Realized Volatility, Heterogeneous Market Hypothesis and the Extended Wold Decomposition

Claudio Tebaldi

Bocconi University and IGIER

Conference in honor of J. Gatheral's 60th birthday
Courant Institute NYU
October 14, 2017

Foreword on Impulse Response Functions

'Dynamic economic models make predictions about impulse responses .....impulse responses quantify the exposure to long run macroeconomic shocks.... Financial markets provide compensations to investors who are exposed to these shocks.'
J. Borovicka L. Hansen 2016

This exotic research program came to my mind...

World Bachelier Conference London 2008 Program

- *08.30-09.30 Plenary lecture: Lars Peter Hansen "Modelling the long run: valuation in dynamic stochastic economies"*
- *09.30-10.30 Plenary lecture: Jim Gatheral "Consistent modelling of VIX and SPX options"*

JIM .... YOU'RE RESPONSIBILE!

Research Group on Impulse Response Functions

- OTT with F. Ortu and A. Tamoni "Long Run Risk and the Persistence of Consumption Shocks " *The Review of Financial Studies* (2013), **26** (11) 2876-2915.
- BPTT with F.Bandi, B. Perron and A.Tamoni "The Scale of Predictability", forthcoming *Journal of Econometrics*.
- OSTT with F.Ortu, F. Severino and A.Tamoni "A persistence-based Wold-type decomposition for stationary time series", Working Paper IGIER under review.
- CVOST Cerreia-Vioglio, S., Ortu, F., Severino, F., Tebaldi, C., 2017. Multivariate Wold Decompositions. Working Paper IGIER n.606, Bocconi University
- Rough Cascades and a potential resolution of Volatility and Interest Rate Puzzles. Daniele D'Ascenzo (now JP Morgan) Daniele D'Arienzo Bocconi PhD Candidate

Motivation of the Talk

- Two well-established stylized facts that characterize economic fluctuations in dynamic economies and financial markets:
  - The 'multiscale' nature of information based agent decisions: intrinsic frequencies range from HFT trading decisions to secular trends.
  - Widespread observation of self-similarity and scale invariance. The 'critical' nature of economic fluctuations
- Research Program: explore the implications of these facts on impulse response functions with particular attention to their normative rather than descriptive implications.
- This talk: I will discuss the implications of these facts on log-volatility IRf to conclude that there are important "structural" motivations that suggest the introduction of a Rough Cascade Volatility Model.

Plan of the talk

- Impulse Response Functions in Dynamic Stochastic Economies.
- Fluctuations Theory and Critical Phenomena: Scaling, Universality and Renormalization Group.
- The Extended Wold Decomposition.
- Heterogeneous Market Hypothesis, Resolution Filtration and Cascades of Shocks.
- Volatility Forecasting with (Rough) Impulse Response Functions.

Literature on IRf

- **Slutsky** (1927), **Yule** (1927) and **Frisch** (1933) formulated the concepts of “propagation” and “impulse” in economic time series. **Wold** (1938) formalizes the notion of IRF for a stationary time series.
- Identification challenges in rational expectation models **Sims** (1980).
- **Hansen, Scheinkmann** and **Borovicka** (2011-2016) introduce the modern non-linear continuous time extensions of IRf and its relation with Malliavin Derivatives and Option sensitivities.
- Extension of the IRf as a relevant 'normative' tool: see **Boijnov Shepard** (2017) ‘Time series experiments and causal estimands’.

Classical Wold Decomposition

Given a zero-mean, regular, weakly stationary time series $\mathbf{x} = \{x_t\}_{t\in\mathbb{Z}}$, we have

$$x_t = \sum_{h=0}^{+\infty} \alpha_h \varepsilon_{t-h} + \nu_t \qquad \forall t \in \mathbb{Z},$$

where the equality is in the L^2-norm.

- The process $\boldsymbol{\varepsilon} = \{\varepsilon_t\}_{t\in\mathbb{Z}}$ is a unit variance white noise.
- α_h is the projection coefficient of x_t on the linear space generated by the innovation ε_{t-h}:

  $$\alpha_h = \mathbb{E}\left[x_t \varepsilon_{t-h}\right], \qquad h \in \mathbb{N}_0.$$

  α_h is the impulse response function of x_t to the shock ε_{t-h}.

- The deterministic component $\boldsymbol{\nu}$ is orthogonal to $\boldsymbol{\varepsilon}$, id est

  $$\mathbb{E}\left[\nu_t \varepsilon_{t-h}\right] = 0 \qquad \forall h \in \mathbb{N}_0.$$

Fluctuations Theory and Critical Phenomena

- (Widom) *scaling* law for Magnetization

$$M(H,T) = |t|^{\beta} F\left(\frac{H}{|t|^{\beta\delta}}\right) \qquad t = \frac{T - T_c}{T_c}$$

  where function F is a *universal* scaling such that:
  - $M \sim H^{1/\delta}$ for $|t| = 0$,
  - $M \sim |t|^{\beta}$ for $H = 0$ and $|t| \to 0$
- Kadanoff's idea was that in the critical regime a thermodynamic system, due to the strong correlations among the microscopic variables, behaves as if constituted by rigid blocks of arbitrary size.
- Wilson Renormalization group: a semigroup of transformations that produces a progressive elimination of the microscopic degrees of freedom to obtain the asymptotic large scale properties of the system.

Renormalization Group in a Nutshell

- Renormalization Transformation and the Central Limit Theorem (Jona-Lasinio Phys. Reports 2001) Consider $\xi_1, ..., \xi_n, ...$ i.i.d. with zero mean and unit variance, and define block variables:

$$\zeta_n^1 = 2^{-\frac{n}{2}} \sum_{i=1}^{2^n} \xi_i,\ \zeta_n^2 = 2^{-\frac{n}{2}} \sum_{i=2^n+1}^{2^{n+1}} \xi_i \text{ , then } \zeta_{n+1} = \frac{\zeta_n^1 + \zeta_n^2}{\sqrt{2}}$$

and correspondingly on the densities:

$$\mathcal{R} p_n^{\zeta} := \int dy p_n^{\zeta} \left(x\sqrt{2} - y \right) p_n^{\zeta} (y)$$

The fixed point of the transformation: $\mathcal{R} p_\infty^{\zeta} = p_\infty^{\zeta}$ selects the standard normal distribution.

