US aggregate output measurement: A common trend approach

Tincho Almuzara
CEMFI

Gabriele Fiorentini
Università di Firenze

Enrique Sentana
CEMFI

ESCoE Conference on Economic Measurement 2018
16 – 17 May 2018, London

Motivation

- Consider the following time series signal-extraction problem

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \\ \vdots \\ y_{N,t} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} x_t + \begin{bmatrix} v_{1,t} \\ v_{2,t} \\ \vdots \\ v_{N,t} \end{bmatrix},$$

  where x_t is an unobserved $I(1)$ signal of interest and $\mathbf{y_t}$ are measurements of x_t contaminated with errors $\mathbf{v_t}$ which we assume to be independent of each other and of the signal at all leads and lags.
- In our empirical application $N = 2$ and
  - x_t is latent US aggregate output (in logs) during quarter t;
  - y_{1t} is the expenditure-based estimate of x_t (GDP);
  - y_{2t} is the income-based estimate of x_t (GDI).
  - We look at US data from 1952Q1 to 2015Q4 (256 obs).

Motivation

- In theory, the expenditure (GDP) and income (GDI) measures of aggregate (real) production should be equal.
- In practice, they differ because they rely on different sources.
- Their difference, known as the "statistical discrepancy", was regarded by some macroeconomists as a curiosity in the National Income and Product Accounts.
- However, the Great Recession has substantially renewed interest in the possibility of obtaining more reliable GDP growth figures by combining those two measures.
- In the early days, some national statistical offices computed a simple equally weighted average.
- More sophisticated methods would give higher weights to the more precise GDP measures.
- Nowadays, the Australian Bureau of Statistics reports a single official GDP figure, and the US Department of Commerce Bureau of Economic Analysis has seriously considered this possibility.

Motivation

- However, dynamic considerations matter.
- In particular, measurement errors may well be serially correlated even though they are covariance stationary, as pointed out by Smith, Weale and Satchell (1998, REStud).
- In addition, these two GDP measures should be cointegrated with the true GDP, with cointegrating vector (1,-1).
- The difference between (log) GDP and (log) GDI seems covariance stationary, although rather persistent. ▸ plot

Motivation

- Therefore, in order to work with an invertible model whose spectral density matrix has full rank at all frequencies, we combine the statistical discrepancy, $y_{2t} - y_{1t}$, and the equally weighted average of the quarterly rates of growth of GDP and GDI, $(\Delta y_{1t} + \Delta y_{2t})/2$.
- In principle, there are infinitely many other asymptotically equivalent stationarity transformations of the two output measures, but (-1,1) and (1-L)(.5,.5) seems rather natural.
- In the time domain, we can avoid this indeterminacy by computing the log-likelihood function in levels using a diffuse prior for the non-stationary component of the initial observations.
- We impose unit loadings to ensure a common average growth rate.
- The usual assumption that the covariance matrix of the common factor innovations, f_t, and the idiosyncratic factor innovations, v_{1t} and v_{2t}, is diagonal turns out to be non-parametrically just identifying in this case (subject to "admissibility").

Motivation

- Most applied papers formulate the model in first differences

$$\begin{bmatrix} \Delta y_{1,t} \\ \Delta y_{2,t} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \Delta x_t + \begin{bmatrix} u_{1,t} \\ u_{2,t} \end{bmatrix},$$

  which amounts to assuming that $d_t = y_{1,t} - y_{2,t}$ is $I(1)$ but implies that $\mathbf{u}_t$ is a strictly nonivertible process with a singular spectral density at the zero frequency under cointegration.
- Aruoba et al. (2016, JoE) specify the model in first differences with serially uncorrelated measurement errors.
- Under the plausible null of cointegration, the effects on estimation and filtering when working with first differences deserve a closer look.

Contribution

- We thoroughly asses the consequences of **not** imposing cointegration when in fact the system **is** cointegrated by:
  1. An extensive simulation study in which we assume a cointegrated DGP but fit models on first differences that do not impose cointegration.
  2. A careful theoretical analysis that compares the MSE of:
     - The Wiener-Kolmogorov filters implied by the misspecified models evaluated a their inconsistent pseudo true values
     - The optimal WK filters derived under the null of cointegration.

