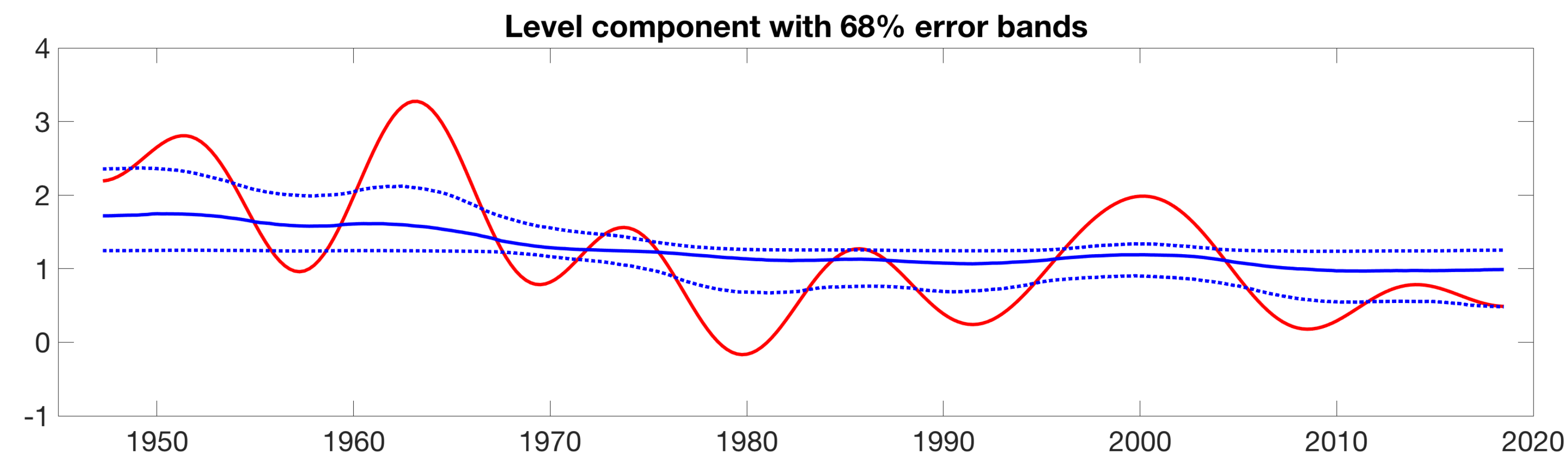
  - what is the value of B_T given X_T $(= A_T + B_T)$

- Parameters, same as LLM but now object of interest is predictive distribution: $f(B_T \mid X_T)$

Example: TFP



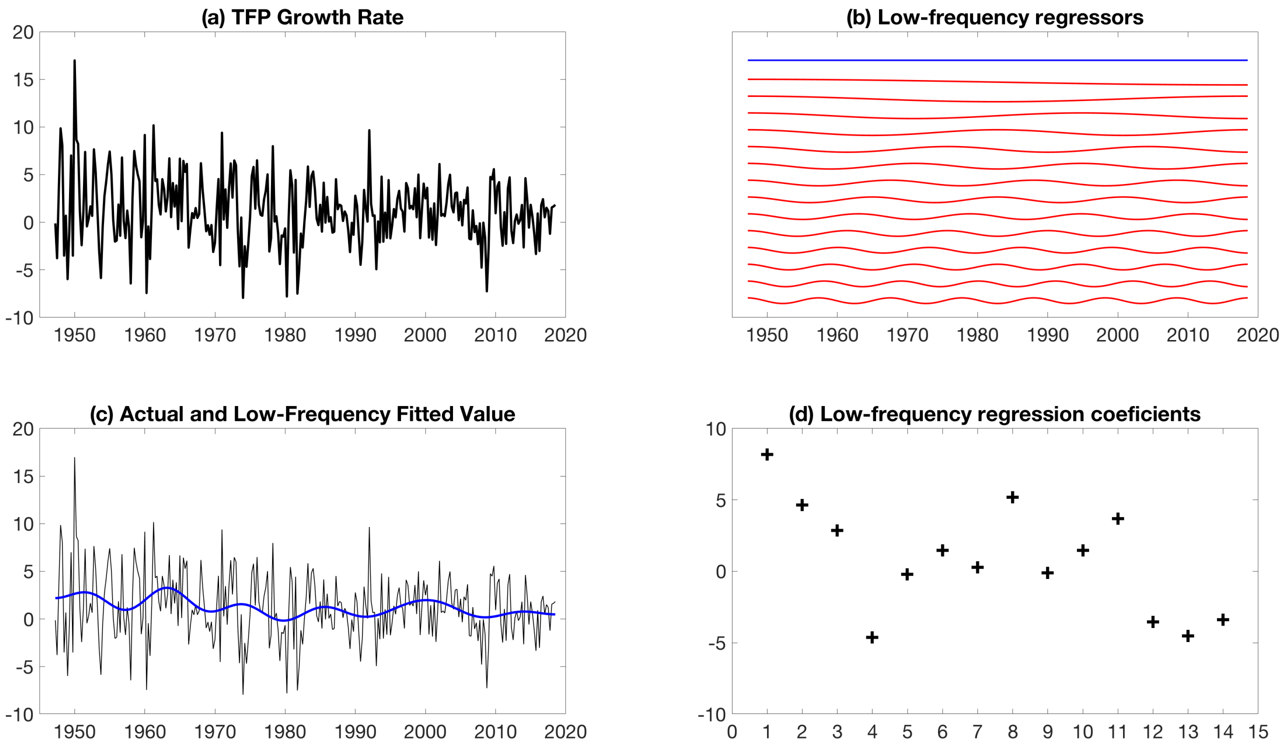
TFP growth rate
20
15
10
5
0
-5
-10
1950
1960
1970
1980
1990
2000
2010
2020
TFP
Low-frequency pojection
Level component



Level component with 68% error bands
4
3
2
1
0
-1
1950
1960
1970
1980
1990
2000
2010
2020

Example: Inference about break in mean

- Time series model: $x_t = \mu_t + u_t$ where $u_t \sim \mathrm{I}(0)$ and $\mu_t = \mu + 1(t > rT)\times\delta$
- Parameters, σ, μ, δ and $0 < r < 1$

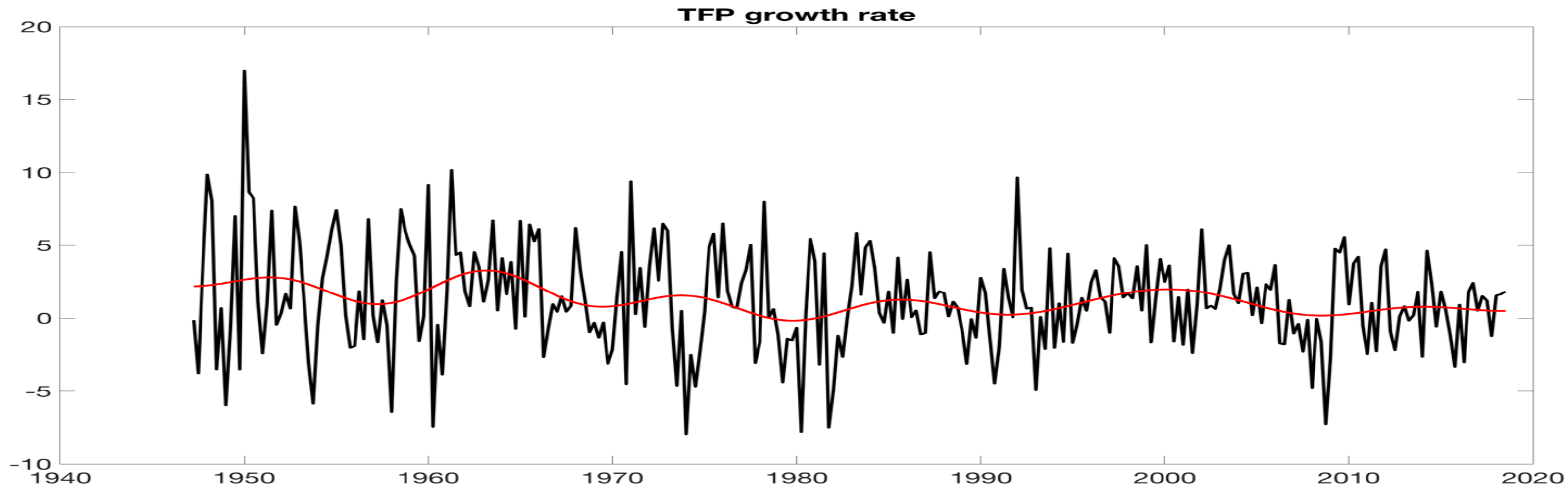


Mean shift induces a non-zero mean in projection coefficients, where mean depends on δ and r.