Renormalization Group in a Nutshell II

- The action of $\mathcal{R}$ in the vicinity of the fixed point:

$$\mathcal{R}p_{\infty}^{\zeta}\left(1+\eta h\right)=p_{\infty}^{\zeta}\left(1+\eta\mathcal{L}h\right)+O\left(\eta^{2}\right)$$

  defines a linear operator $\mathcal{L}$:

$$\mathcal{L}h\left(x\right):=\frac{2}{\sqrt{\pi}}\int dye^{-y^{2}}h\left(x\sqrt{2}-y\right)$$

- Eigenfunctions are given by Hermite functions h_n
- Eigenvalues $\lambda_n = 2^{1-\frac{n}{2}}$, $0 \le n < 2$ relevant directions, $n = 2$ marginal, $n > 2$ irrelevant.

Renormalization Group in a Nutshell III

- More generally the Renormalization Semigroup establishes a modified stochastic limiting procedure for stochastic correlated variables. E.g. Sinai defines an $H-$ self similar process as a fixed point of a rescaling transformation: $\zeta_{n+1} = \frac{\zeta_n^1 + \zeta_n^2}{2^H}$.
- Critical properties are determined by fixed points of proper RG procedures that connect microscopic model critical correlations to macroscopic observable behavior.
- Information on eigenvalues and eigenfunctions of the linearized operator provide information on critical indices and on finite size scaling functions.

Rescaling Transformation on Time Series and Discrete Haar filter

Multiresolution decomposition:

$$x_t = \sum_{j=1}^{J} \breve{g}_t^{(j)} + \breve{\pi}_t^{(J)} \qquad \forall t \in \mathbb{Z}.$$

- $\breve{\pi}_t^{(j)}$ is the average of size 2^j of past values of x_t:

$$\breve{\pi}_t^{(j)} = \frac{1}{2^j} \sum_{p=0}^{2^j-1} x_{t-p}.$$

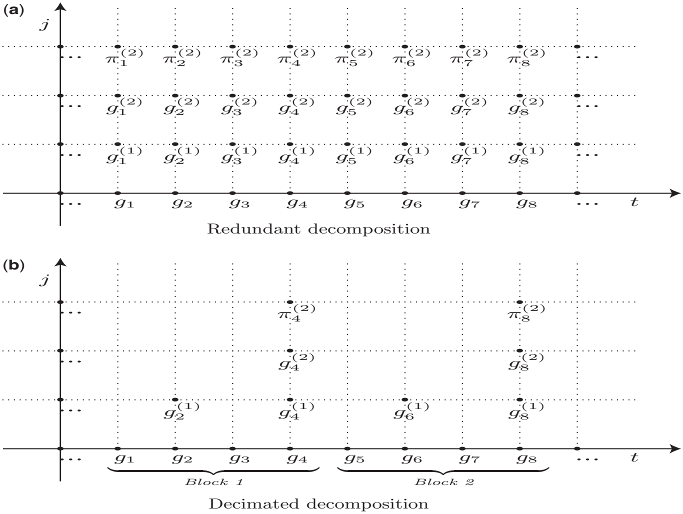
- $\breve{g}_t^{(j)}$ is the difference between these averages:

$$\breve{g}_t^{(j)} = \breve{\pi}_t^{(j-1)} - \breve{\pi}_t^{(j)}.$$

Rescaling Transformation and Persistence

- Variables $\breve{g}_t^{(j)}$ is associated with the level of persistence j: it captures fluctuations of x_t with half-life in $[2^{j-1}, 2^j)$. In this way, disentangle low-frequency shocks from high-frequency fluctuations.
- Decimation procedure is necessary to get rid of the spurious correlation due to the overlapping of observations in the construction of $\breve{g}_t^{(j)}$. Decimation selects the relevant degrees of freedom removing redundant statistics.
- Decimated (detail) components $\breve{g}_{t-2^j k}^{(j)}$ are proportional to Haar Transform of the original time series and are in one-to-one relationship with the original time series.
- However, even after decimation, variables $\breve{g}_t^{(j)}$ may be correlated, not useful to define an IRf.

Redundant vs Decimated observations



The Abstract Wold Theorem

A general approach to derive orthogonal decompositions of time series follows from the Abstract Wold Theorem, that involves an isometric operator on a Hilbert space.

Theorem (Abstract Wold Theorem)

Consider a Hilbert space $\mathcal{H}$ and an isometry $\mathbf{V} : \mathcal{H} \longrightarrow \mathcal{H}$, i.e.

$$\langle \mathbf{V}x, \mathbf{V}y \rangle = \langle x, y \rangle \qquad \forall x, y \in \mathcal{H}.$$

Then $\mathcal{H}$ decomposes into an orthogonal sum $\mathcal{H} = \hat{\mathcal{H}} \oplus \tilde{\mathcal{H}}$, where

$$\hat{\mathcal{H}} = \bigcap_{j=0}^{+\infty} \mathbf{V}^j \mathcal{H}, \qquad \tilde{\mathcal{H}} = \bigoplus_{j=0}^{+\infty} \mathbf{V}^j \mathcal{L}^{\mathbf{V}}$$

and the wandering subspace $\mathcal{L}^{\mathbf{V}}$ is defined as $\mathcal{L}^{\mathbf{V}} = \mathcal{H} \ominus \mathbf{V}\mathcal{H}$.

The Classical Wold Decomposition from the Abstract Wold Theorem

- The Classical Wold Decomposition follows from the Abstract Wold Theorem by considering the Hilbert space

$$\mathcal{H}_t(\mathbf{x}) = \mathrm{cl}\left\{ \sum_{k=0}^{+\infty} a_k x_{t-k} : \quad \sum_{k=0}^{+\infty} \sum_{h=0}^{+\infty} a_k a_h \gamma(k-h) < +\infty \right\},$$

  where $\gamma : \mathbb{Z} \longrightarrow \mathbb{R}$ is the autocovariance function of $\mathbf{x}$.