Contribution

- We find that results depend on
  1. The degree of observability of the signal. In this respect, we define $R^2(\lambda)$ and R^2 indices to measure observability.
  2. The dynamics of the rate of growth of the signal and, in particular, its persistence.
  3. The accuracy of the approximating misspecified model.
- In particular
  1. When the measurement error is almost negligible the signal becomes almost observable and a bad filter may produce a good measure.
  2. The dynamic properties of the signal matter
  3. The specification of high order AR processes for the measurement errors in the model in first differences will mitigate the effects of not imposing cointegration

Characterizing observability

- The spectral representation of the first differences of the observed processes is

$$\mathrm{d}z_{\Delta\mathbf{y}}(\lambda) = \mathrm{d}z_{\Delta x}(\lambda)\ell_N + (1 - e^{-i\lambda})\mathrm{d}z_{\mathbf{v}}(\lambda),$$

with spectral density matrix

$$F_{\Delta\mathbf{y}}(\lambda) = \ell_N \ell_N' f_{\Delta x}(\lambda) + F_{\Delta\mathbf{v}}(\lambda)$$

- The GLS estimator of $\mathrm{d}z_{\Delta x}(\lambda)$

$$\begin{aligned}\mathrm{d}z_{\Delta x}^{\mathrm{GLS}}(\lambda) &= \left[\ell_N' F_{\Delta\mathbf{v}}^{-1}(\lambda)\ell_N\right]^{-1} \ell_N' F_{\Delta\mathbf{v}}^{-1}(\lambda)\mathrm{d}z_{\Delta\mathbf{y}}(\lambda) \\ &= \mathrm{d}z_{\Delta x}(\lambda) + \left[\ell_N' F_{\Delta\mathbf{v}}^{-1}(\lambda)\ell_N\right]^{-1} \ell_N' F_{\Delta\mathbf{v}}^{-1}(\lambda)\mathrm{d}z_{\Delta\mathbf{v}}(\lambda)\end{aligned}$$

is minimal sufficient for $\mathrm{d}z_x$ (Fiorentini and Sentana (2017)).

$$R^2(\lambda) = \{1 + f_{\Delta x}^{-1}(\lambda)(\ell_N' F_{\Delta\mathbf{v}}^{-1}(\lambda)\ell_N)^{-1}\}^{-1}$$

We take $R^2 = \int_{\lambda\in[-\pi,\pi]} R^2(\lambda)\,\mathrm{d}\lambda$.

Characterizing observability

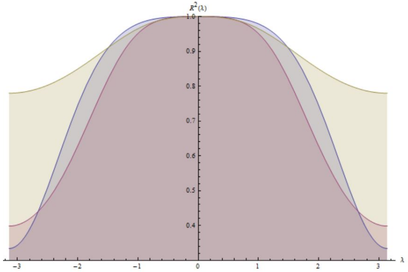


Figure: $R^2(\lambda)$ measure of observability of GDP growth

(1984Q2-2007Q2 (blue), 1984Q2-2017Q4 (red), 1952Q1-2017Q4 (yellow))

Sub-period	1984Q2-2007Q2	1984Q2-2017Q4	1952Q2-2017Q4
R^2	0.86	0.88	0.94

Simulation study

We consider the following cointegrated DGP

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} x_t + \begin{bmatrix} v_{1,t} \\ v_{2,t} \end{bmatrix},$$

$$\Delta x_t = \rho_x \Delta x_{t-1} + \sigma_x f_t$$

$$v_{it} = \rho_{v_i} v_{it-1} + \sigma_{v_i} e_{it} \qquad i = 1, 2$$

$$\sigma_x^2 = 1 - \rho_x^2, \qquad \rho_{v_1} = \rho_{v_2} = \rho_v, \qquad \sigma_{v_1}^2 = \sigma_{v_2}^2 = (1 - \rho_v^2)(1 - R^2)/R^2$$

$$\varepsilon_t \equiv (f_t, e_{1t}, e_{2t})' \quad iid \quad \sim \mathrm{N}(\mathbf{0}, \mathbf{I}_3)$$

In our simulation study we consider
$T = 120$, $\rho_x \in \{0.5, 0.9\}$, $\rho_v \in \{0.0, 0.75, 0.875\}$, $R^2 \in \{0.25, 0.5, 0.85\}$.