$$x_t = \bar{x}_{1:T} + \sum_{j=1}^{q} \underbrace{\sqrt{2}\cos\left(j\pi\left(\frac{t-0.5}{T}\right)\right)}_{\text{Regressor (period}=2T/j)} \underbrace{X_{jT}}_{\text{OLS Coefficient}} + \ residual$$

$$T^{1/2}\left[\begin{pmatrix} \bar{x}_{1:T} \\ X_T \end{pmatrix} - m(\mu,\delta,r)\right] \sim N(0,\sigma^2 I)$$

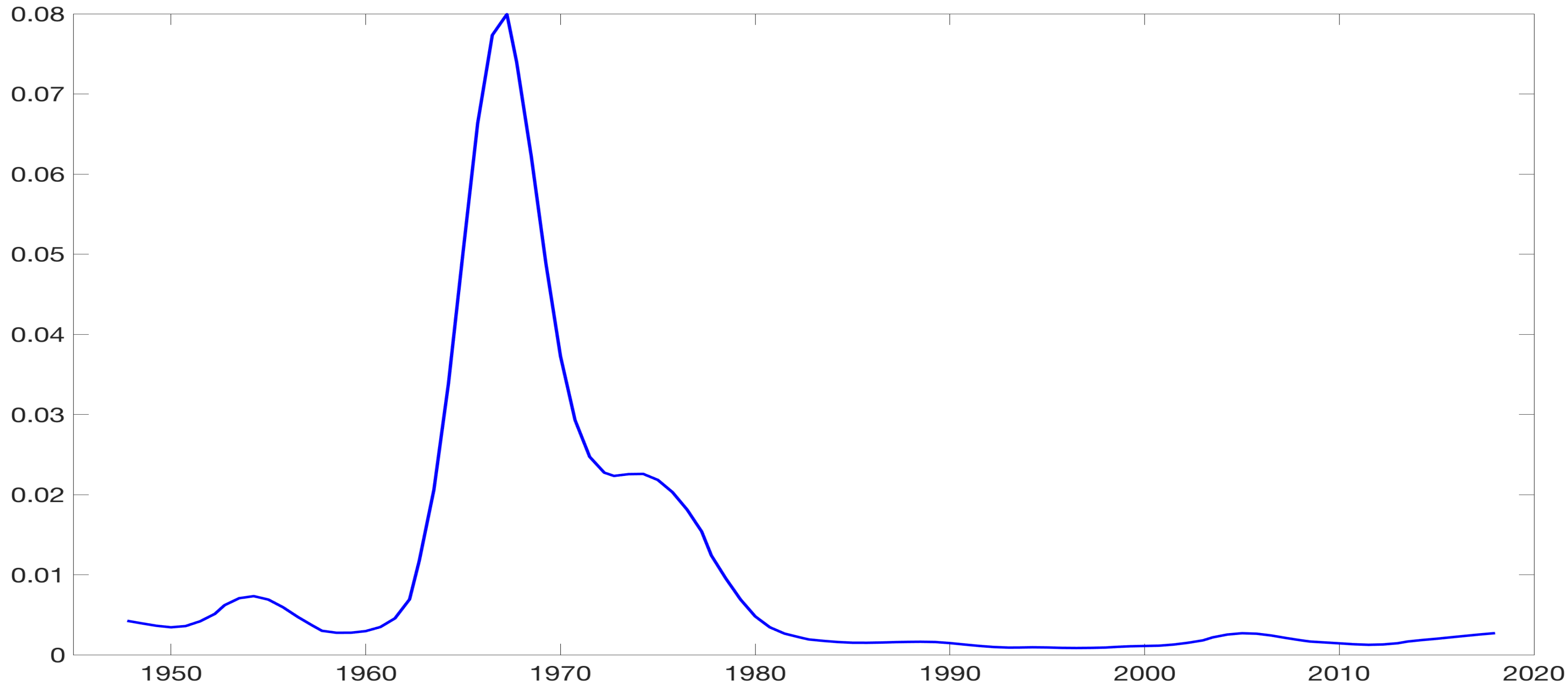
Example: TFP (Shortest period = 10 years, q = 14)



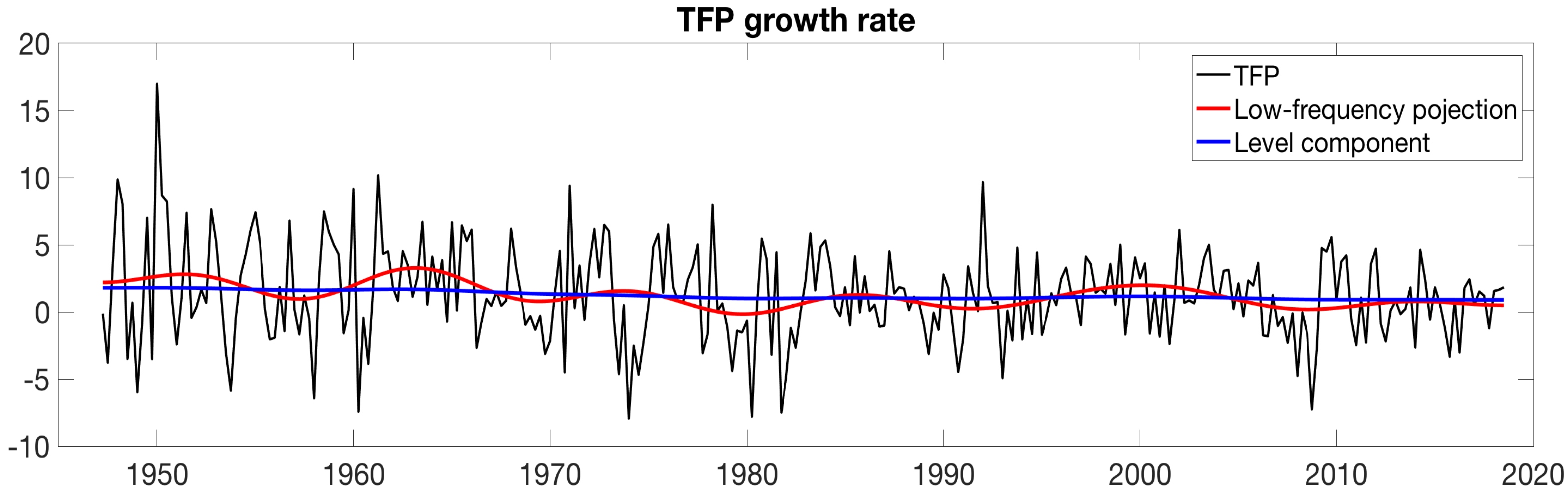
$$x_t = \mu_t + u_t \text{ where } u_t \sim \mathrm{I}(0) \text{ and } \mu_t = \mu + 1(t > rT)\times\delta$$