- The isometric operator that works on $\mathcal{H}_t(\mathbf{x})$ to obtain the Classical Wold Decomposition is the lag operator:

$$\mathbf{L} : \qquad \sum_{k=0}^{+\infty} a_k x_{t-k} \longmapsto \sum_{k=0}^{+\infty} a_k x_{t-1-k}.$$

The Extended Wold Decomposition: the set-up

The instrument is the Abstract Wold Theorem. Which isometric operator?

- We employ the Hilbert space

$$\mathcal{H}_t(\boldsymbol{\varepsilon}) = \left\{ \sum_{k=0}^{+\infty} a_k \varepsilon_{t-k} : \quad \sum_{k=0}^{+\infty} a_k^2 < +\infty \right\},$$

  i.e. the space spanned by the classical Wold innovations of $\mathbf{x}$.

- Inspired by RG ideas, the isometry is the scaling operator

$$\mathbf{R} : \quad \sum_{k=0}^{+\infty} a_k \varepsilon_{t-k} \quad \longmapsto \quad \sum_{k=0}^{+\infty} \frac{a_k}{\sqrt{2}} \left(\varepsilon_{t-2k} + \varepsilon_{t-2k-1} \right).$$

Theorem

$\mathcal{H}_t(\varepsilon)$ decomposes into the orthogonal sum

$$\mathcal{H}_t(\varepsilon) = \bigoplus_{j=1}^{+\infty} \mathbf{R}^{j-1}\mathcal{L}_t^{\mathbf{R}},$$

where

$$\mathbf{R}^{j-1}\mathcal{L}_t^{\mathbf{R}} = \left\{ \sum_{k=0}^{+\infty} b_k^{(j)} \varepsilon_{t-k2^j}^{(j)} \in \mathcal{H}_t(\varepsilon) : \quad b_k^{(j)} \in \mathbb{R} \right\},$$

with

$$\varepsilon_t^{(j)} = \frac{1}{\sqrt{2^j}} \left(\sum_{i=0}^{2^{j-1}-1} \varepsilon_{t-i} - \sum_{i=0}^{2^{j-1}-1} \varepsilon_{t-2^{j-1}-i} \right).$$

The decomposition of x_t

- Observe that (the purely non-deterministic part of) x_t belongs to $\mathcal{H}_t(\varepsilon)$.
- Denote, then, by $g_t^{(j)}$ the projection of x_t on $\mathbf{R}^{j-1}\mathcal{L}_t^{\mathbf{R}}$.

Proposition

Under the above conditions,

$$x_t = \sum_{j=1}^{+\infty} g_t^{(j)} = \sum_{j=1}^{+\infty} \sum_{k=0}^{+\infty} \beta_k^{(j)} \varepsilon_{t-k2^j}^{(j)},$$

where $\beta_k^{(j)} = \mathbb{E}\left[x_t \varepsilon_{t-k2^j}^{(j)}\right]$ is given by

$$\beta_k^{(j)} = \frac{1}{\sqrt{2^j}} \left(\sum_{i=0}^{2^{j-1}-1} \alpha_{k2^j+i} - \sum_{i=0}^{2^{j-1}-1} \alpha_{k2^j+2^{j-1}+i} \right).$$

The persistence-based Wold-type Decomposition Theorem

Theorem

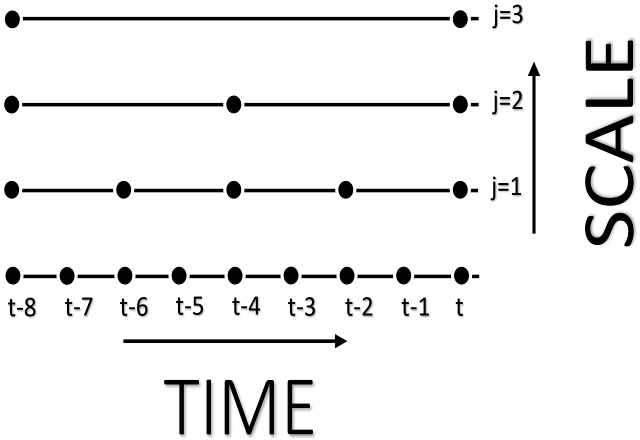
Given a zero-mean, weakly stationary purely non-deterministic time series $\mathbf{x} = \{x_t\}_{t \in \mathbb{Z}}$, *then*

$$x_t = \sum_{j=1}^{+\infty} g_t^{(j)} = \sum_{j=1}^{+\infty} \sum_{k=0}^{+\infty} \beta_k^{(j)} \varepsilon_{t-k2^j}^{(j)}.$$

- *For any scale* j, *the process* $\left\{ \varepsilon_{t-k2^j}^{(j)} \right\}_{k \in \mathbb{Z}}$ *is a unit variance white noise.*
- *The multiscale impulse responses* $\beta_k^{(j)}$ *are unique, they do not depend on* t *and* $\sum_k \left(\beta_k^{(j)} \right)^2 < +\infty$.
- $\mathbb{E}\left[g_{t-p}^{(j)} g_{t-q}^{(l)} \right]$ *depends at most on* $j, l, p-q$ *and*

$$\mathbb{E}\left[g_{t-m2^j}^{(j)} g_{t-n2^l}^{(l)} \right] = 0 \qquad \forall j \neq l, \quad \forall m, n \in \mathbb{N}_0.$$

Resolution Filtration



The decomposition of x_t: remarks

- From now on, we call persistent component at scale j the quantity

$$g_t^{(j)} = \sum_{k=0}^{+\infty} \beta_k^{(j)} \varepsilon_{t-k2^j}^{(j)}.$$

- When t is fixed, innovations of $g_t^{(j)}$ have support

$$S_t^{(j)} = \{t - k2^j : \quad k \in \mathbb{Z}\},$$

that becomes sparser and sparser as j increases.

