Boosting the Hodrick-Prescott Filter

Peter C. B. Phillips and Zhentao Shi

Yale-Auckland-SMU-Soton and CUHK

NBER-NSF Time Series Conference, August 2019

HP Filter

- History: .... Whittaker (1923), Henderson (1924), Aitken (1925), Leser(1961), Hodrick & Prescott (1980/1997)
- Asymptotics: Phillips & Jin (2002/2016)
- Critiques: .... Cogley & Nelson (1997), Hamilton (2018)
- Computation: Sakarya & de Jong (2016, 2017), Cornea-Madeira (2017)

- The trend: "most probable" trend given smoothness-prior condition

$$\left(\widehat{f}_t^{\mathrm{HP}}\right) = \arg\min_{(f_t)} \left\{ \sum_{t=1}^{n} (x_t - f_t)^2 + \lambda \sum_{t=2}^{n} \left(\Delta^2 f_t\right)^2 \right\}$$

- The residual ("cyclical" component)

$$\left(\widehat{c}_t^{\mathrm{HP}}\right) = \left(x_t - \widehat{f}_t^{\mathrm{HP}}\right)$$

HP Filter

- History: .... Whittaker (1923), Henderson (1924), Aitken (1925), Leser(1961), Hodrick & Prescott (1980/1997)
- Asymptotics: Phillips & Jin (2002/2016)
- Critiques: .... Cogley & Nelson (1997), Hamilton (2018)
- Computation: Sakarya & de Jong (2016, 2017), Cornea-Madeira (2017)

- The trend: "most probable" trend given smoothness-prior condition

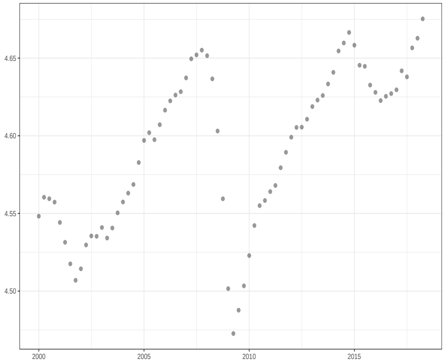
$$\left(\widehat{f}_t^{\mathrm{HP}}\right) = \arg\min_{(f_t)} \left\{ \sum_{t=1}^{n} (x_t - f_t)^2 + \lambda \sum_{t=2}^{n} \left(\Delta^2 f_t\right)^2 \right\}$$

- The residual ("cyclical" component)

$$\left(\widehat{c}_t^{\mathrm{HP}}\right) = \left(x_t - \widehat{f}_t^{\mathrm{HP}}\right)$$

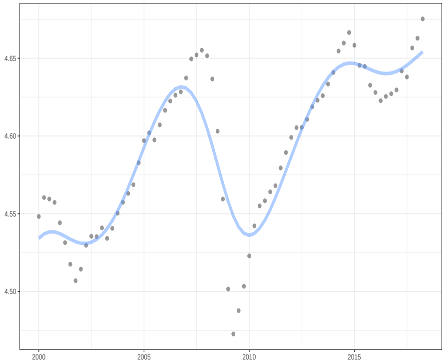
US Industrial Production Index: 21st Century

Quarterly Industrial Production data 2000:Q1 - 2018:Q4



US Industrial Production: 21st Century

Quarterly IP data with fitted HP filter 2000:Q1 - 2018:Q4



US Industrial Production: 21st Century

Quarterly IP data with fitted HP filter & AR(4) 2000:Q1 - 2018:Q4

What do we learn from HP filtering and AR regression?

- HP smooths data and captures downturns and upward directions
- AR(4) reproduces the data with one period delay and near mds error
- Do we distinguish trend and cycle by these methods? How do we define trend and cycle?

US Industrial Production: 21st Century

Quarterly IP data with fitted HP filter & AR(4) 2000:Q1 - 2018:Q4

What do we learn from HP filtering and AR regression?

- HP smooths data and captures downturns and upward directions
- AR(4) reproduces the data with one period delay and near mds error
- Do we distinguish trend and cycle by these methods? How do we define trend and cycle?

Tuning Parameter of the HP Filter

- **Gold standard:** $\lambda = 1600$ for quarterly data
  - ubiquitous usage with fixed λ setting
  - independent of sample size
  - in contrast to ....
- Nonparametric function estimation
  - employs sample size dependent tuning parameters
  - to achieve consistent estimation & remove bias
- Is this possible for the HP filter?
  - the HP filter is a nonparametric trend function estimator
  - asymptotic theory is needed for tuning parameter guidance
  - what choice of λ assures consistent estimation?
  - And for what types of trend?

go ML

Tuning Parameter of the HP Filter

- Gold standard: $\lambda = 1600$ for quarterly data
  - ubiquitous usage with fixed λ setting
  - independent of sample size
  - in contrast to ....
- Nonparametric function estimation
  - employs sample size dependent tuning parameters
  - to achieve consistent estimation & remove bias
- Is this possible for the HP filter?
  - the HP filter is a nonparametric trend function estimator
  - asymptotic theory is needed for tuning parameter guidance
  - what choice of λ assures consistent estimation?
  - And for what types of trend?

go ML

Tuning Parameter of the HP Filter

- Gold standard: $\lambda = 1600$ for quarterly data
  - ubiquitous usage with fixed λ setting
  - independent of sample size
  - in contrast to ....
- Nonparametric function estimation
  - employs sample size dependent tuning parameters
  - to achieve consistent estimation & remove bias
- Is this possible for the HP filter?
  - the HP filter is a nonparametric trend function estimator
  - asymptotic theory is needed for tuning parameter guidance
  - what choice of λ assures consistent estimation?
  - And for what types of trend?

go ML

Machine Learning via Boosting

Boosting algorithms: use a weak base learner to iterate and improve

- statistical iteration - twicing (Tukey, 1977)
- L_2 boosting in regression – Buhlmann & Yu (2003)
- economic applications:
  - twicing kernel – Newey et al (2004)
  - forecasting – Ng (2014)