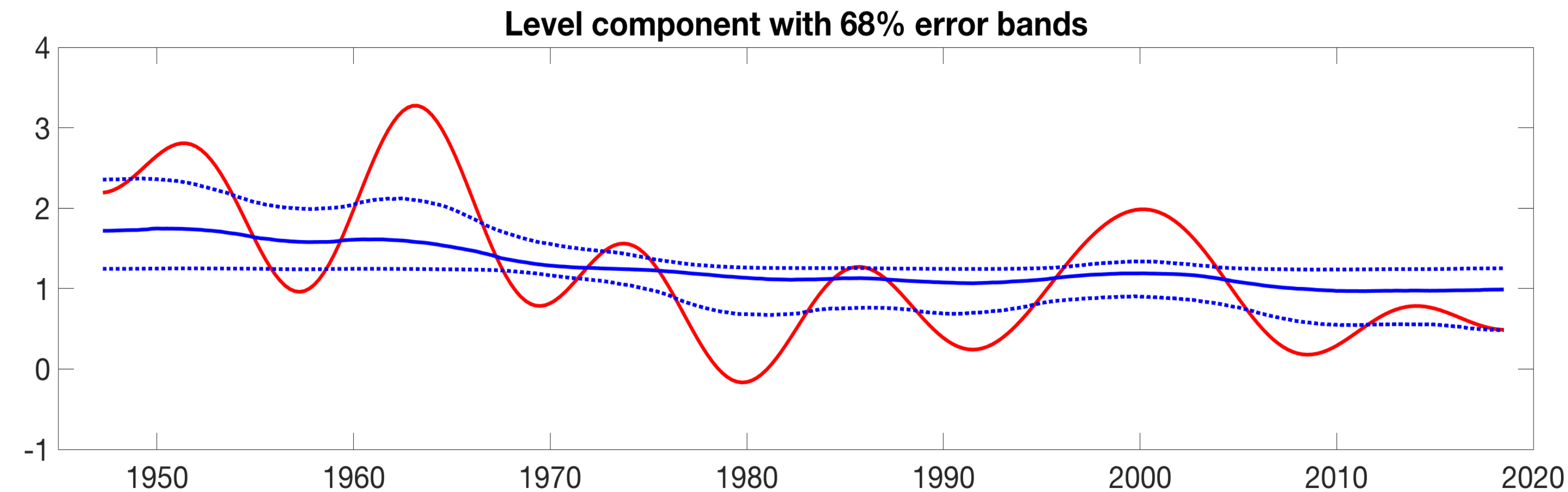
Diffuse prior for σ, μ, δ; $r \sim \mathrm{U}(0,1)$ (discrete approximation)

Break date posterior



	Posterior percentiles				
Parameter	0.05	1/6	0.50	5/6	0.95
μ_{pre}	1.32	1.69	2.12	2.51	2.87
μ_{post}	0.40	0.62	0.86	1.10	1.30
$\mu_{pre} - \mu_{post}$	0.37	0.81	1.29	1.73	2.08


TFP growth rate
TFP
Low-frequency pojection
Level component
20
15
10
5
0
-5
-10
1950
1960
1970
1980
1990
2000
2010
2020


Level component with 68% error bands
4
3
2
1
0
-1
1950
1960
1970
1980
1990
2000
2010
2020

Questions:

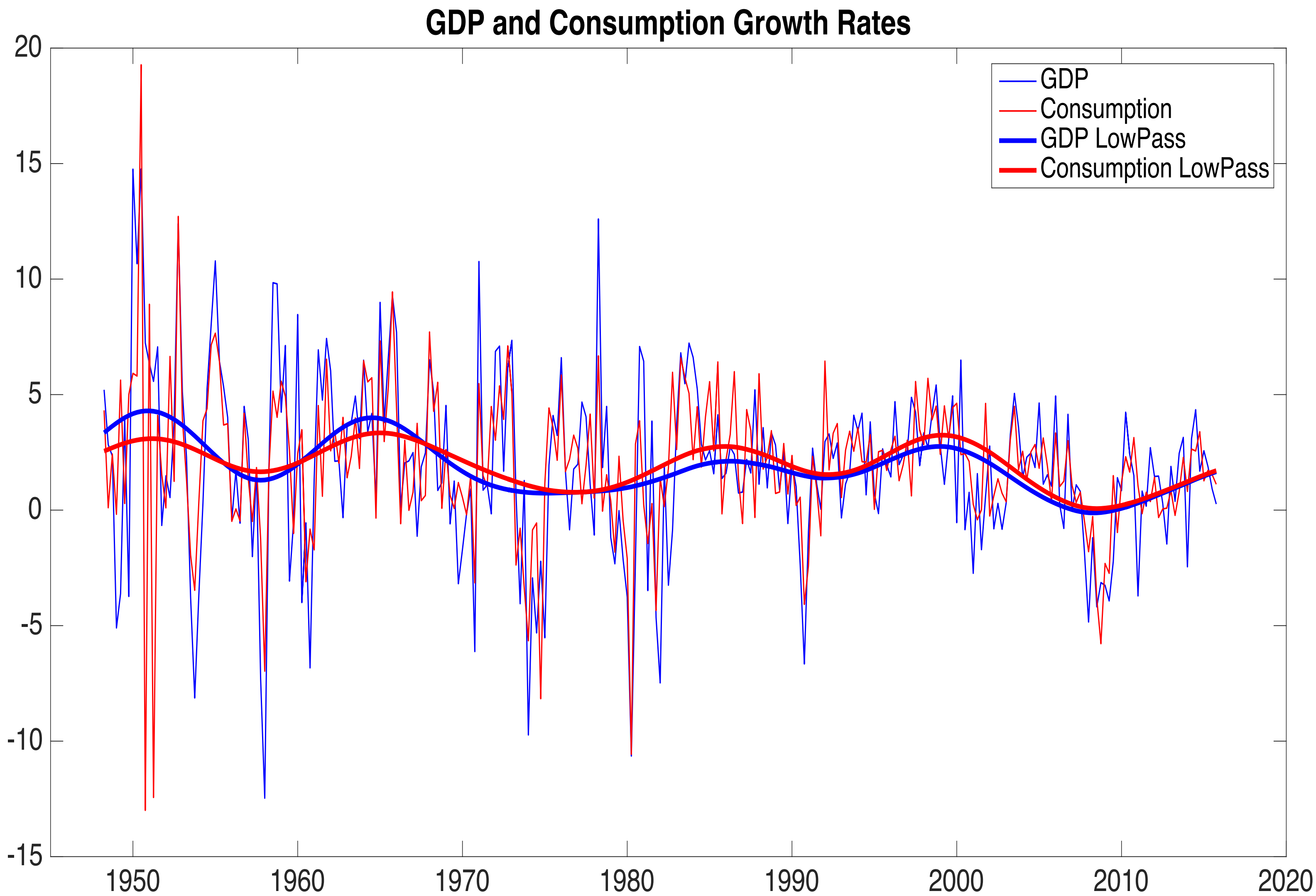
~~1. Long-run level ('mean')~~

~~2. Long-run persistence ('halflife')~~

3. Long-run correlation (correlation, linear regression, IV, etc.)

4. Long-run predictions (point forecasts and uncertainty bands)

Two Series


GDP and Consumption Growth Rates
GDP
Consumption
GDP LowPass
Consumption LowPass
20
15
10
5
0
-5
-10
-15
1950
1960
1970
1980
1990
2000
2010
2020

Variances and Covariances

$$\hat{x}_{1:T} = \bar{x}_{1:T} l_T + \Psi_T X_T$$

$$\hat{y}_{1:T} = \bar{y}_{1:T} l_T + \Psi_T Y_T$$

Let

$$\tilde{x}_t = \hat{x}_t - \bar{x}_{1:T}$$

Variances and Covariances

$$\Omega_T = E\left[T^{-1}\sum_{t=1}^{t}\begin{pmatrix}\tilde{x}_t & \tilde{y}_t\end{pmatrix}\begin{pmatrix}\tilde{x}_t \\ \tilde{y}_t\end{pmatrix}\right]$$

$$= E\left[T^{-1}\begin{pmatrix}X_T' & Y_T'\end{pmatrix}\Psi_T'\Psi_T\begin{pmatrix}X_T \\ Y_T\end{pmatrix}\right]$$

$$= E\begin{bmatrix}X_T'X_T & X_T'Y_T \\ Y_T'X_T & Y_T'Y_T\end{bmatrix}$$

Ψ_T are 'special': $T^{-1}\,\Psi_T'\Psi_T = I_q$

Multivariate CLT:

$$T^{1/2}\begin{bmatrix} X_T \\ Y_T \end{bmatrix} \Rightarrow \begin{bmatrix} X \\ Y \end{bmatrix} \sim N\left(0, \begin{bmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{bmatrix}\right)$$

Variances and Covariances

$$\Omega_T = E\begin{bmatrix} X_T{}'X_T & X_T{}'Y_T \\ Y_T{}'X_T & Y_T{}'Y_T \end{bmatrix} = \begin{bmatrix} Tr(\Sigma_{XX,T}) & Tr(\Sigma_{XY,T}) \\ Tr(\Sigma_{XY,T}) & Tr(\Sigma_{YY,T}) \end{bmatrix}$$