- $\beta_k^{(j)}$ is the multiscale impulse response associated with the innovation at scale j and time shift $k2^j$.
- Importantly, components at different scales are uncorrelated:

$$\mathbb{E}\left[g_t^{(j)} g_t^{(l)}\right] = 0, \qquad j \neq l.$$

EWD forecasting and extensions

- Multivariate extension based on the theory of modules: $\beta_k^{(j)}$ matrix of impulse responses (OSTT and COST)
- Extension of the Beveridge Nelson permanent-transitory decomposition (OST).
- Linear Forecasting theory for Wold decomposition induced by isometry R: is a discretized version of the linear forecasting theory for wide sense self-similar processes by Nuzman and Poor (2000, 2001).
- The dyadic non-linear extension of the IRf definition: Gaussian Stochastic Calculus of Variations (Malliavin Thalmeier 2006) see also Stroock construction of Malliavin Calculus.

Estimation of multiscale IRFs

Given a weakly stationary time series $\mathbf{x} = \{x_t\}_t$ we estimate multiscale IRFs in three steps:

1. estimate a suitable autoregressive form for $\mathbf{x}$ by exploiting AIC of BIC;
2. by operator inversion, turn the AR process into an MA and estimate classical IRFs α_h;
3. from coefficients α_h, estimate multiscale IRFs $\beta_k^{(j)}$.

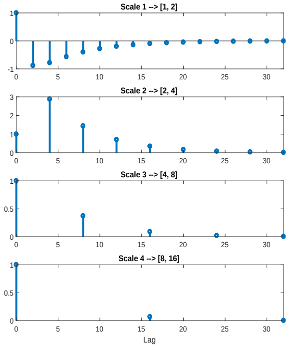
Simulations: multiscale IRF of an AR(2)

- Consider a weakly stationary purely non-deterministic $AR(2)$ processes

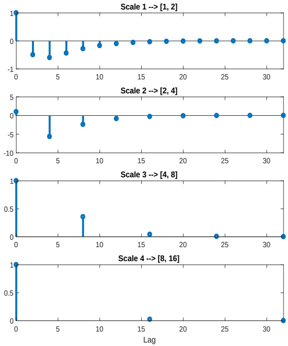
$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \varepsilon_t$$

- Compare two specifications $\phi_1 = 1.16$ and $\phi_2 = -0.27$ vs $\phi_1 = 1.3$ and $\phi_2 = -0.41$.
    - IRf at scale level $j = 2$, denote overreaction of the $AR(2)$ process
    - Scale levels $j \geq 3$ have the same behaviour as the multiscale impulse response functions of an $AR(1)$.

Simulations: multiscale IRFs of an AR(2)



(c) Multiscale impulse responses.



(d) Multiscale impulse responses.

Predictability of consumption growth components

- Run a regression component by component, namely:

$$\Delta c_{j,t+1,t+2^j} = \beta_0 + \beta_1 pd_{j,t} + \epsilon_{t+2^j,j}$$

	Scale j							
	1	2	3	4	5	6	7	8
pd_t	0.31	-0.49	**-0.73**	0.16	-0.17	**-0.35**	**0.28**	0.40
	(0.74)	(-1.75)	(-2.88)	(0.50)	(-0.85)	(-1.93)	(2.56)	(1.51)
	[0.00]	[0.01]	[0.06]	[0.01]	[0.02]	[0.24]	[0.38]	[0.01]

Table: The table reports OLS estimates of the regressors, corrected t-statistics in parentheses and adjusted R^2 statistics in square brackets. The sample spans the period 1947Q2-2009Q4.

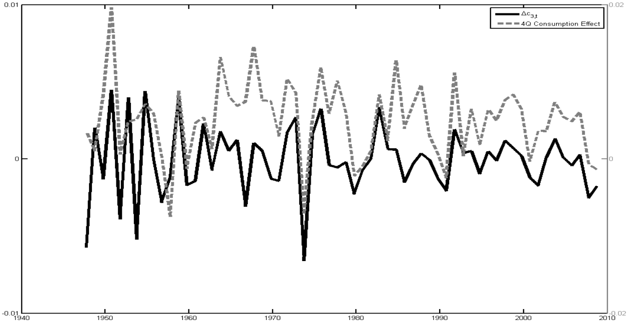
- Long lasting cycles in consumption growth are forecasted by cycles of corresponding length in asset prices scaled by dividends

Identification of consumption drivers

- Which are the economic drivers of the predictable components of consumption growth?
- To determine these drivers we search for time series that are
    - characterized by an half-life close to the one of the component they are to drive
    - correlated with such component

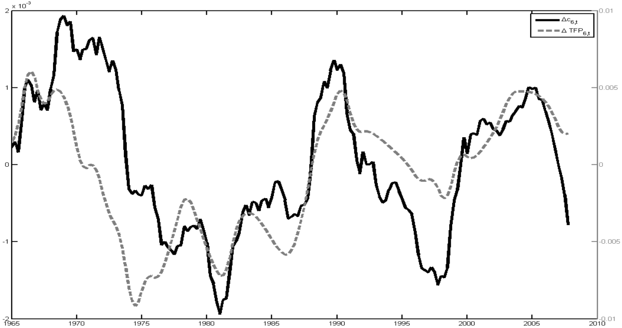
Identification - 3rd Component

- Our third component captures with a correlation of about 60% the alignment between consumption and investment decisions in the fourth quarter (fourth-quarter effect, e.g. Moller and Rangvid, 2010).



Identification - 6th Component

- Predictable components of consumption that occur at cycles between 8 and 16 years reveal the position of the economy with respect to the technological cycle (e.g. Garleanu et al., 2009 and Kung and Schmid, 2011), with a correlation of 64%.



Identification - 7th Component

- Our seventh component captures the alternating twenty-year periods of booms and busts of US live births, with a correlation of about 44%.