Boosting the HP filter:

- use the HP filter as the weak base from which to learn about trend
- adaptively boost the HP filter by iterating on the residuals

Machine Learning via Boosting

Boosting algorithms: use a weak base learner to iterate and improve

- statistical iteration - twicing (Tukey, 1977)
- L_2 boosting in regression – Buhlmann & Yu (2003)
- economic applications:
  - twicing kernel – Newey et al (2004)
  - forecasting – Ng (2014)

Boosting the HP filter:

- use the HP filter as the weak base from which to learn about trend
- adaptively boost the HP filter by iterating on the residuals

Machine Learning via Boosting

Boosting algorithms: use a weak base learner to iterate and improve

- statistical iteration - twicing (Tukey, 1977)
- L_2 boosting in regression – Buhlmann & Yu (2003)
- economic applications:
  - twicing kernel – Newey et al (2004)
  - forecasting – Ng (2014)

Boosting the HP filter:

- use the HP filter as the weak base from which to learn about trend
- adaptively boost the HP filter by iterating on the residuals

The HP Trend and Cycle Operators

- The HP fitted trend is

$$\widehat{f}^{\text{HP}} = Sx = \left(I_n + \lambda DD'\right)^{-1} x$$

where $D = D_2$ is the HP second differencing matrix.

- The 'cycle' residual is

$$\widehat{c}^{\text{HP}} = x - \widehat{f}^{\text{HP}} = (I_n - S)\, x$$

- Now boost the filter by repeated operation

The HP Trend and Cycle Operators

- The HP fitted trend is

$$\widehat{f}^{\mathrm{HP}} = Sx = \left(I_n + \lambda DD'\right)^{-1} x$$

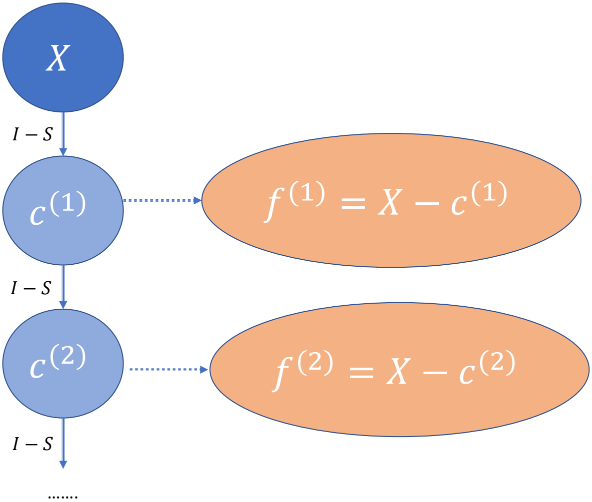
  where $D = D_2$ is the HP second differencing matrix.

- The 'cycle' residual is

$$\widehat{c}^{\mathrm{HP}} = x - \widehat{f}^{\mathrm{HP}} = (I_n - S)\, x$$

- Now boost the filter by repeated operation

The HP Boosting Algorithm



Analytic Form of the Boosted Filter

- Repeat the HP filter on the residuals

$$\begin{aligned}\widehat{c}^{(2)} &= (I_n - S)\,\widehat{c}^{\mathrm{HP}} = (I_n - S)^2\,x \\ \widehat{f}^{(2)} &= x - \widehat{c}^{(2)} = \left(I_n - (I_n - S)^2\right)x\end{aligned}$$

- Iterate m times

$$\begin{aligned}\widehat{c}^{(m)} &= (I_n - S)^m\,x \\ \widehat{f}^{(m)} &= x - \widehat{c}^{(m)} = B_m x\end{aligned}$$

where $B_m = I_n - (I_n - S)^m$

- the base learner operator is

$$S = \left(I_n + \lambda DD'\right)^{-1}$$

- the boost operator is

$$B_m(S) = \sum_{j=0}^{m-1} S(I-S)^j = (I - (I-S)^m)$$

go proof

Analytic Form of the Boosted Filter

- Repeat the HP filter on the residuals

$$\begin{aligned} \widehat{c}^{(2)} &= (I_n - S)\,\widehat{c}^{\mathrm{HP}} = (I_n - S)^2\,x \\ \widehat{f}^{(2)} &= x - \widehat{c}^{(2)} = \left(I_n - (I_n - S)^2\right)x \end{aligned}$$

- Iterate m times

$$\begin{aligned} \widehat{c}^{(m)} &= (I_n - S)^m\,x \\ \widehat{f}^{(m)} &= x - \widehat{c}^{(m)} = B_m x \end{aligned}$$

where $B_m = I_n - (I_n - S)^m$

- the base learner operator is

$$S = \left(I_n + \lambda DD'\right)^{-1}$$

- the boost operator is

$$B_m(S) = \sum_{j=0}^{m-1} S(I-S)^j = (I - (I-S)^m)$$

go proof

Analytic Form of the Boosted Filter

- Repeat the HP filter on the residuals

$$\begin{aligned}\widehat{c}^{(2)} &= (I_n - S)\,\widehat{c}^{\mathrm{HP}} = (I_n - S)^2\, x \\ \widehat{f}^{(2)} &= x - \widehat{c}^{(2)} = \left(I_n - (I_n - S)^2\right) x\end{aligned}$$

- Iterate m times

$$\begin{aligned}\widehat{c}^{(m)} &= (I_n - S)^m\, x \\ \widehat{f}^{(m)} &= x - \widehat{c}^{(m)} = B_m x\end{aligned}$$

where $B_m = I_n - (I_n - S)^m$

- the base learner operator is

$$S = \left(I_n + \lambda D D'\right)^{-1}$$

- the boost operator is

$$B_m(S) = \sum_{j=0}^{m-1} S(I - S)^j = (I - (I - S)^m)$$

go proof

Analytic Form of the Boosted Filter

- Repeat the HP filter on the residuals

$$
\begin{aligned}
\widehat{c}^{(2)} &= (I_n - S)\,\widehat{c}^{\text{HP}} = (I_n - S)^2\,x \\
\widehat{f}^{(2)} &= x - \widehat{c}^{(2)} = \left(I_n - (I_n - S)^2\right)x
\end{aligned}
$$