The large-sample limit of this is

$$\Omega = E\begin{bmatrix} X'X & X'Y \\ Y'X & Y'Y \end{bmatrix} = \begin{bmatrix} Tr(\Sigma_{XX}) & Tr(\Sigma_{XY}) \\ Tr(\Sigma_{XY}) & Tr(\Sigma_{YY}) \end{bmatrix}$$

Straightforward to do inference about Ω

Functions of Ω

(1) Correlations, partial correlations

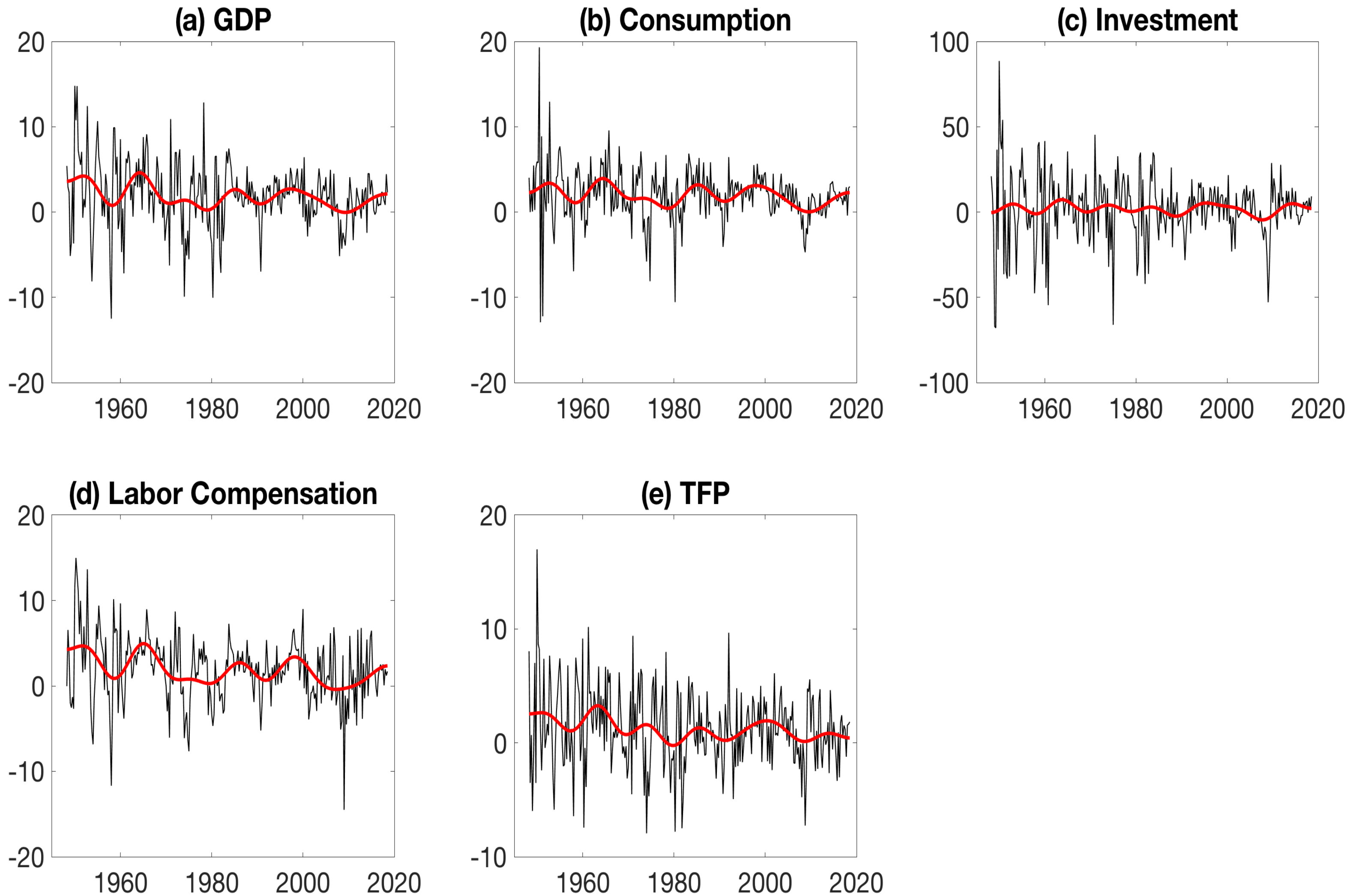
(2) Linear regression coefficients

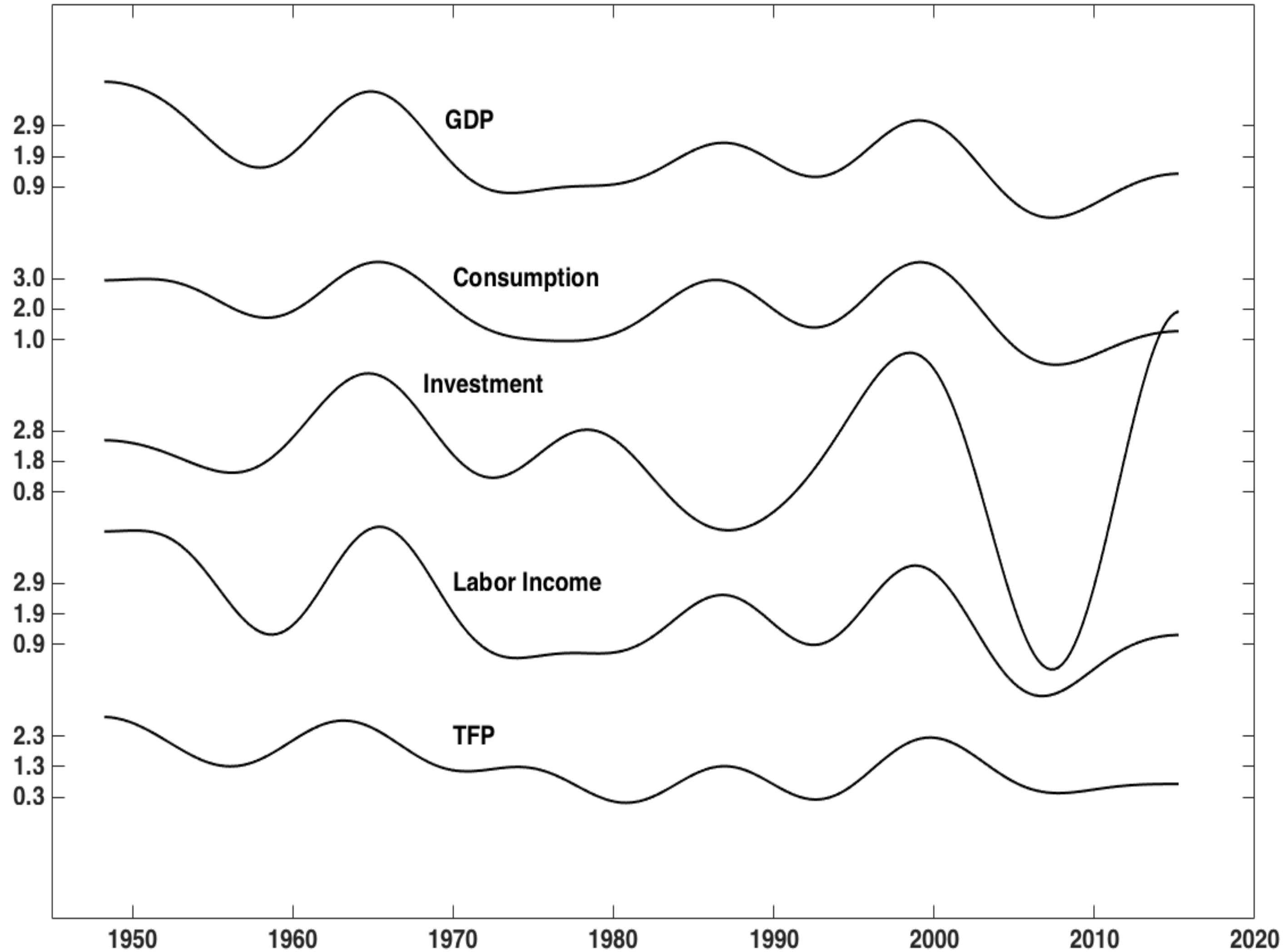
(3) Linear IV

Challenge: How to parameterize

$$\begin{bmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{bmatrix}$$

Example: 5 Macro Variables




GDP
Consumption
Investment
Labor Income
TFP
2.9
1.9
0.9
3.0
2.0
1.0
2.8
1.8
0.8
2.9
1.9
0.9
2.3
1.3
0.3
1950
1960
1970
1980
1990
2000
2010
2020