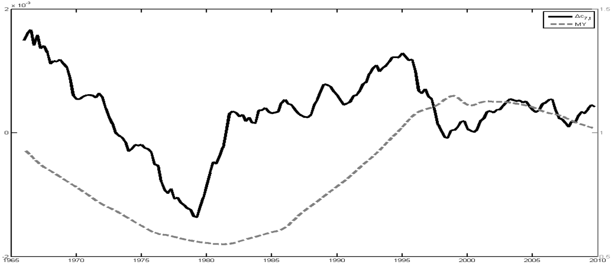


Figure: The seventh component of consumption growth, $g_{7,t}$ along with a demographic variable, MY_t, the middle-aged to young ratio proposed in Geanakoplos et al. (2004).

The Equity Premium

- Premium on the market return satisfies:

$$E\left[r_{m,t+1} - r_{f,t}\right] + \frac{\sigma_m^2}{2} \quad = \quad \lambda_\eta \sigma_\eta^2 + \kappa_{1,m} \underline{\lambda}_\varepsilon' \mathbf{Q} \underline{A_m}$$

$$\underline{A_m} = \left(\mathbb{I}_J - \kappa_{1,m} diag\left(\underline{\rho}\right)\right)^{-1} \left(\underline{\phi} - \frac{1}{\psi}\underline{1}\right) \quad \mathbf{Q} = \mathbf{E}_t\left[\varepsilon_{t+1}\varepsilon_{t+1}'\right]$$

where $\underline{\rho} = (\rho_1, \ldots, \rho_j, \ldots, \rho_J)'$

- The vector $\underline{\lambda}_\varepsilon$ determines the term structure of risk prices.
- The exposure to risks depends on $\mathbf{Q}\underline{A_m}$.

A calibration for equity premia

Use the estimate of $\psi = 5$ and calibrate $\gamma = 5$. Then the equity premia at different scales are:

Scale $j =$	Half-life (Years)	Q_j $(1.0e-005)$	Risk Exposure (1.0e-003)	Risk Price	Risk Premium (%)
1	0.08	0.31	1.072	4.67	0.50
2	0.44	0.18	0.712	12.12	0.86
3	1.52	0.15	0.592	32.33	1.91
4	3.63	0.12	0.652	96.03	6.29
5	4.57	0.07	0.288	168.69	4.86
6	12.5	0.05	0.140	181.71	2.51
7	18.77	0.05	0.068	183.28	1.25
8	33.27	0.07	0.016	183.84	0.26

Table: This table reports equity premium (in %) $E_t[r_{m,t+1} - r_{f,t}]$ decomposed by level of persistence.

Reconstructing a time series from its scale components - 1

Question: given the dynamics of the components at different scales, what can we say about the process $\mathbf{x}$ built by summing up such components?

- In order to make the sum feasible, we need to assume a common innovation process $\boldsymbol{\varepsilon}$ defined for any $t \in \mathbb{Z}$.
- At each scale level j, we define the detail process $\boldsymbol{\varepsilon}^{(\boldsymbol{j})} = \left\{\varepsilon_t^{(j)}\right\}_{t\in\mathbb{Z}}$ as a $MA(2^j - 1)$ driven by innovations $\boldsymbol{\varepsilon}$:

$$\varepsilon_t^{(j)} = \sum_{i=0}^{2^j-1} \delta_i^{(j)} \varepsilon_{t-i}, \qquad \delta_i^{(j)} \in \mathbb{R}, \quad i = 0, \dots, 2^j - 1.$$

- Extending the renormalization argument to non trivial fixed points.

Reconstructing a time series from its scale components - 2

We consider the processes $\mathbf{g}^{(\mathbf{j})} = \left\{ g_t^{(j)} \right\}_{t\in\mathbb{Z}}$, with degree of persistence j, such that:

1)

$$g_t^{(j)} = \sum_{k=0}^{+\infty} \beta_k^{(j)} \varepsilon_{t-k2^j}^{(j)}$$

2)

$$\sum_{j=1}^{+\infty} \sum_{h=0}^{+\infty} \left(\beta^{(j)}_{\left\lfloor \frac{h}{2^j} \right\rfloor} \delta^{(j)}_{h-2^j\left\lfloor \frac{h}{2^j} \right\rfloor} \right)^2 < +\infty$$

3) $\mathbb{E}\left[g_{t-p}^{(j)} g_{t-q}^{(l)} \right]$ depends at most on $j, l, p-q$ and

$$\mathbb{E}\left[g_{t-m2^j}^{(j)} g_{t-n2^l}^{(l)} \right] = 0 \qquad \forall j \neq l, \quad \forall m, n \in \mathbb{N}_0.$$

The Reconstruction Theorem

Theorem

Under the above assumptions, the process $\mathbf{x} = \{x_t\}_{t\in\mathbb{Z}}$ *defined by*

$$x_t = \sum_{j=1}^{+\infty} g_t^{(j)}$$

is zero-mean, weakly stationary purely non-deterministic and

$$x_t = \sum_{h=0}^{+\infty} \alpha_h \varepsilon_{t-h},$$

with

$$\alpha_h = \sum_{j=1}^{+\infty} \beta^{(j)}_{\left\lfloor \frac{h}{2^j} \right\rfloor} \delta^{(j)}_{h-2^j\left\lfloor \frac{h}{2^j} \right\rfloor}.$$

Reconstruction Theorem Remarks

- Different selection of $\delta^{(j)}_{h-2^j\left\lfloor \frac{h}{2^j}\right\rfloor}$ select different renormalization schemes that impact EWD and IRf.
- These coefficients are economically determined by the information flows along the resolution filtration.
- Open Question: How competition shapes this flow in financial markets?
- Randomizes allocations of weights can be used to generate intermittency and clustering. Close to **Kahane** and **Peyriere** random cascade models, **Mandelbrot Calvet** and **Fisher**'s multifractal volatility.