- Iterate m times

$$
\begin{aligned}
\widehat{c}^{(m)} &= (I_n - S)^m\,x \\
\widehat{f}^{(m)} &= x - \widehat{c}^{(m)} = B_m x
\end{aligned}
$$

where $B_m = I_n - (I_n - S)^m$

- the base learner operator is

$$S = \left(I_n + \lambda DD'\right)^{-1}$$

- the boost operator is

$$B_m(S) = \sum_{j=0}^{m-1} S(I - S)^j = (I - (I - S)^m)$$

go proof

Connection to L_2 Boosting

- Buhlmann and Yu (2003) – regression and discrete classification
- Boosting a base learner in ridge regression: $\widehat{y} = S^{\text{ridge}} y$
- the base learner operator is

$$S^{\text{ridge}} = Z(Z'Z + \lambda I_p)^{-1} Z'$$

- the boost operator is

$$B_m(S^{\text{ridge}}) = \sum_{j=0}^{m} S^{\text{ridge}} (I - S^{\text{ridge}})^j = (I - (I - S^{\text{ridge}})^{m+1})$$

Assumes iid data and shows MSE improvement from boosting

Connection to L_2 Boosting

- Buhlmann and Yu (2003) – regression and discrete classification
- Boosting a base learner in ridge regression: $\widehat{y} = S^{\text{ridge}} y$
- the base learner operator is

$$S^{\text{ridge}} = Z(Z'Z + \lambda I_p)^{-1} Z'$$

- the boost operator is

$$B_m(S^{\text{ridge}}) = \sum_{j=0}^{m} S^{\text{ridge}} (I - S^{\text{ridge}})^j = (I - (I - S^{\text{ridge}})^{m+1})$$

Assumes iid data and shows MSE improvement from boosting

Contribution of this Paper

Boosting algorithm for trend determination

- Apply refitting to nonstationary time series
- Allow for stochastic and deterministic trends and trend breaks
- Easy-to-implement procedure

Limit theory for boosted HP filter

- gives consistent estimation of stochastic and deterministic trends
- gives a consistent estimates of trend break points

Numerics and empirics

- international analysis of Okun's law
- international business cycles
- historical trend determination for US industrial production

Contribution of this Paper

Boosting algorithm for trend determination

- Apply refitting to nonstationary time series
- Allow for stochastic and deterministic trends and trend breaks
- Easy-to-implement procedure

Limit theory for boosted HP filter

- gives consistent estimation of stochastic and deterministic trends
- gives a consistent estimates of trend break points

Numerics and empirics

- international analysis of Okun's law
- international business cycles
- historical trend determination for US industrial production

Contribution of this Paper

Boosting algorithm for trend determination

- Apply refitting to nonstationary time series
- Allow for stochastic and deterministic trends and trend breaks
- Easy-to-implement procedure

Limit theory for boosted HP filter

- gives consistent estimation of stochastic and deterministic trends
- gives a consistent estimates of trend break points

Numerics and empirics

- international analysis of Okun's law
- international business cycles
- historical trend determination for US industrial production

Stochastic Trend Determination

- Suppose x_t follows a stochastic trend and $\frac{x_{\lfloor nr \rfloor}}{\sqrt{n}} \rightsquigarrow B(r)$
- Applying the HP filter to a trajectory of x_t leads to a smooth curve
- but gives an inconsistent estimate of the trend unless λ is very small ($\lambda = o(n)$ as $n \to \infty$)

Form of the Brownian motion limit process

- The Brownian motion limit function has Karhunen-Loève Representation

$$B(r) = \sum_{k=1}^{\infty} \sqrt{\lambda_k} \varphi_k(r) \xi_k = \sqrt{2} \sum_{k=1}^{\infty} \frac{\sin\left[(k - \frac{1}{2})\pi r\right]}{\left(k - \frac{1}{2}\right)\pi} \xi_k \equiv BM(\omega^2)$$

- where $\xi_k \sim \text{i.i.d. } N\left(0, \omega^2\right)$, $\sqrt{\lambda_k} = \frac{1}{(k - \frac{1}{2})\pi}$ and $\varphi_k(r) = \sqrt{2} \sin\left[(k - \frac{1}{2})\pi r\right]$

The limit form of the HP filter has a related, but smooth, KL representation

Stochastic Trend Determination

- Suppose x_t follows a stochastic trend and $\frac{x_{\lfloor nr \rfloor}}{\sqrt{n}} \rightsquigarrow B(r)$
- Applying the HP filter to a trajectory of x_t leads to a smooth curve
- but gives an inconsistent estimate of the trend unless λ is very small ($\lambda = o(n)$ as $n \to \infty$)

Form of the Brownian motion limit process

- The Brownian motion limit function has Karhunen-Loève Representation

$$B(r) = \sum_{k=1}^{\infty} \sqrt{\lambda_k} \varphi_k(r) \xi_k = \sqrt{2} \sum_{k=1}^{\infty} \frac{\sin\left[(k - \frac{1}{2})\pi r\right]}{\left(k - \frac{1}{2}\right)\pi} \xi_k \equiv BM(\omega^2)$$

- where $\xi_k \sim \text{i.i.d. } N\left(0, \omega^2\right)$, $\sqrt{\lambda_k} = \frac{1}{(k - \frac{1}{2})\pi}$ and $\varphi_k(r) = \sqrt{2} \sin\left[(k - \frac{1}{2})\pi r\right]$