Simulation study

The researcher fits the model in first differences

$$\begin{bmatrix} \Delta y_{1,t} \\ \Delta y_{2,t} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \Delta x_t + \begin{bmatrix} u_{1,t} \\ u_{2,t} \end{bmatrix},$$

$$\Delta x_t = \alpha_x \Delta x_{t-1} + \gamma_x f_t$$

$$u_{it} = \sum_{k=1}^{p} \alpha_{u_i} u_{it-k} + \gamma_{u_i} e_{it} \qquad i = 1, 2$$

$$\varepsilon_t \equiv (f_t, e_{1t}, e_{2t})' \quad iid \quad \sim \ \mathrm{N}(\mathbf{0}, \mathbf{I}_3)$$

We first report results for the misspecified model $(p = 0)$ and then for the model that imposes cointegration

Simulation study $p = 0$

Design: $\rho_x = 0.5$		$\mathbb{E}[\hat{\alpha}_x]$	$\sqrt{T}\text{sd}(\hat{\alpha}_x)$	$\sqrt{T}\text{sd}(\hat{\rho}_x)$
$\rho_v = 0.0$	$R^2 = 0.25$	-0.36	2.10	1.77
	$R^2 = 0.50$	0.04	2.00	1.30
	$R^2 = 0.85$	0.56	1.84	0.96
$\rho_v = 0.75$	$R^2 = 0.25$	0.42	2.50	1.19
	$R^2 = 0.50$	0.49	1.72	0.97
	$R^2 = 0.85$	0.53	1.29	0.88
$\rho_v = 0.875$	$R^2 = 0.25$	0.48	1.83	1.03
	$R^2 = 0.50$	0.52	1.44	0.92
	$R^2 = 0.85$	0.53	1.22	0.88

For some DGPs we notice a pronounced bias in $\hat{\alpha}_x$ and, in general, a much greater sampling variability

Simulation study $p = 0$

Design: $\rho_x = 0.9$		$\mathbb{E}[\hat{\alpha}_x]$	$\sqrt{T}\mathrm{sd}(\hat{\alpha}_x)$	$\sqrt{T}\mathrm{sd}(\hat{\rho}_x)$
	$R^2 = 0.25$	-0.27	5.85	0.55
$\rho_v = 0.0$	$R^2 = 0.50$	0.77	5.28	0.51
	$R^2 = 0.85$	0.91	0.47	0.47
	$R^2 = 0.25$	0.88	1.12	0.54
$\rho_v = 0.75$	$R^2 = 0.50$	0.88	0.76	0.47
	$R^2 = 0.85$	0.88	0.57	0.45
	$R^2 = 0.25$	0.87	0.85	0.50
$\rho_v = 0.875$	$R^2 = 0.50$	0.88	0.66	0.46
	$R^2 = 0.85$	0.89	0.52	0.44

Results are qualitatively similar to the previous slide and signal to noise ratio matters

Simulation study $p = 0$

RMSE of concurrent estimator of the rate of growth of the signal $\Delta x_{T|T}$

Design: $\rho_x = 0.5$		First-diffs	Cointegration
$\rho_v = 0.0$	$R^2 = 0.25$	1.56	0.76
	$R^2 = 0.50$	0.84	0.62
	$R^2 = 0.85$	0.50	0.37
$\rho_v = 0.75$	$R^2 = 0.25$	0.76	0.62
	$R^2 = 0.50$	0.52	0.44
	$R^2 = 0.85$	0.28	0.21
$\rho_v = 0.875$	$R^2 = 0.25$	0.59	0.51
	$R^2 = 0.50$	0.40	0.33
	$R^2 = 0.85$	0.24	0.14

Simulation study $p = 0$

RMSE of concurrent estimator of the rate of growth of the signal $\Delta x_{T|T}$

Design: $\rho_x = 0.9$		First-diffs	Cointegration
$\rho_v = 0.0$	$R^2 = 0.25$	1.25	0.50
	$R^2 = 0.50$	0.51	0.42
	$R^2 = 0.85$	0.33	0.28
$\rho_v = 0.75$	$R^2 = 0.25$	0.52	0.47
	$R^2 = 0.50$	0.37	0.35
	$R^2 = 0.85$	0.20	0.19
$\rho_v = 0.875$	$R^2 = 0.25$	0.43	0.41
	$R^2 = 0.50$	0.30	0.29
	$R^2 = 0.85$	0.14	0.13

Misspecification analysis

- We explain how to obtain pseudo-true values for the parameters of the misspecified model.
- We report asymptotic biases $|\alpha_\infty - \alpha_0|$ and $|\gamma_{x\infty} - \gamma_{x0}|$ (for fixed p) and find that they are larger when the signal is more persistent and less observable
- We also detail how to compute standard errors which are robust to misspecification
- We find that the sample variability of parameter estimators in the misspecified models remains high even when the measurement error processes are allowed to follow high order $AR(p)$.