Stochastic (log) volatility modelling

- Levy construction of the BM is that continuous time limit of the EWD generates the class of Brownian SemiStationary processes

$$x_t = \lim_{J \to +\infty} \sum_{j=-J}^{+\infty} x_t^{(j)} = \lim_{J \to +\infty} \sum_{j=-J}^{+\infty} \sum_{k=0}^{+\infty} \beta_k^{(j)} \varepsilon_{t-k2^j}^{(j)} = \int_{-\infty}^{t} g\left(t-s\right) dW_s$$

- EWD Forecasting formulas:

$$E_t\left[x_{t+\Delta}\right] = E_t\left[\sum_{j=-\infty}^{+\infty} x_{t+\Delta}^{(j)}\right] = \sum_{j=-\infty}^{+\infty} \sum_{k=0}^{+\infty} \beta_{k,\Delta}^{(j)} \varepsilon_{t-k2^j}^{(j)}$$

Data and Realized Volatility

- Tick-by-tick series of USD/CHF exchange rate.
- Range: Dec1998 to Dec2003.
- Spot logarithmic middle prices computed (by Corsi) as averages of log bid and ask quotes.
- Returns are used to estimate daily realized volatility, as in Andersen, Bollerslev Diebold and Labys (2003):

$$d_t = \sqrt{\sum_{j=0}^{M-1} r_{t-j/M}^2}, \qquad M = 12.$$

Persistence-based forecasting

- The Corsi forecasting equation exploits short-term lags of Realized Volatility:

$$d_{t+1} = a_0 + a_d d_t + a_w w_t + a_m m_t + \nu_t. \tag{1}$$

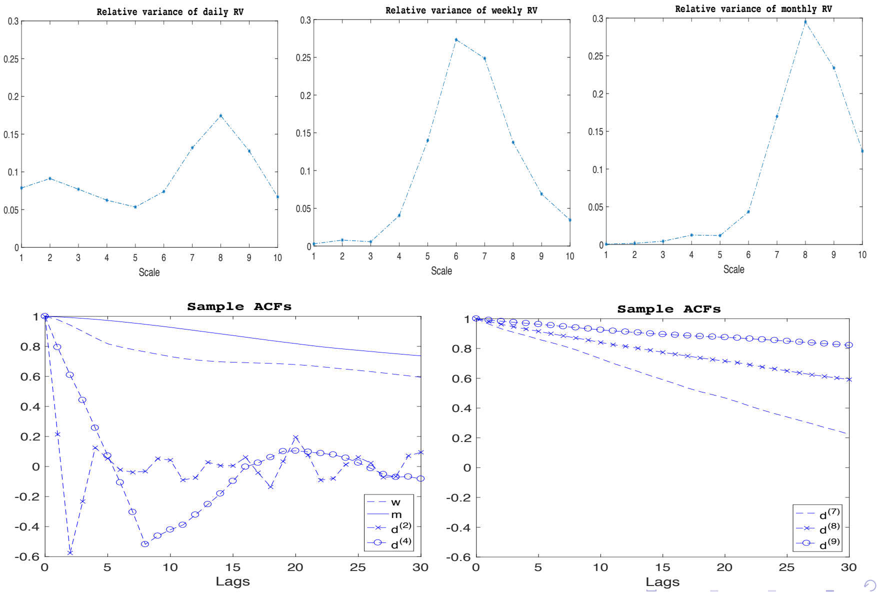
- Forecasting method based on the persistent components of d_t that explain the most variance:

$$d_{t+1} = a^{(0)} + a^{(7)} \mathbb{E}_t \left[d_{t+1}^{(7)} \right] + a^{(8)} \mathbb{E}_t \left[d_{t+1}^{(8)} \right] + a^{(9)} \mathbb{E}_t \left[d_{t+1}^{(9)} \right] + \xi_t.$$

- These three scales explain 44,6% of total variance. Same forecasting power as Equation (1), but uncorrelated persistent components.

	RMSE	MAE	R^2
HAR-RV	2.607	1.757	0.565
Extended Wold (3)	2.537	1.788	0.494
Extended Wold (10)	2.292	1.556	0.588

Variance decomposition of Realized Volatility



Variance decomposition: remarks

- Volatility persistence is associated with the heterogeneous information arrivals in the market (Andersen and Bollerslev 1997) and with the presence of heterogeneous degree of persistence of information based trading (Müller et al. 1997)
- Evidence of the Heterogeneous Market Hypothesis. Data allows the estimation of 10 uncorrelated scales, which overall explain roughly 95% of total sample variance.
- Persistence of the shocks: Shocks with degree of persistence associated to scales 7, 8 and 9 which involve that last 128, 256 and 512 working days explain most of the variance variability.

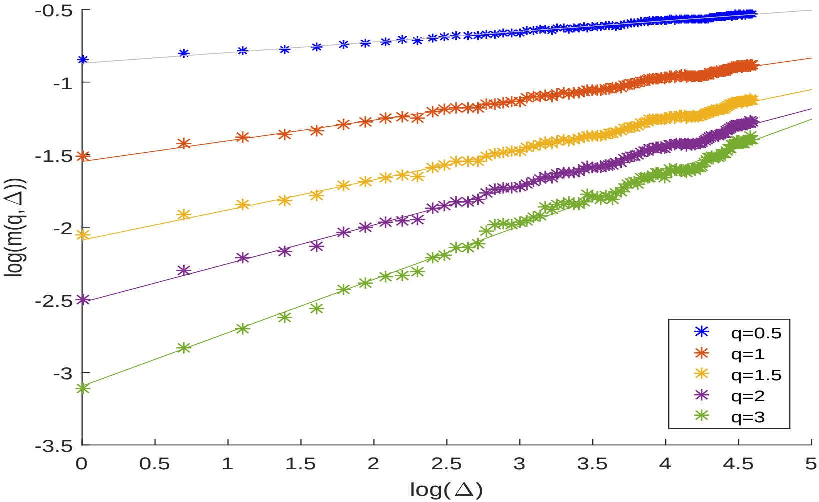
Structural vs Descriptive interpretation of the EWD

An observation from a smart but inattent Referee:

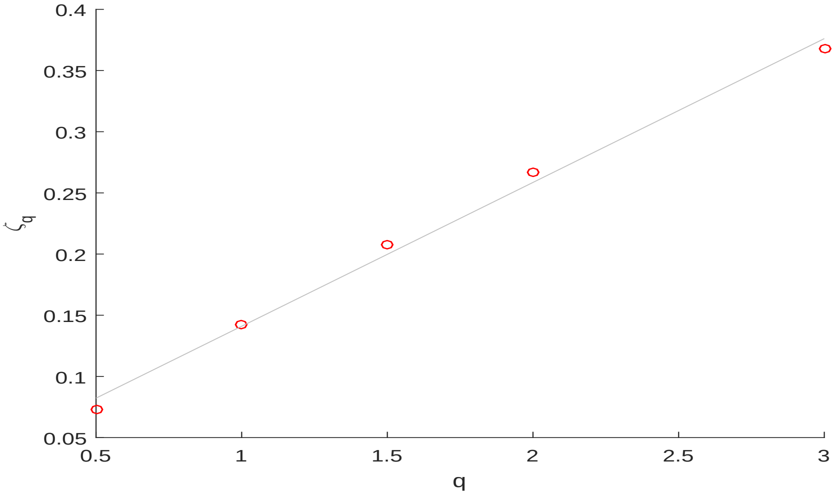
'In order to show the lack of structural interpretation, assume that the data generating process is at a daily frequency. One could either observe the data at a daily frequency or at a weekly one. Each frequency will lead to a different DWT and I think that they are not strongly related, that is for each scale (or horizon) the variables will be quite different.'

MAIN TAKEAWAY: A necessary condition for the decomposition to have a structural interpretation is a scale invariant aggregation scheme, i.e. the existence of an RG fixed point.

Scale Invariance-1



Scale Invariance-2



Rough Volatility Cascade Model

- "Volatility is rough" statement must be interpreted as the empirical observation that the macroscopic log-volatility dynamics is invariant w.r.t to a suitable RG scheme in the high frequency limit.
- This Prescription is sufficient to computation of the log-vol IRf.

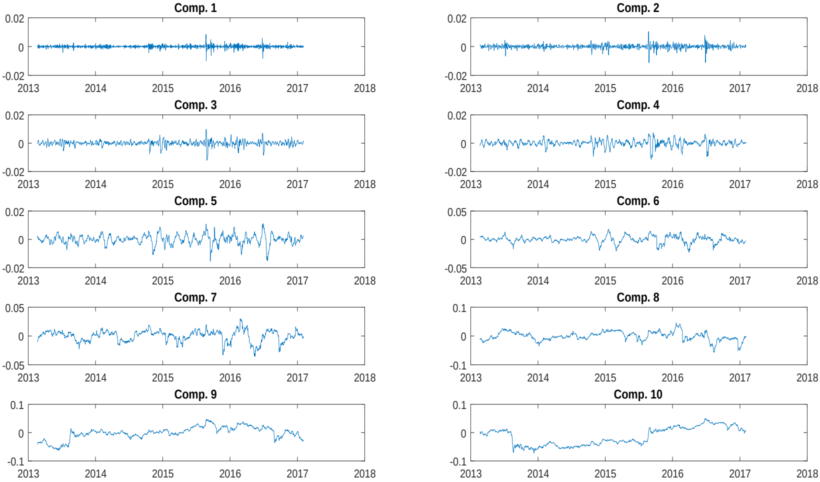
$$\mathbf{R}^H : \quad \sum_{k=0}^{+\infty} a_k \varepsilon_{t-k} \quad \longmapsto \quad \sum_{k=0}^{+\infty} \frac{a_k}{2^H} \left(\varepsilon_{t-2k} + \varepsilon_{t-2k-1}\right)$$

will generate a properly defined EWD

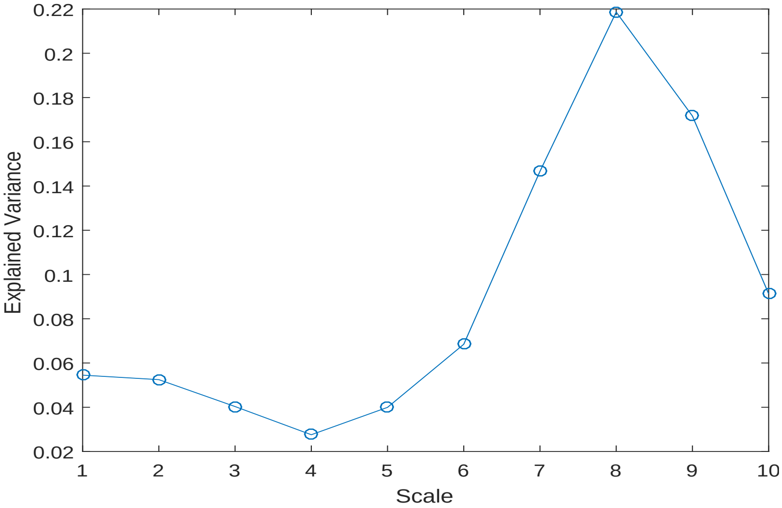
$$\log \sigma_t = \sum_{j=1}^{+\infty} g_t^{(j)} = \sum_{j=1}^{+\infty} \sum_{k=0}^{+\infty} \beta_k^{(j)} \varepsilon_{t-k2^j}^{(j)}.$$

- The model in continuous time would look like the cascade model of Calvet Fisher and Wu for interest rates with parametric restrictions induced by the "Roughness Hypothesis" .