The limit form of the HP filter has a related, but smooth, KL representation

Stochastic Trend Determination

- Suppose x_t follows a stochastic trend and $\frac{x_{\lfloor nr \rfloor}}{\sqrt{n}} \rightsquigarrow B(r)$
- Applying the HP filter to a trajectory of x_t leads to a smooth curve
- but gives an inconsistent estimate of the trend unless λ is very small ($\lambda = o(n)$ as $n \to \infty$)

Form of the Brownian motion limit process

- The Brownian motion limit function has Karhunen-Loève Representation

$$B(r) = \sum_{k=1}^{\infty} \sqrt{\lambda_k} \varphi_k(r) \xi_k = \sqrt{2} \sum_{k=1}^{\infty} \frac{\sin\left[(k - \frac{1}{2})\pi r\right]}{\left(k - \frac{1}{2}\right)\pi} \xi_k \equiv BM(\omega^2)$$

- where $\xi_k \sim \text{i.i.d. } N\left(0, \omega^2\right)$, $\sqrt{\lambda_k} = \frac{1}{(k - \frac{1}{2})\pi}$ and $\varphi_k(r) = \sqrt{2} \sin\left[(k - \frac{1}{2})\pi r\right]$

The limit form of the HP filter has a related, but smooth, KL representation

HP limit theory for unit root stochastic trend data

Theorem – Phillips & Jin, 2015

If $\frac{x_{\lfloor nr \rfloor}}{\sqrt{n}} \rightsquigarrow B(r)$ and $\lambda = \mu n^4$

$$\frac{\widehat{f}^{HP}_{\lfloor nr \rfloor}}{\sqrt{n}} \rightsquigarrow f^{HP}(r) = \sum_{k=1}^{\infty} \frac{\lambda_k^2}{\mu + \lambda_k^2} \sqrt{\lambda_k} \varphi_k(r) \xi_k,$$

Implications

- $\frac{\widehat{f}^{HP}_{\lfloor nr \rfloor}}{\sqrt{n}} \rightsquigarrow f^{HP}(r) \neq B(r)$ and $\frac{\lambda_k^2}{\mu+\lambda_k^2} = \frac{1}{1+\frac{\mu}{\lambda_k^2}} = O(k^{-4})$
- Limit process $f^{HP}(r) \in C^4$ is a smooth function
- The smooth HP limit process implies predictability
- Similar results hold for diffusion limit processes
- Setting $\lambda = \mu n^4$ matches empirical results well for $\lambda = 1600$

Can boosting improve this performance of the HP filter?

HP limit theory for unit root stochastic trend data

Theorem – Phillips & Jin, 2015

If $\frac{x_{\lfloor nr \rfloor}}{\sqrt{n}} \rightsquigarrow B(r)$ and $\lambda = \mu n^4$

$$\frac{\widehat{f}^{\mathrm{HP}}_{\lfloor nr \rfloor}}{\sqrt{n}} \rightsquigarrow f^{\mathrm{HP}}(r) = \sum_{k=1}^{\infty} \frac{\lambda_k^2}{\mu + \lambda_k^2} \sqrt{\lambda_k} \varphi_k(r) \xi_k,$$

Implications

- $\frac{\widehat{f}^{\mathrm{HP}}_{\lfloor nr \rfloor}}{\sqrt{n}} \rightsquigarrow f^{\mathrm{HP}}(r) \neq B(r)$ and $\frac{\lambda_k^2}{\mu+\lambda_k^2} = \frac{1}{1+\frac{\mu}{\lambda_k^2}} = O(k^{-4})$
- Limit process $f^{\mathrm{HP}}(r) \in C^4$ is a smooth function
- The smooth HP limit process implies predictability
- Similar results hold for diffusion limit processes
- Setting $\lambda = \mu n^4$ matches empirical results well for $\lambda = 1600$

Can boosting improve this performance of the HP filter?

HP limit theory for unit root stochastic trend data

Theorem – Phillips & Jin, 2015

If $\frac{x_{\lfloor nr \rfloor}}{\sqrt{n}} \rightsquigarrow B(r)$ and $\lambda = \mu n^4$

$$\frac{\widehat{f}^{\mathrm{HP}}_{\lfloor nr \rfloor}}{\sqrt{n}} \rightsquigarrow f^{\mathrm{HP}}(r) = \sum_{k=1}^{\infty} \frac{\lambda_k^2}{\mu + \lambda_k^2} \sqrt{\lambda_k} \varphi_k(r) \xi_k,$$

Implications

- $\frac{\widehat{f}^{\mathrm{HP}}_{\lfloor nr \rfloor}}{\sqrt{n}} \rightsquigarrow f^{\mathrm{HP}}(r) \neq B(r)$ and $\frac{\lambda_k^2}{\mu+\lambda_k^2} = \frac{1}{1+\frac{\mu}{\lambda_k^2}} = O(k^{-4})$
- Limit process $f^{\mathrm{HP}}(r) \in C^4$ is a smooth function
- The smooth HP limit process implies predictability
- Similar results hold for diffusion limit processes
- Setting $\lambda = \mu n^4$ matches empirical results well for $\lambda = 1600$

Can boosting improve this performance of the HP filter?

Boosted HP Filter Limit Theory for Stochastic Trends

Theorem 1

Suppose that x_t satisfies

$$\frac{x_{\lfloor nr \rfloor}}{\sqrt{n}} \rightsquigarrow B(r)$$

and the HP filter with $\lambda = \mu n^4$ is iterated m times according to the boosted HP algorithm. If $m \to \infty$ as $n \to \infty$ then

$$\frac{\widehat{f}^{(m)}_{\lfloor nr \rfloor}}{\sqrt{n}} \rightsquigarrow B(r).$$