One-factor model

$$x_t = \begin{bmatrix} gdp \\ consumption \\ investment \\ w \\ tfp \end{bmatrix} = \begin{bmatrix} 1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ \lambda_5 \end{bmatrix} f_t + \begin{bmatrix} u_{1,t} \\ u_{2,t} \\ u_{3,t} \\ u_{4,t} \\ u_{5,t} \end{bmatrix}$$

$f_t \sim$ LLM: $f_t = a_t + b_t$ with $b_t = b_0 + (\theta / T)\sum_{i=1}^{t} e_i$

$u_{i,t} \sim \mathrm{I}(d)$

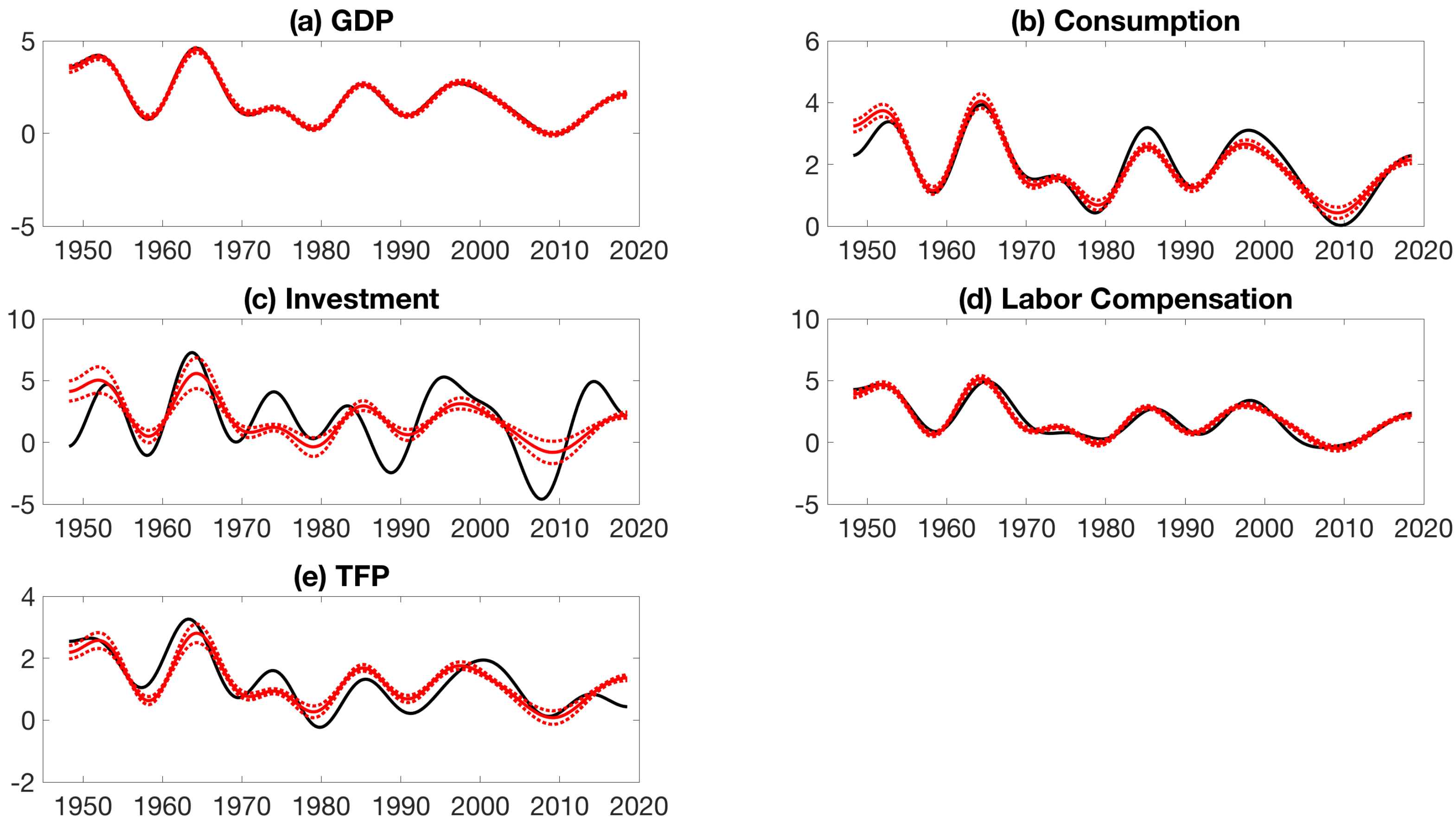
Results: Posterior Median and 68% credible sets

Variable	λ	$R^2(f)$	d
GDP	1.00 (1.00, 1.00)	0.93 (0.68, 0.99)	-0.22 (-0.35,-0.02)
Cons	0.80 (0.71, 0.89)	0.87 (0.51, 0.98)	0.11 (-0.07, 0.25)
Investment	1.41 (0.95, 1.88)	0.26 (0.03, 0.79)	0.04 (-0.22, 0.27)
W	1.25 (1.14, 1.34)	0.70 (0.27, 0.95)	-0.16 (-0.32, 0.07)
TFP	0.60 (0.49, 0.71)	0.70 (0.52, 0.83)	0.12 (-0.06, 0.27)

f_t process parameters

Parameter	Median and 68% credible set
θ	2.27 (0.34 10.92)
θ/T	0.010 (0.001 0.039)
$\sigma \times \theta/T$	0.05 (0.01 0.15)

Low-frequency projection: Series and common component



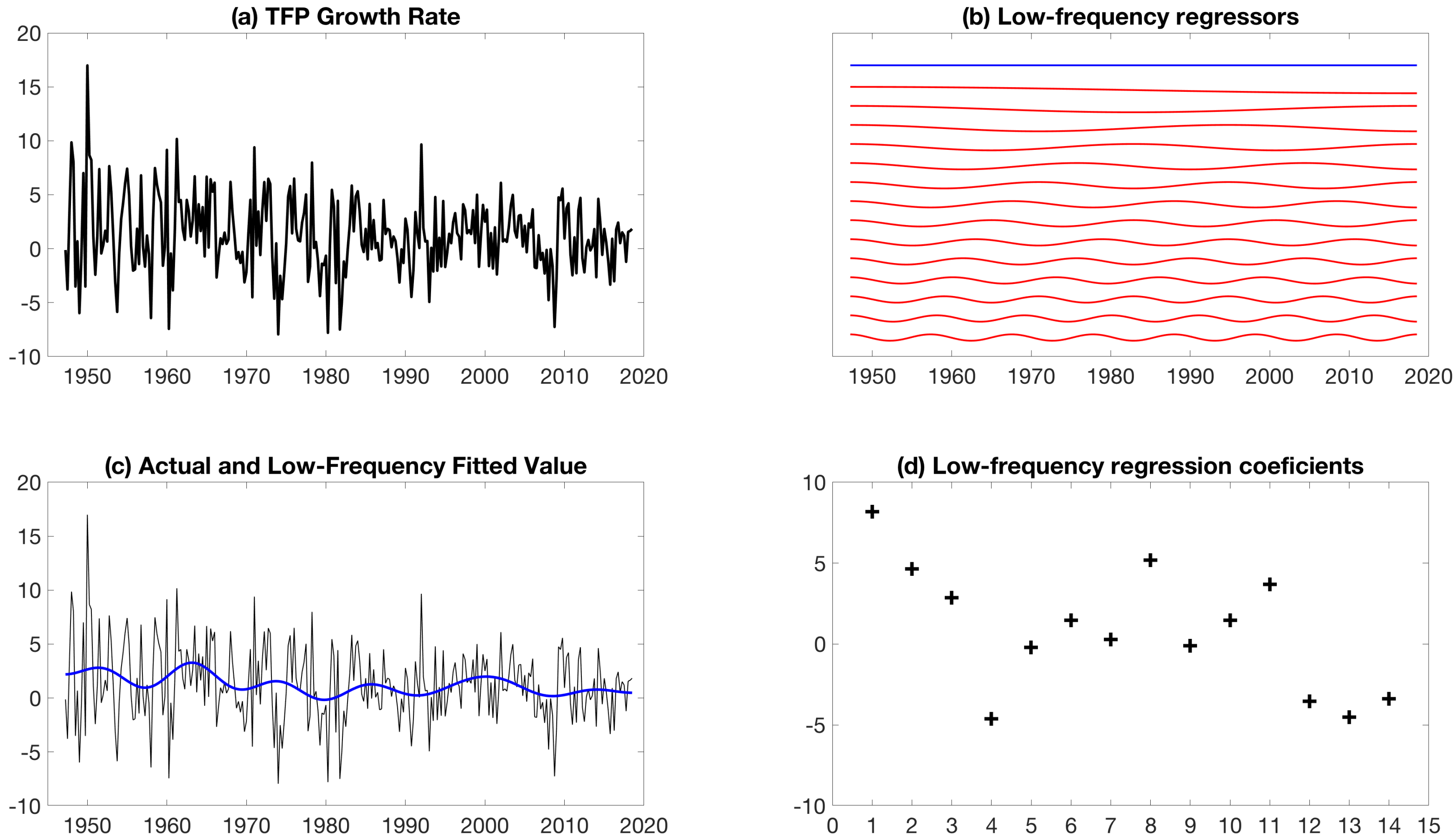
Questions:

~~1. Long-run level ('mean')~~

~~2. Long-run persistence ('halflife')~~

~~3. Long-run correlation (correlation, linear regression, IV, etc.)~~

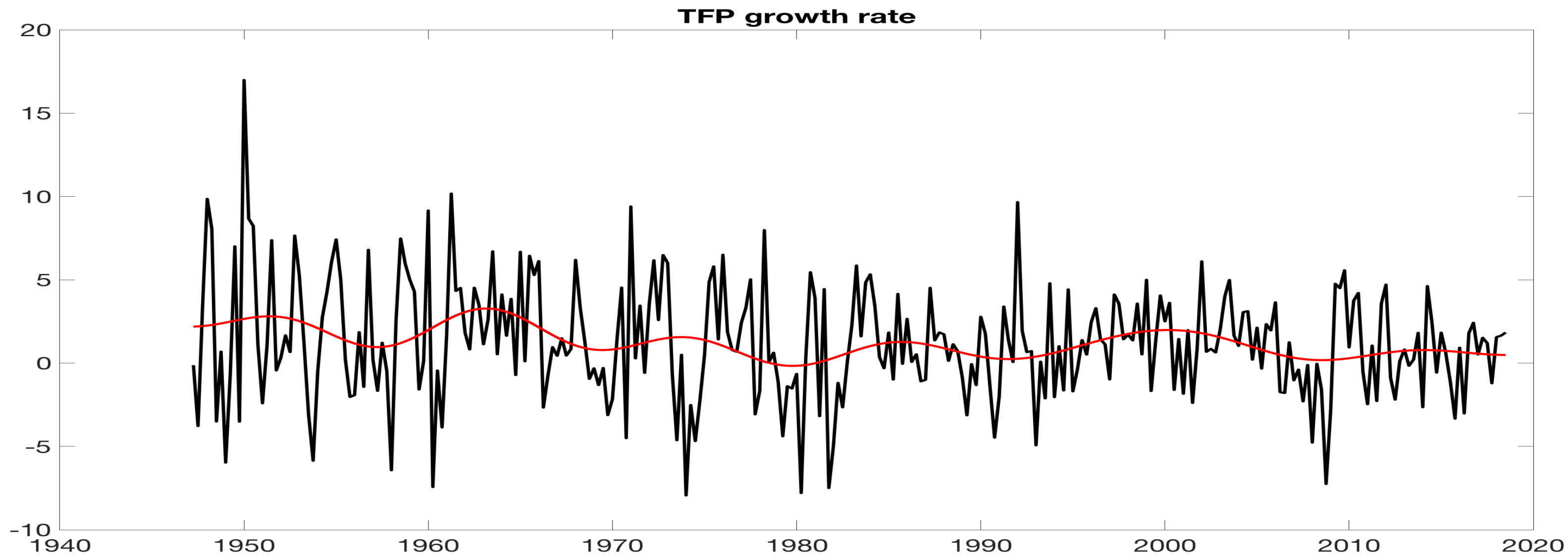
4. Long-run predictions (point forecasts and uncertainty bands)



(a) TFP Growth Rate
20
15
10
5
0
-5
-10
1950 1960 1970 1980 1990 2000 2010 2020
(b) Low-frequency regressors
1950 1960 1970 1980 1990 2000 2010 2020
(c) Actual and Low-Frequency Fitted Value
20
15
10
5
0
-5
-10
1950 1960 1970 1980 1990 2000 2010 2020
(d) Low-frequency regression coeficients
10
5
0
-5
-10
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

$$T^{1/2}\begin{bmatrix} \left(\bar{x}_{T+1:T+h} - \bar{x}_{1:T}\right) \\ X_T \end{bmatrix} \Rightarrow Y = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} \sim N(0,\Sigma)$$

Prediction: $f(Y_1 \mid Y_2)$



T = 70 years; h = 35; h/T = 0.5

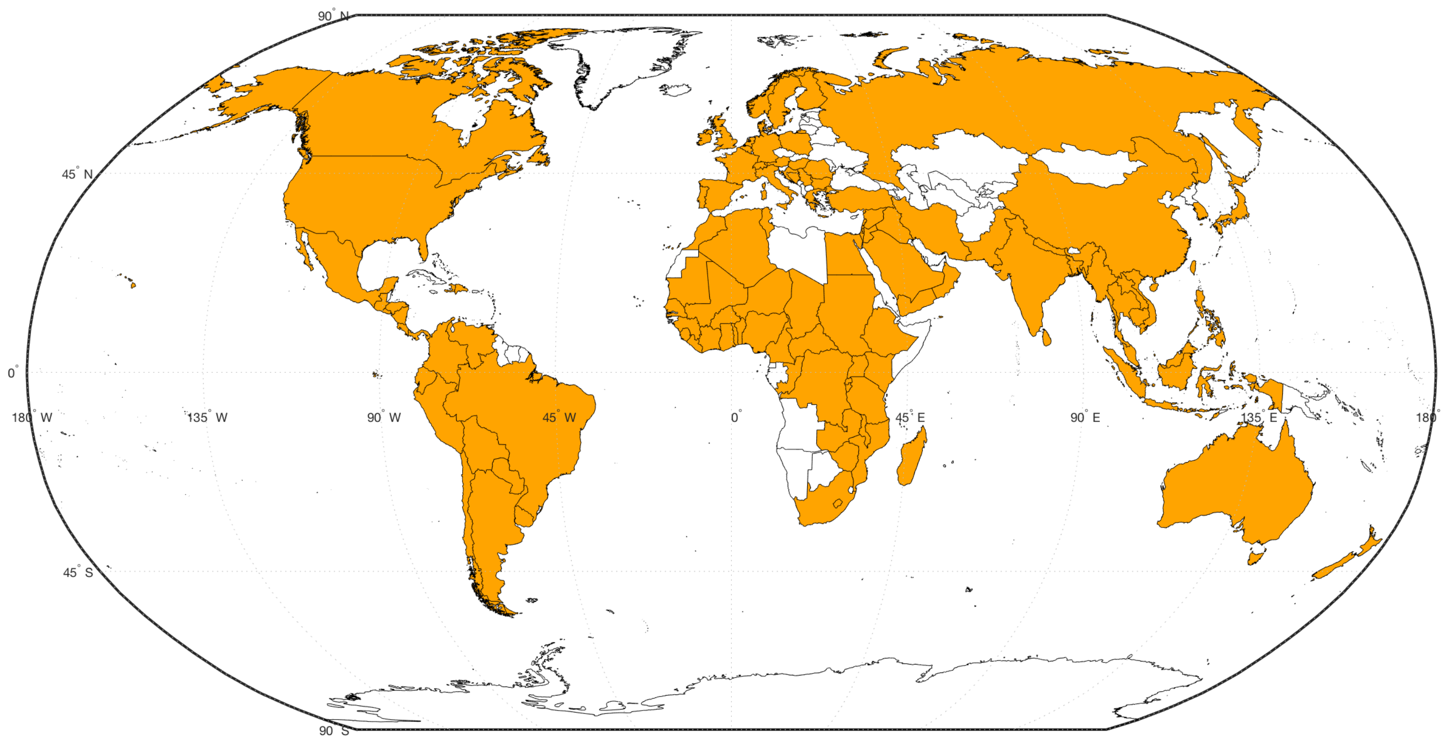
$\bar{x}_{1:T}$ = 1.24; s^2 = 14.8; HAR-SE = $s/T^{1/2}$ = 0.23; q (= df) = 14