Misspecification analysis

- For large T the ML estimator $\hat{\boldsymbol{\theta}}_T$ approximately satisfies

$$0 = \frac{1}{|\Lambda_T|} \sum_{\lambda \in \Lambda_T} \frac{\partial \text{vec}'(f_{\Delta \mathbf{y}}(\lambda|\hat{\boldsymbol{\theta}}_T))}{\partial \boldsymbol{\theta}} \left[f_{\Delta \mathbf{y}}^{-1}(\lambda|\hat{\boldsymbol{\theta}}_T) \otimes (f_{\Delta \mathbf{y}}^{-1}(\lambda|\hat{\boldsymbol{\theta}}_T))' \right] \times \text{vec}\left[I'_{\Delta \mathbf{y}}(\lambda) - f'_{\Delta \mathbf{y}}(\lambda|\hat{\boldsymbol{\theta}}_T) \right]$$

where $f_{\Delta \mathbf{y}}$ is the spectral density of the estimated model and $I_{\Delta \mathbf{y}}$ is the periodogram of both first differences

Misspecification analysis

Monte Carlo design	Theoretical $a_{x\infty}$	Theoretical q_∞	MC average $\hat{a}_{x,\infty}$	MC average $\hat{q}_\infty$
$\rho_{x0} = 0.20,\ R_0^2 = 0.25$	-0.55	0.27	-0.55	0.29
$\rho_{x0} = 0.20,\ R_0^2 = 0.50$	-0.05	0.99	-0.06	0.95
$\rho_{x0} = 0.20,\ R_0^2 = 0.85$	0.16	5.64	0.16	5.42
$\rho_{x0} = 0.50,\ R_0^2 = 0.25$	-0.55	0.22	-0.52	0.25
$\rho_{x0} = 0.50,\ R_0^2 = 0.50$	0.29	0.98	0.28	0.93
$\rho_{x0} = 0.50,\ R_0^2 = 0.85$	0.46	5.62	0.46	5.52
$\rho_{x0} = 0.80,\ R_0^2 = 0.25$	-0.53	0.21	-0.48	0.22
$\rho_{x0} = 0.80,\ R_0^2 = 0.50$	0.80	0.77	0.72	0.74
$\rho_{x0} = 0.80,\ R_0^2 = 0.85$	0.79	5.56	0.78	5.44

$$\rho_v = 0, \qquad p = 0, \qquad q = R^2/(1 - R^2)$$

A comparison of asymptotic standard errors for $\hat{a}_x$ and $\hat{\gamma}_x$

Estimated model	$\text{std}(\hat{a}_{xT})$	$\text{std}(\hat{g}_{xT})$	$\text{std}(\hat{a}_{xT})$ –MC–	$\text{std}(\hat{g}_{xT})$ –MC–
M_0	1.7469	2.3287	1.8351	2.4270
M_1	1.3976	2.1569	1.3818	2.1194
M_2	1.3548	2.1538	1.3386	2.1252
M_∞	0.9115	1.4855	0.8964	1.4317

$$(\text{std}^2(\hat{\boldsymbol{\theta}}_T) \equiv \lim_{T\to\infty} \text{var}(\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_\infty))$$

(Monte Carlo design: $\rho_x = 0.5$, $\gamma_{x0} = 0.50$, $\rho_v = 0$ and $R_0^2 = 0.5$)

Misspecification analysis: Filtering the signal

- $\Delta\hat{x}_{t|T}^{(p)} = \mathbb{E}\left[\Delta x_t \mid \mathcal{Y}_T, \hat{\theta}_T, M_p\right]$
- The expectation is computed under model M_p, i.e. $\mathbf{u}$ is (diagonal) VAR(p)
- How do filters/smoothers of the signal $\Delta\hat{x}_{t|T}^{(p)}$ behave?
- As expected, the Kalman smoother under M_∞ achieves highest accuracy regardless of parameter uncertainty and it consistently estimates the variance of $\xi_{t|T}^{(\infty)}$
- The Kalman smoother based on M_p has no significant bias but is substantially more variable than the optimal filter and gives misleading inference on variance of $\xi_{t|T}^{(p)}$
- The RMSE of $\xi_{t|T}^{(p)}$ is larger when the signal is more persistent and less observable
- As p increases the RMSE of $\xi_{t|T}^{(p)}$ decreases towards that of $\xi_{t|T}^{(\infty)}$