Practical Implications

- Boosting smartens the HP filter with standard setting $\lambda = \mu n^4 \sim 1600$, making the boosted filter consistent for a stochastic trend
- Implementation requires only peforming the iterative algorithm after fitting the HP filter with the standard setting
- Remaining issue - choosing a stopping time and selecting m

go DST

Boosted HP Filter Limit Theory for Stochastic Trends

Theorem 1

Suppose that x_t satisfies

$$\frac{x_{\lfloor nr \rfloor}}{\sqrt{n}} \rightsquigarrow B(r)$$

and the HP filter with $\lambda = \mu n^4$ is iterated m times according to the boosted HP algorithm. If $m \to \infty$ as $n \to \infty$ then

$$\frac{\widehat{f}^{(m)}_{\lfloor nr \rfloor}}{\sqrt{n}} \rightsquigarrow B(r).$$

Practical Implications

- Boosting smartens the HP filter with standard setting $\lambda = \mu n^4 \sim 1600$, making the boosted filter consistent for a stochastic trend
- Implementation requires only peforming the iterative algorithm after fitting the HP filter with the standard setting
- Remaining issue - choosing a stopping time and selecting m

Heuristic Outline

- Asymptotic form of HP operators (ignoring end corrections)

$$\widehat{f}^{\mathrm{HP}} = Sx = (I_n + \lambda DD')^{-1} x \approx G_\lambda x := \left(\frac{1}{\lambda L^{-2} (1-L)^4 + 1} \right) x$$

$$\widehat{c}^{\mathrm{HP}} \approx (1 - G_\lambda) x = \left(\frac{\lambda L^{-2} (1-L)^4}{\lambda L^{-2} (1-L)^4 + 1} \right) x$$

- Induced effect of the operator $1 - G_\lambda$ on the ON components is

$$(1 - G_\lambda) \varphi_k \left(\frac{t}{n} \right) \approx \frac{\mu}{\mu + \lambda_k^2} \varphi_k \left(\frac{t}{n} \right)$$

- Boosting with m iterations induces the operator $(1 - G_\lambda)^m$ and

$$(1 - G_\lambda)^m \varphi_k \left(\frac{t}{n} \right) \approx \left(\frac{\mu}{\mu + \lambda_k^2} \right)^m \varphi_k \left(\frac{t}{n} \right) \to 0, \quad \text{as } m \to \infty$$

proof

Heuristic Outline

- Asymptotic form of HP operators (ignoring end corrections)

$$\widehat{f}^{\mathrm{HP}} = Sx = (I_n + \lambda DD')^{-1} x \approx G_\lambda x := \left(\frac{1}{\lambda L^{-2} (1-L)^4 + 1} \right) x$$

$$\widehat{c}^{\mathrm{HP}} \approx (1 - G_\lambda)\, x = \left(\frac{\lambda L^{-2} (1-L)^4}{\lambda L^{-2} (1-L)^4 + 1} \right) x$$

- Induced effect of the operator $1 - G_\lambda$ on the ON components is

$$(1 - G_\lambda)\, \varphi_k \left(\frac{t}{n} \right) \approx \frac{\mu}{\mu + \lambda_k^2} \varphi_k \left(\frac{t}{n} \right)$$

- Boosting with m iterations induces the operator $(1 - G_\lambda)^m$ and

$$(1 - G_\lambda)^m\, \varphi_k \left(\frac{t}{n} \right) \approx \left(\frac{\mu}{\mu + \lambda_k^2} \right)^m \varphi_k \left(\frac{t}{n} \right) \to 0, \quad \text{as } m \to \infty$$

proof

Heuristic Outline

- Asymptotic form of HP operators (ignoring end corrections)

$$\widehat{f}^{\mathrm{HP}} = Sx = (I_n + \lambda DD')^{-1} x \approx G_\lambda x := \left(\frac{1}{\lambda L^{-2} (1-L)^4 + 1} \right) x$$

$$\widehat{c}^{\mathrm{HP}} \approx (1 - G_\lambda) x = \left(\frac{\lambda L^{-2} (1-L)^4}{\lambda L^{-2} (1-L)^4 + 1} \right) x$$

- Induced effect of the operator $1 - G_\lambda$ on the ON components is

$$(1 - G_\lambda) \varphi_k \left(\frac{t}{n} \right) \approx \frac{\mu}{\mu + \lambda_k^2} \varphi_k \left(\frac{t}{n} \right)$$

- Boosting with m iterations induces the operator $(1 - G_\lambda)^m$ and

$$(1 - G_\lambda)^m \varphi_k \left(\frac{t}{n} \right) \approx \left(\frac{\mu}{\mu + \lambda_k^2} \right)^m \varphi_k \left(\frac{t}{n} \right) \to 0, \quad \text{as } m \to \infty$$

proof

Heuristic Outline 2

- The induced boosted HP ON components on the trend are then

$$\left[1-(1-G_\lambda)^m\right]\varphi_k\left(\frac{t}{n}\right)\approx\varphi_k\left(\frac{t}{n}\right)$$

  for all $k\leq K_n$ as $(m,K_n,n)\to\infty$ with $K_n^4/m\to 0$

- leading to the following form of the boosted HP filter

$$\begin{aligned}\frac{\widehat{f}_t^{(m)}}{\sqrt{n}} &= \left[1-(1-G_\lambda)^m\right]\frac{x_t}{\sqrt{n}}\approx\left[1-(1-G_\lambda)^m\right]B^{K_n}\left(\frac{t}{n}\right)\\ &= \sum_{k=1}^{K_n}\sqrt{\lambda_k}\left[1-(1-G_\lambda)^m\right]\varphi_k\left(\frac{t}{n}\right)\xi_k\\ &= \sum_{k=1}^{K_n}\sqrt{\lambda_k}\varphi_k\left(\frac{t}{n}\right)\xi_k+o_{a.s.}(1)=B\left(\frac{t}{n}\right)+o_{a.s.}(1)\end{aligned}$$