I(0) prediction interval: $\bar{x}_{1:T} \pm t_{q,1-\alpha/2}\sqrt{s^2/T}\times\sqrt{1+r^{-1}}$; 68% interval: 0.84 to 1.64

LLM prediction interval: 0.2 to 1.6

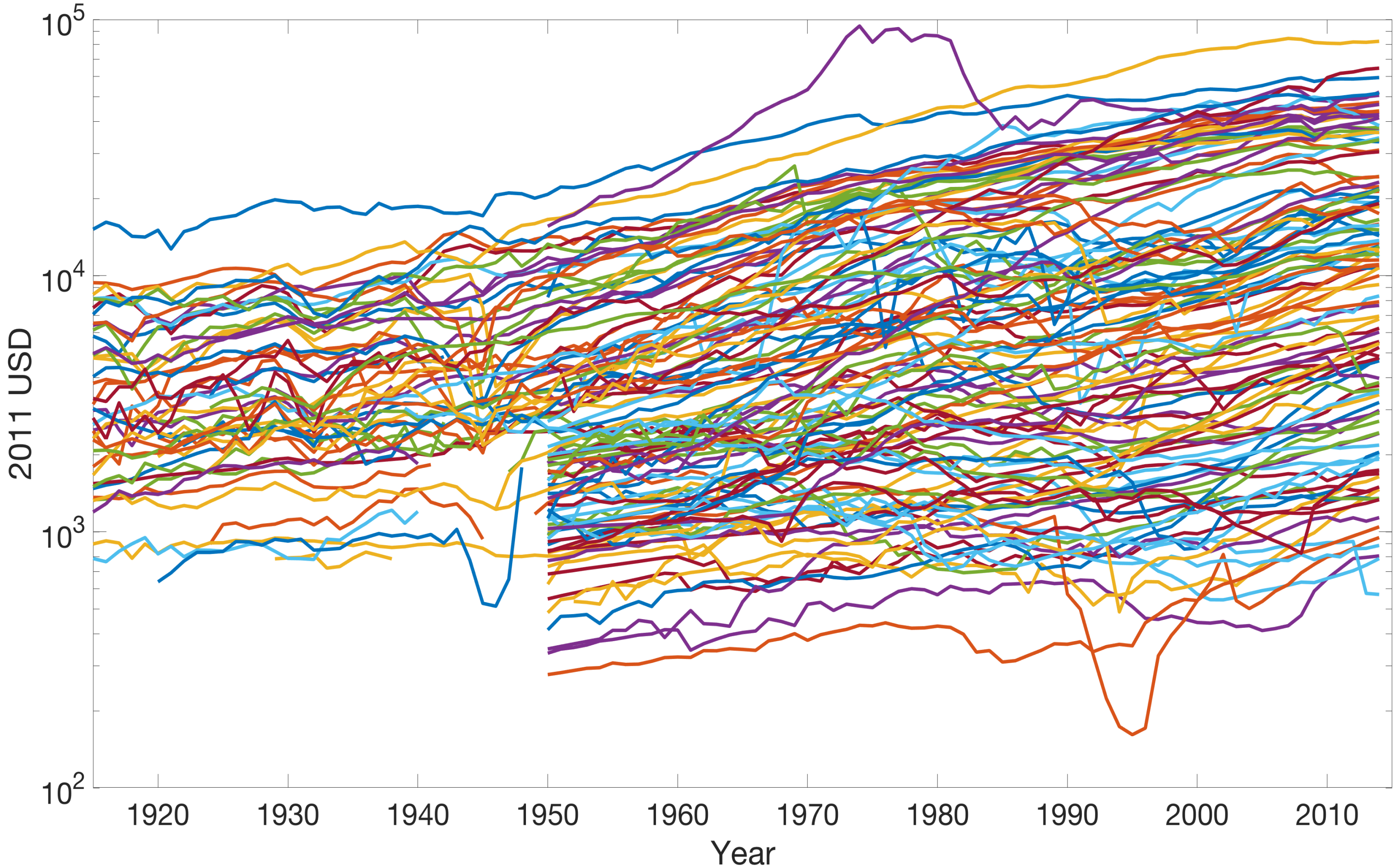
A more ambitious prediction exercise: Annual Data 1915-2014 for 112 countries

(Merged: PWT 1950-2014 and Maddison 1915-1949
countries with at least 50 years of post-1949 data and population > 3 million)

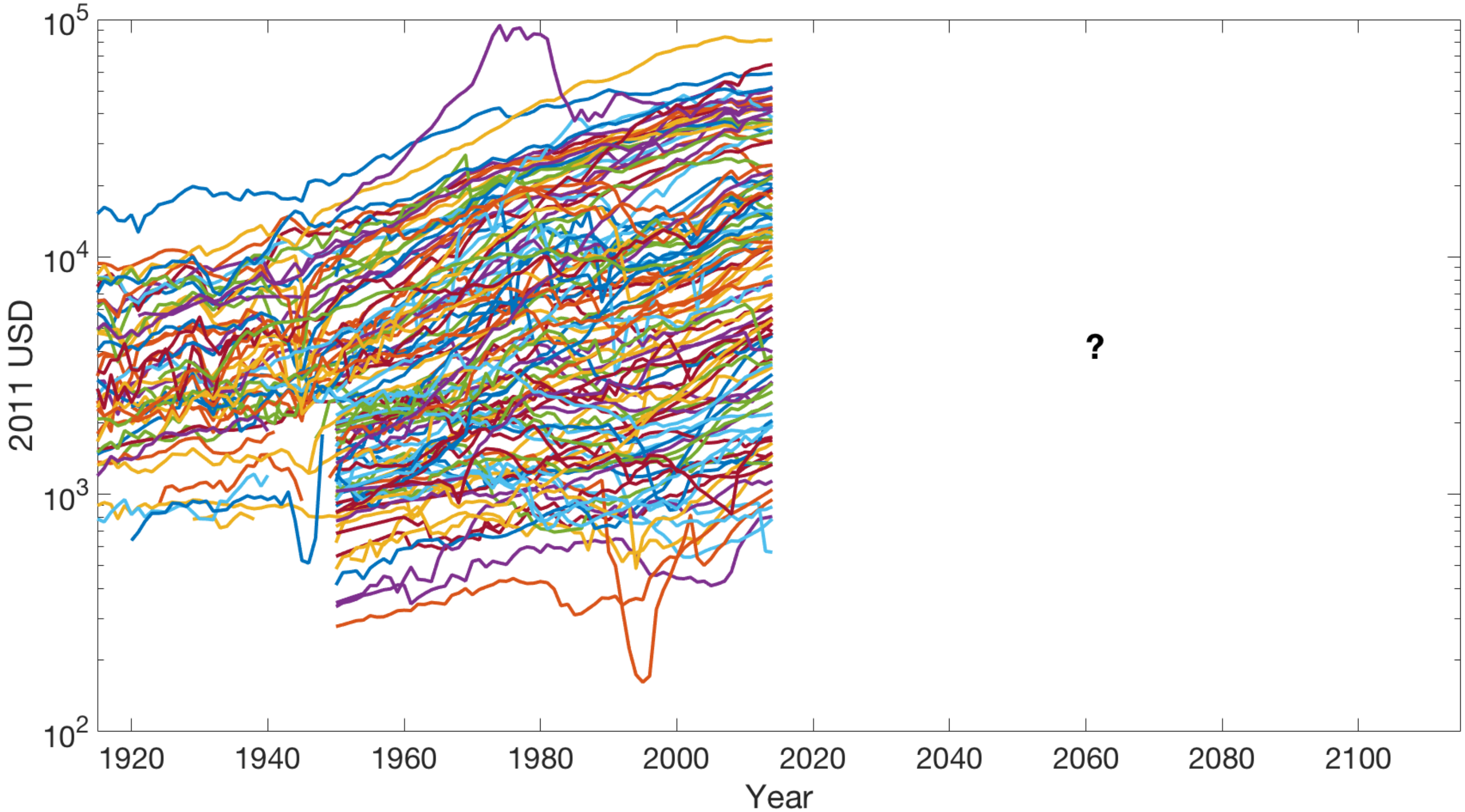


- 97% of World GDP in 2014 and 96% of World Population
- Unbalanced Panel (39-52 countries before 1950, 107 in 1950, 110 in 1952 and 112 in 1960)