Misspecification analysis

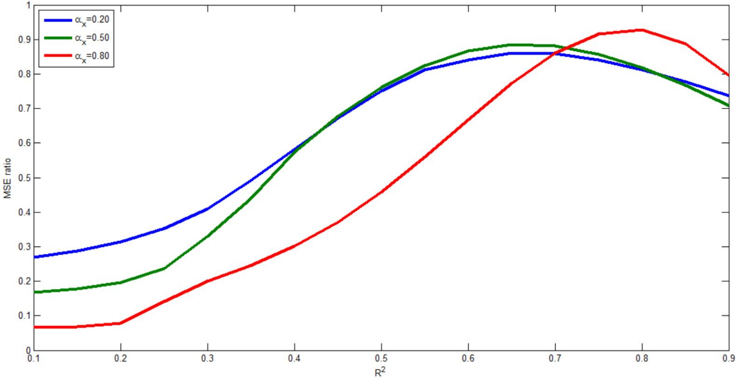


Figure: Ratio of MSE of efficient filter to implied filter in M_0

Misspecification analysis: filtering

$\text{var}(\xi_{t|T}^{(\infty)})$ as a percentage of $\text{var}(\xi_{t|T}^{(p)})$

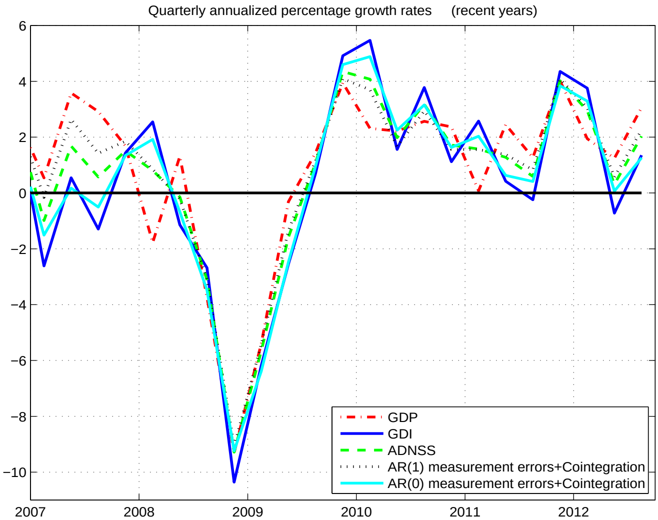
Monte Carlo design	AR(0) at true values	AR(1) at true values	AR(0) at estimates	AR(1) at estimates
$\rho_{x0} = 0.20,\ R_0^2 = 0.25$	46.6	78.8	45.6	75.0
$\rho_{x0} = 0.20,\ R_0^2 = 0.50$	75.2	95.6	73.7	93.3
$\rho_{x0} = 0.20,\ R_0^2 = 0.85$	95.6	98.8	90.3	87.9
$\rho_{x0} = 0.50,\ R_0^2 = 0.25$	32.9	83.6	32.1	65.7
$\rho_{x0} = 0.50,\ R_0^2 = 0.50$	76.3	94.0	73.7	91.8
$\rho_{x0} = 0.50,\ R_0^2 = 0.85$	94.3	98.4	90.3	92.7
$\rho_{x0} = 0.80,\ R_0^2 = 0.25$	16.7	83.3	16.3	64.1
$\rho_{x0} = 0.80,\ R_0^2 = 0.50$	80.6	94.5	70.6	91.9
$\rho_{x0} = 0.80,\ R_0^2 = 0.85$	92.3	97.5	88.0	76.1

Misspecification analysis: forecasting

Relative efficiency in forecasting $\Delta\mathbf{y}$

Monte Carlo design	AR(0) at true values	AR(1) at true values	AR(0) at estimates	AR(1) at estimates
$\rho_{x0} = 0.20,\ R_0^2 = 0.25$	75.4	86.2	75.5	86.5
$\rho_{x0} = 0.20,\ R_0^2 = 0.50$	87.4	92.9	87.7	93.5
$\rho_{x0} = 0.20,\ R_0^2 = 0.85$	96.2	97.9	96.1	98.5
$\rho_{x0} = 0.50,\ R_0^2 = 0.25$	77.4	84.6	77.5	87.5
$\rho_{x0} = 0.50,\ R_0^2 = 0.50$	85.9	92.6	86.3	93.3
$\rho_{x0} = 0.50,\ R_0^2 = 0.85$	95.3	97.5	95.4	98.1
$\rho_{x0} = 0.80,\ R_0^2 = 0.25$	74.1	85.2	74.2	86.4
$\rho_{x0} = 0.80,\ R_0^2 = 0.50$	81.5	90.4	81.8	91.2
$\rho_{x0} = 0.80,\ R_0^2 = 0.85$	91.9	95.9	91.4	96.1