- using the fact that, as $(K_n,n)\to\infty$ with $\frac{K_n}{n}\to 0$,

$$B^{K_n}(r):=\sum_{k=1}^{K_n}\sqrt{\lambda_k}\varphi_k\left(\frac{t}{n}\right)\xi_k=B(r)+o_{a.s.}(1)$$

proof

Deterministic and Stochastic Trends

Polynomial trend break function with stochastic trend

$$g_n(t) = \begin{cases} \alpha_n^0 + \beta_{n,1}^0 t + \ldots + \beta_{n,J}^0 t^J & t < \tau_0 = \lfloor nr_0 \rfloor \\ \alpha_n^1 + \beta_{n,1}^1 t + \ldots + \beta_{n,J}^1 t^J & t \geq \tau_0 = \lfloor nr_0 \rfloor \end{cases}$$

Limiting break function form

$$n^{-1/2} g_n(\lfloor nr \rfloor) \to g(r) = \begin{cases} \alpha^0 + \beta_1^0 r + \ldots + \beta_J^0 r^J & r < r_0 \\ \alpha^1 + \beta_1^1 r + \ldots + \beta_J^1 r^J & r \geq r_0 \end{cases},$$

DGP is $x_t = g_n(t) + x_t^0$ with stochastic trend x_t^0 and limit process

$$n^{-1/2} x_{t=\lfloor nr \rfloor} \rightsquigarrow g(r) + B(r) =: B_g(r)$$

- x_t^0 follows the functional law $\frac{x_{\lfloor nr \rfloor}^0}{\sqrt{n}} \rightsquigarrow B(r)$
- $g_n(t)$ is a nonrandom polynomial break function
- limit function $g(r)$ is piecewise smooth

Deterministic and Stochastic Trends

Polynomial trend break function with stochastic trend

$$g_n(t) = \begin{cases} \alpha_n^0 + \beta_{n,1}^0 t + \ldots + \beta_{n,J}^0 t^J & t < \tau_0 = \lfloor nr_0 \rfloor \\ \alpha_n^1 + \beta_{n,1}^1 t + \ldots + \beta_{n,J}^1 t^J & t \geq \tau_0 = \lfloor nr_0 \rfloor \end{cases}$$

Limiting break function form

$$n^{-1/2} g_n(\lfloor nr \rfloor) \to g(r) = \begin{cases} \alpha^0 + \beta_1^0 r + \ldots + \beta_J^0 r^J & r < r_0 \\ \alpha^1 + \beta_1^1 r + \ldots + \beta_J^1 r^J & r \geq r_0 \end{cases},$$

DGP is $x_t = g_n(t) + x_t^0$ with stochastic trend x_t^0 and limit process

$$n^{-1/2} x_{t=\lfloor nr \rfloor} \rightsquigarrow g(r) + B(r) =: B_g(r)$$

- x_t^0 follows the functional law $\frac{x_{\lfloor nr \rfloor}^0}{\sqrt{n}} \rightsquigarrow B(r)$
- $g_n(t)$ is a nonrandom polynomial break function
- limit function $g(r)$ is piecewise smooth

Boosted HP Filter Limit Theory for Deterministic & Stochastic Trends

Theorem 2

Let $x_t = g_n(t) + x_t^0$. Iterate the HP filter m times with $\lambda = \mu n^4$. If $(m, n) \to \infty$ with $m/n \to 0$

$$\frac{\widehat{f}_{\lfloor nr \rfloor}^{(m)}}{\sqrt{n}} \rightsquigarrow g^{\text{bHP}}(r) + B(r) := \begin{cases} g(r) + B(r), & r \neq r_0 \\ \frac{1}{2}\left\{g\left(r_0^-\right) + g\left(r_0^+\right)\right\} + B(r_0), & r = r_0 \end{cases}. \tag{1}$$

for each $r \in [0, 1]$ and $r_0 \in (0, 1)$

Implications

- Boosted HP filter is consistent for stochastic trends with polynomial drift
- Consistency holds for polynomial break function except at break point
- limit function $g^{bHP}(r)$ is piecewise smooth just like the true limit $g(r)$ function
- discontinuity in $g^{bHP}(r)$ at r^0 matches that of $g(r)$ function
- consistency for all $r \neq r^0$ ensures consistent estimation of break point r^0
- Similar consistency result holds for any piecewise smooth deterministic trend with a finite number of break points

Proof via Fourier representation

Boosted HP Filter Limit Theory for Deterministic & Stochastic Trends

Theorem 2

Let $x_t = g_n(t) + x_t^0$. Iterate the HP filter m times with $\lambda = \mu n^4$. If $(m, n) \to \infty$ with $m/n \to 0$

$$\frac{\widehat{f}_{\lfloor nr \rfloor}^{(m)}}{\sqrt{n}} \rightsquigarrow g^{\mathrm{bHP}}(r) + B(r) := \begin{cases} g(r) + B(r), & r \neq r_0 \\ \frac{1}{2}\left\{g\left(r_0^-\right) + g\left(r_0^+\right)\right\} + B(r_0), & r = r_0 \end{cases}. \quad (1)$$

for each $r \in [0, 1]$ and $r_0 \in (0, 1)$

Implications

- Boosted HP filter is consistent for stochastic trends with polynomial drift
- Consistency holds for polynomial break function except at break point
- limit function $g^{bHP}(r)$ is piecewise smooth just like the true limit $g(r)$ function
- discontinuity in $g^{bHP}(r)$ at r^0 matches that of $g(r)$ function
- consistency for all $r \neq r^0$ ensures consistent estimation of break point r^0
- Similar consistency result holds for any piecewise smooth deterministic trend with a finite number of break points

Proof via Fourier representation

Boosted HP Filter Limit Theory for Deterministic & Stochastic Trends

Theorem 2

Let $x_t = g_n(t) + x_t^0$. Iterate the HP filter m times with $\lambda = \mu n^4$. If $(m, n) \to \infty$ with $m/n \to 0$

$$\frac{\widehat{f}_{\lfloor nr \rfloor}^{(m)}}{\sqrt{n}} \rightsquigarrow g^{\mathrm{bHP}}(r) + B(r) := \begin{cases} g(r) + B(r), & r \neq r_0 \\ \frac{1}{2}\left\{g\left(r_0^-\right) + g\left(r_0^+\right)\right\} + B(r_0), & r = r_0 \end{cases}. \tag{1}$$