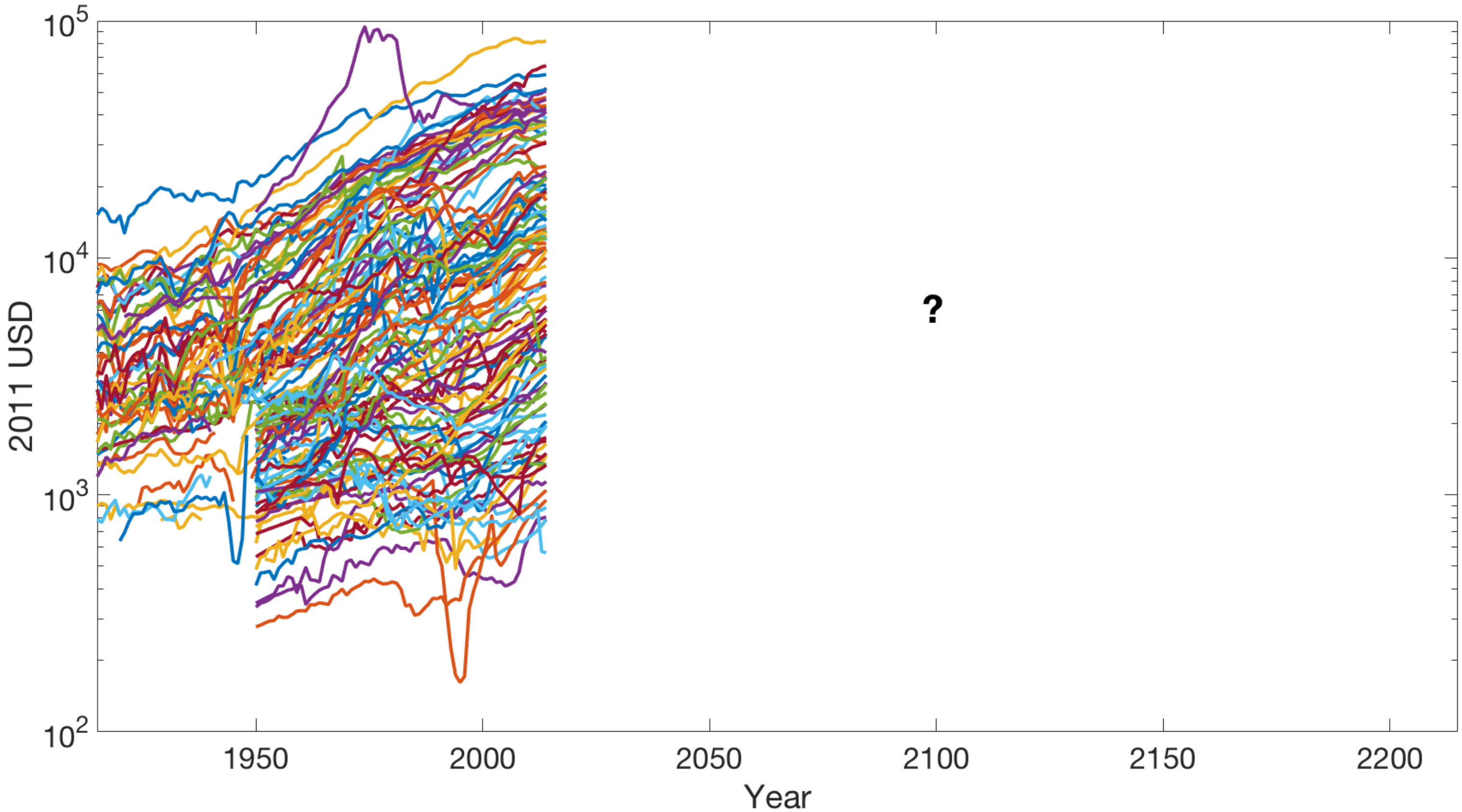
Data: GDP/Population for 112 countries



Long-Run Forecasting Problem



or



Course Topics

1. Time series refresher (MW)
2. Heteroskedasticity and autocorrelation consistent/robust (HAC, HAR) standard errors (JS)
3. Dynamic causal effects (JS)
4. Weak instruments/weak identification in IV and GMM (JS)
5. Dynamic factor models and prediction with large datasets (MW)
6. Low-frequency analysis of economic time series (MW)

References

Anderson, T. W. (1984). *An introduction to multivariate statistics.* Wiley, New York.

Baillie, R. T. (1996). "Long Memory Processes and, Frational Integration in Econometrics," *Journal of Econometrics*, 73, 5—59.

Cavanagh, C. (1985), "Roots Local to Unity," manuscript, Department of Economics, Harvard University.

Chan, H.H. and C.Z. Wei (1987), "Asymptotic Inference for Nearly Non-Stationary AR(1) Processes," *Annals of Statistics*, 15, 1050-1063.

Dickey, D.A.. and W.A. Fuller (1979). " "Distribution of the Estimators for Autoregressive Time Series with a Unit Root" *Journal of the American Statistical Association*. **74** (366): 427–431.

Elliott, G., T. J. Rothenberg, and J. H. Stock (1996). "Efficient Tests for an Autoregressive Unit Root," *Econometrica*, 64, 813—836.

Elliott, G. (1998). "The Robustness of Cointegration Methods When Regressors Almost Have Unit Roots," *Econometrica*, 66, 149—158.

Engle, R. F. (1974). "Band spectrum regression," *International Economic Review*, pp. 1—11.

Engle, Robert F. and Clive W.J. Granger (1987), "Co-Integration and Error-Correction: Representation, Estimation and Testing," *Econometrica*, 55, pp. 251-276.

Geweke, J. (2005). *Contemporary Bayesian Econometrics and Statstics*, Wiley.

Granger, C. W. J., AND P. Newbold (1974): "Spurious Regressions in Econometrics," *Journal of Econometrics*, 2, 111–120.

Grenander,U., and M. Rosenblatt (1957). *Statistical Analysis of Stationary Time Series*, NewYork:Wiley.

Hannan, E.J. (1963). "Regression for Time Series," in M. Rosenblatt ed., Time Series Analysis, New York: John Wiley, pp. 14-37.

Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter.* Cambridge University Press.

Johansen, Soren. (1988), "Statistical Analysis of Cointegration Vectors," *Journal of Economic Dynamics and Control*, 12, pp. 231-254.

Koop, G. (2003). *Bayesian Econometrics*, Wiley

Müller, U. K. (2004). "A Theory of Robust Long-Run Variance Estimation," Working paper, Princeton University.

Müller, U. K. and M.W. Watson (2016). "Measuring Uncertainty about Long-Run Predictions," *Review of Economic Studies,* 83 (4), October 2016, pp. 1711-1740.

Müller, U. K. and M.W. Watson (2017). "Low-Frequency Econometrics," in *Advances in Economics and Econometrics Vol. II,* B. Honore, A. Pakes, M. Piazzesi, and L. Samuelson (eds). Cambridge University Press, 2017.

Müller, U. K., J.H. Stock and M.W. Watson (2019). "International Long-run Growth Dynamics," manuscript, in progress.

Phillips, P. C. B. (1986). "Understanding Spurious Regression in Econometrics," *Journal of Econometrics*, 33, pp. 311-340.

Phillips, P. C. B. (1998). "New Tools for Understanding Spurious Regressions," *Econometrica*, Vol. 66m Bo. 6m pp. 1299-1325.

Robinson, P. M. (2003). "Long-Memory Time Series," in *Time Series with Long Memory*, ed. by P. M. Robinson, pp. 4—32. Oxford University Press, Oxford.

Stock, J. H. (1994). "Unit Roots, Structural Breaks and Trends," in *Handbook of Econometrics*, ed. by R. F. Engle, and D. McFadden, vol. 4, pp. 2740—2841. North Holland, New York.

Yule, G.U. (1926). "Why Do We Sometimes Get Nonsense Correlations Between Time Series? A Study in Sampling and the Nature of Time Series," *Journal of the Royal Statistical Society*, 89, pp. 1-69.