Empirical Application: Smoothed US aggregated growth



Understanding bias

- Consider fitting the spectral density of observables $f_{\Delta \boldsymbol{y}\Delta \boldsymbol{y}}$ choosing a spectral density $g_{\Delta x}$ and constants $g_1, \ldots, g_N$ to minimize

$$\mathcal{Q}(g_{\Delta x}, g_1, \ldots, g_N) = \frac{1}{2}\int_{\lambda\in[-\pi,\pi]} \log\det(g_{\Delta \boldsymbol{y}\Delta \boldsymbol{y}}(\lambda))\, \mathrm{d}\lambda + \frac{1}{2}\int_{\lambda\in[-\pi,\pi]} \mathrm{tr}(g^{-1}_{\Delta \boldsymbol{y}\Delta \boldsymbol{y}}(\lambda) f_{\Delta \boldsymbol{y}\Delta \boldsymbol{y}}(\lambda))\, \mathrm{d}\lambda,$$

$$\text{where } g_{\Delta \boldsymbol{y}\Delta \boldsymbol{y}}(\lambda) = g_{\Delta x}(\lambda)\boldsymbol{\ell}_N\boldsymbol{\ell}'_N + \mathrm{G}, \text{ all } \lambda \in [-\pi,\pi],$$
$$\text{and } \mathrm{G} = \mathrm{diag}(g_1, \ldots, g_N).$$

- This program captures the estimation of the spectral density of the signal $f_{\Delta x\Delta x}$ by means of $g_{\Delta x}$ when measurement errors are restricted to be white noise uncorrelated to each other.

Understanding bias

- When $N = 2$ we obtain the following bias

$$g_{\Delta x}(\lambda) - f_{\Delta x \Delta x}(\lambda) = \frac{g_2^2}{(g_1 + g_2)^2}(f_{\Delta v_1 \Delta v_1}(\lambda) - g_1) + \frac{g_1^2}{(g_1 + g_2)^2}(f_{\Delta v_2 \Delta v_2}(\lambda) - g_2).$$

- If we further assume $f_{vv}(\lambda) \equiv f_{v_1 v_1}(\lambda) = f_{v_2 v_2}(\lambda)$, all λ,

$$g_{\Delta x}(\lambda) - f_{\Delta x \Delta x}(\lambda) = \frac{1}{2}(f_{\Delta v \Delta v}(\lambda) - g),$$

$$g = \left(\int_{\lambda \in [-\pi, \pi]} (f_{\Delta x \Delta x}(\lambda) + f_{\Delta v \Delta v}(\lambda)/2)\, \mathrm{d}\lambda \right)^{-1} \times \left(\int_{\lambda \in [-\pi, \pi]} f_{\Delta v \Delta v}(\lambda)(f_{\Delta x \Delta x}(\lambda) + f_{\Delta v \Delta v}(\lambda)/2)\, \mathrm{d}\lambda \right)$$

Understanding bias

- g is as a projection coefficient of $f_{\Delta v \Delta v}(\lambda)$ onto a constant in a Hilbert space induced by the weighted inner product with $\omega(\lambda) \equiv f_{\Delta x \Delta x}(\lambda) + f_{\Delta v \Delta v}(\lambda)/2$.
- $\omega(\lambda)$ is proportional to $\det(f_{\Delta \boldsymbol{y} \Delta \boldsymbol{y}}(\lambda))/f_{\Delta v \Delta v}(\lambda)$ and to the spectral density of the GLS estimator of the signal.
- Bias would be identically zero if the assumption that measurement errors in first-difference are white noise were correct.
- Any departure from this assumption means re-distributing power across frequencies for the pseudo-true spectral density $g_{\Delta x}$ relative to $f_{\Delta x \Delta x}$.
- If $f_{vv}(\lambda)$ is constant, g will be some intermediate value in the range of $f_{\Delta v \Delta v}(\lambda)$ implying a distortion that tends to depress the spectrum for low frequencies and emphasize the spectrum at high frequencies.
- Bias disappears as the R^2 measure of observability tends to unity.

Statistical discrepancy

b