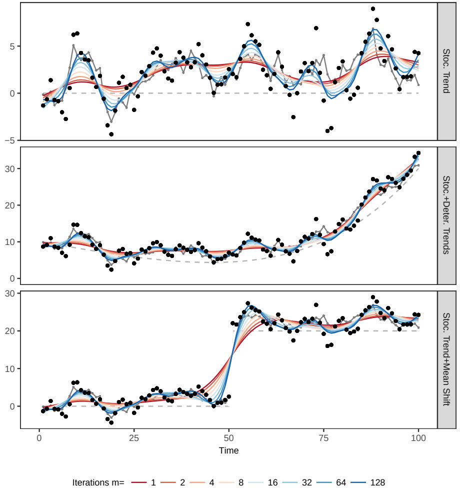
for each $r \in [0, 1]$ and $r_0 \in (0, 1)$

Implications

- Boosted HP filter is consistent for stochastic trends with polynomial drift
- Consistency holds for polynomial break function except at break point
- limit function $g^{bHP}(r)$ is piecewise smooth just like the true limit $g(r)$ function
- discontinuity in $g^{bHP}(r)$ at r^0 matches that of $g(r)$ function
- consistency for all $r \neq r^0$ ensures consistent estimation of break point r^0
- Similar consistency result holds for any piecewise smooth deterministic trend with a finite number of break points

Proof via Fourier representation

Illustration: fitting (breaking) stochastic trends with stationary errors



- Data (dots); determinist trend function (broken grey line); full trend (grey line)
- HP filter (red line); bHP filter (lines shaded orange to blue)

Computational Complexity

- Simple HP filter: involves search over λ
  - A grid system $(\lambda^{(1)}, \lambda^{(2)}, \ldots, \lambda^{(m)})$
  - Matrix inverse unless faster algorithms used
- Boosted HP filter: search for m given a big λ (a weak base learner)
  - Matrix-vector multiplication

Stopping Criteria

- Employ a unit root test
  - Augmented Dickey-Fuller test on residuals at each iteration
- Employ an information criterion
  - BIC-type criterion balancing fit against relative dimension

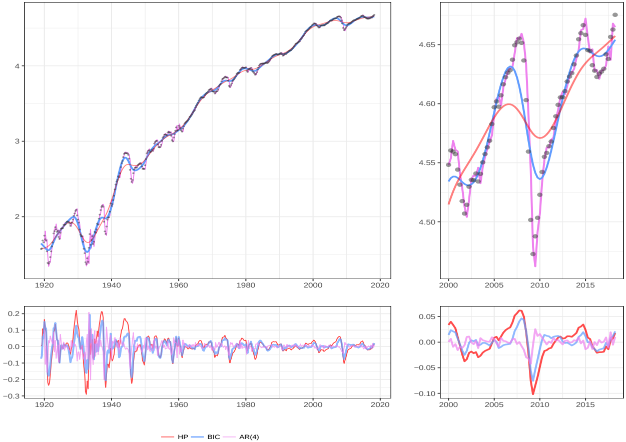
$$IC(m) = \widehat{c}^{(m)\prime}\widehat{c}^{(m)} + \log(n) \times \frac{\widehat{c}^{\mathrm{HP}\prime}\widehat{c}^{\mathrm{HP}}}{\mathrm{tr}(I_n - S)} \times \mathrm{tr}(B_m).$$

Empirical Applications

1. US Industrial Production Index
2. Okun's Law: panel of 40 time series
3. International business cycles: panel of 26 time series (small open economies)

US Industrial Production Index over 100 years

Comparisons of HP and bHP with fitted AR(4) (Hamilton, 2018)



- HP (red) - gentle flexible ruler smoothing, misses obvious trend lines
- bHP (blue) - discriminate smoothing, follows obvious trend lines
- AR(4) (magenta) - fits data closely with random residuals

go cycles sumup

Okun's Law

- Okun (1963): linear relationship between output gap and unemployment rate
- Empirical Regression

$$U_t - U_t^* = \beta \left(Y_t - Y_t^* \right) + \varepsilon_t$$

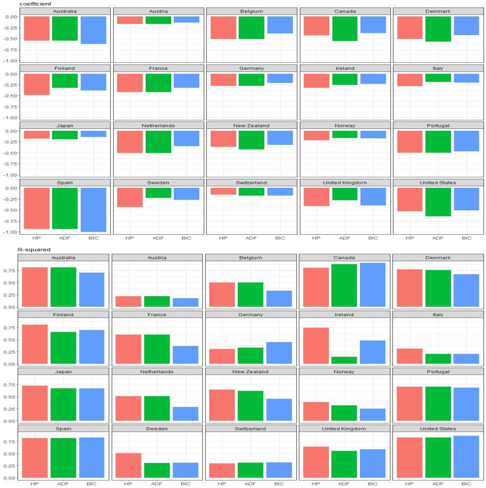
- Avoid spurious regression by focusing on business cycle
- Ball, Leigh, and Loungani (2017)
  - 20 OECD countries
  - Annual data 1980–2016.

Okun's Law

- Okun (1963): linear relationship between output gap and unemployment rate
- Empirical Regression

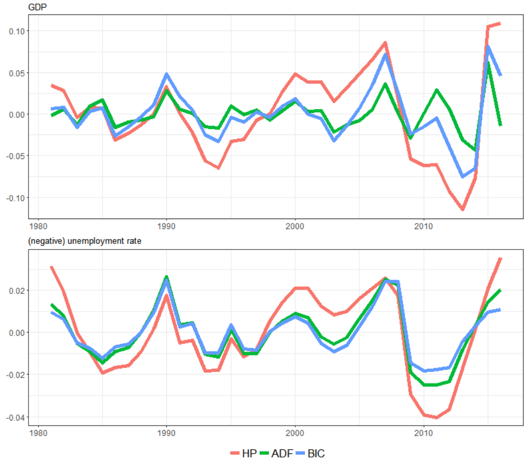
$$U_t - U_t^* = \beta \left(Y_t - Y_t^* \right) + \varepsilon_t$$

- Avoid spurious regression by focusing on business cycle
- Ball, Leigh, and Loungani (2017)
  - 20 OECD countries
  - Annual data 1980–2016.



Coefficients β and R^2 for fitted regression $U_t - U_t^* = \beta\,(Y_t - Y_t^*) + \varepsilon_t$

Legend: Red = HP; Green = bHP-ADF; Blue = bHP-BIC



Ireland: cyclical components of GDP and (negative) unemployment rate over 1980-2016

International Business Cycles

- Aguiar and Gopinath (2007):
  - "Importance of trend shocks distinguishes emerging markets from developed small open economies"
  - "emerging markets are subject to extremely volatile shocks to the stochastic trend"
  - "For emerging markets the cycle is the trend"
- 26 small open economies (40+ Quarterly observations)
  - 13 developed economies (Australia, Belgium, Canada, Finland, New Zealand,...)
  - 13 emerging economies (Argentina, Brazil,Mexico, Peru, Thailand,...)
- Quarterly data
- Time series length varies (median: 54 vs 94)

grfs

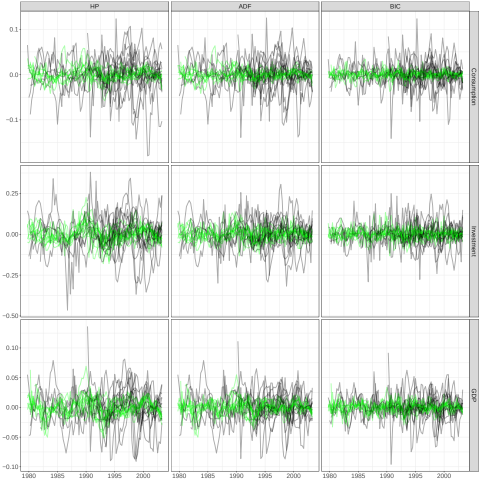
International Business Cycles

- Aguiar and Gopinath (2007):
  - “Importance of trend shocks distinguishes emerging markets from developed small open economies”
  - “emerging markets are subject to extremely volatile shocks to the stochastic trend”
  - “For emerging markets the cycle is the trend”
- 26 small open economies (40+ Quarterly observations)
  - 13 developed economies (Australia, Belgium, Canada, Finland, New Zealand,...)
  - 13 emerging economies (Argentina, Brazil,Mexico, Peru, Thailand,...)
- Quarterly data
- Time series length varies (median: 54 vs 94)

grfs

Variance and Correlation

	HP		ADF		BIC	
	emer.	dev.	emer.	dev.	emer.	dev.
	Number of iterations					
Y	1	1	2	3	10	7
C	1	1	2	2	10	7
I	1	1	4	2	12	6
	Variance and correlation coefficient					
$\sigma(Y_t)$	0.0251	0.0134	0.0228	0.0094	0.0173	0.0076
$\sigma(C_t)$	0.0339	0.0127	0.0325	0.0093	0.0233	0.0070
$\sigma(I_t)$	0.0960	0.0413	0.0786	0.0332	0.0607	0.0268
$\rho(C_t, Y_t)$	0.7592	0.7234	0.6284	0.4772	0.6610	0.4370
$\rho(I_t, Y_t)$	0.8327	0.7024	0.7527	0.5435	0.7177	0.5135
$\rho(Y_t, Y_{t-1})$	0.7608	0.7528	0.6201	0.5624	0.5444	0.4475



Cyclical components of GDP, consumption & Investment by HP, bHP-ADF, bHP-BIC filters.

Developed nations (green); Emerging nations (black)

Sum Up

Trend determination by filtering

- Boosting turns a weak base learner (a dumb HP filter) into a smarter version
  - HP filter is inconsistent for stochastic trends and general deterministic trends
  - bHP filter is consistent for stochastic trends, general deterministic trends, and consistently estimates multiple trend breaks
- Boosting works with nonstationary time series, not just iid data
- Computationally straightforward with existing software and standard setting $\lambda = 1600$

Does boosting resolve trend/cycle decomposition?

- bHP filter has good consistency properties for most standard trend forms
- Economic time series show irregular cycles (both amplitude and duration)
  - fitted regular cycles from AR models are less able to capture irregularities
  - near unit roots play a big role in AR modeling
  - long AR models reduce time series to mds sequences not cycles
- bHP filter accommodates irregular cycles by function estimation of trend

Sum Up

Trend determination by filtering

- Boosting turns a weak base learner (a dumb HP filter) into a smarter version
  - HP filter is inconsistent for stochastic trends and general deterministic trends
  - bHP filter is consistent for stochastic trends, general deterministic trends, and consistently estimates multiple trend breaks
- Boosting works with nonstationary time series, not just iid data
- Computationally straightforward with existing software and standard setting $\lambda = 1600$

Does boosting resolve trend/cycle decomposition?

- bHP filter has good consistency properties for most standard trend forms
- Economic time series show irregular cycles (both amplitude and duration)
  - fitted regular cycles from AR models are less able to capture irregularities
  - near unit roots play a big role in AR modeling
  - long AR models reduce time series to mds sequences not cycles
- bHP filter accommodates irregular cycles by function estimation of trend

Conclude: on a recent provocative directive

Hamilton (2018): “Why you should never use the HP filter”

We are much more optimistic!

- Use bHP - a function space machine learning method
- Designed to estimate trends of various unknown forms
- Delivers consistent trend determination in a wide class
- And can even estimate trend breaks

Conclude: on a recent provocative directive

Hamilton (2018): “Why you should never use the HP filter”

We are much more optimistic!

- Use bHP - a function space machine learning method
- Designed to estimate trends of various unknown forms
- Delivers consistent trend determination in a wide class
- And can even estimate trend breaks