RCEWD-1

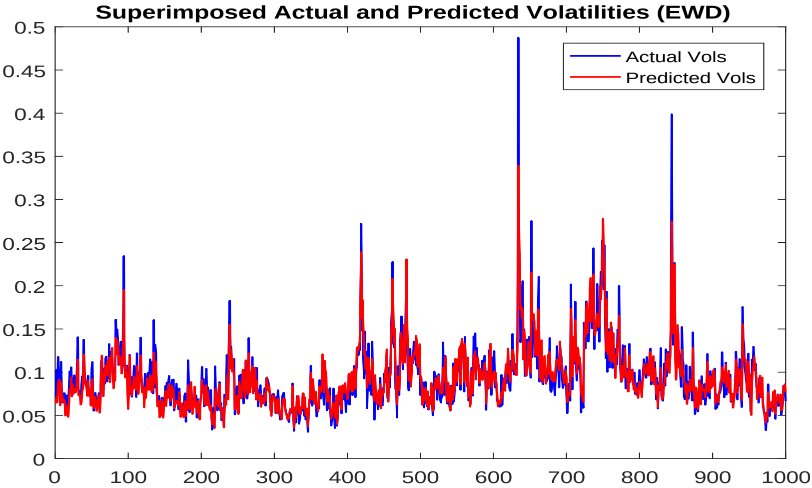

Comp. 1
Comp. 2
Comp. 3
Comp. 4
Comp. 5
Comp. 6
Comp. 7
Comp. 8
Comp. 9
Comp. 10
0.02
0
-0.02
0.05
-0.05
0.1
-0.1
2013
2014
2015
2016
2017
2018

RCEWD-2



Explained Variance
Scale
0.22
0.2
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
1
2
3
4
5
6
7
8
9
10

RCEWD-3



Superimposed Actual and Predicted Volatilities (EWD)
Actual Vols
Predicted Vols
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0
100
200
300
400
500
600
700
800
900
1000

Conclusions and Future Developments

- Non-parametric discrimination among different Rough Vol Models
- A potential resolution of the long-term excess volatility puzzle and rough volatility cascade model.
- Ross recovery problem and Rough Volatility cascade model: better understanding of the role of risk neutral vs historical measure
- Extension to the fractional case of the construction of the non-linear IRf extension Gaussian Stochastic Calculus of Variations and application to Option Hedging?

Barolo is good but Amarone is not bad ...

Happy Birthday Gino!

The decomposition of $\mathcal{H}_t(\mathbf{x})$ induced by $\mathbf{L}$

- For any $j \in \mathbb{N}$, we have $\mathbf{L}^j \mathcal{H}_t(\mathbf{x}) = \mathcal{H}_{t-j}(\mathbf{x})$ and so

$$\hat{\mathcal{H}}_t(\mathbf{x}) = \bigcap_{j=0}^{+\infty} \mathcal{H}_{t-j}(\mathbf{x}).$$

- The wandering subspace is $\mathcal{L}_t^{\mathbf{L}} = \operatorname{span}\left\{x_t - \mathcal{P}_{\mathcal{H}_{t-1}(\mathbf{x})} x_t\right\}$ and

$$\mathbf{L}^j \mathcal{L}_t^{\mathbf{L}} = \operatorname{span}\left\{x_{t-j} - \mathcal{P}_{\mathcal{H}_{t-j-1}(\mathbf{x})} x_{t-j}\right\}.$$

As a result,

$$\tilde{\mathcal{H}}_t(\mathbf{x}) = \bigoplus_{j=0}^{+\infty} \operatorname{span}\left\{x_{t-j} - \mathcal{P}_{\mathcal{H}_{t-j-1}(\mathbf{x})} x_{t-j}\right\}.$$

Comparison with the multiresolution approach - 1

We compare the scaling operator $\mathbf{R} : \mathcal{H}_t(\varepsilon) \longrightarrow \mathcal{H}_t(\varepsilon)$

$$\mathbf{R} : \quad \sum_{k=0}^{+\infty} a_k \varepsilon_{t-k} \quad \longmapsto \quad \sum_{k=0}^{+\infty} \frac{a_k}{\sqrt{2}} \left(\varepsilon_{t-2k} + \varepsilon_{t-2k-1}\right) = \sum_{k=0}^{+\infty} \frac{a_{\lfloor \frac{k}{2} \rfloor}}{\sqrt{2}} \varepsilon_{t-k}$$

with the operator underlying OTT multiresolution-based decomposition $\mathbf{R}_{\mathbf{x}} : \mathcal{S}_t(\mathbf{x}) \longrightarrow \mathcal{S}_t(\mathbf{x})$

$$\mathbf{R}_{\mathbf{x}} : \quad \sum_{k=0}^{N} a_k x_{t-k} \quad \longmapsto \quad \sum_{k=0}^{N} \frac{a_k}{\sqrt{2}} \left(x_{t-2k} + x_{t-2k-1}\right) = \sum_{k=0}^{2N+1} \frac{a_{\lfloor \frac{k}{2} \rfloor}}{\sqrt{2}} x_{t-k}$$

where $\mathcal{S}_t(\mathbf{x})$ is the subspace of $\mathcal{H}_t(\mathbf{x})$ of all finite linear combinations of variables x_{t-k}.

Comparison with the multiresolution approach - 2

- In case $\mathbf{x}$ is purely non-deterministic, $\mathcal{S}_t(\mathbf{x}) \subset \mathcal{H}_t(\varepsilon)$.
- Therefore, both $\mathbf{R}$ and $\mathbf{R_x}$ act on x_t, providing two different decompositions.
  - $\mathbf{R}$ is isometric and delivers the Extended Wold Decomposition;
  - $\mathbf{R_x}$ in general is not isometric and provides a multiresolution- based decomposition that does not rule out correlation across scales.
- The main difference between the two decompositions is that the lag operator and the scaling operator do not commute:

$$\mathbf{RL} = \mathbf{L}^2\mathbf{R}.$$

Comparison with the multiresolution approach - 3

Assume that $\lim_n \gamma(n) = 0$.

- In the Extended Wold Decomposition,

$$x_t = \sum_{j=1}^{+\infty} g_t^{(j)} \qquad \text{with} \qquad g_t^{(j)} = \sum_{k=0}^{+\infty} \beta_k^{(j)} \varepsilon_{t-k2^j}^{(j)}.$$

- In the multiresolution-based decomposition,

$$x_t = \sum_{j=1}^{+\infty} \breve{g}_t^{(j)} \qquad \text{with} \qquad \breve{g}_t^{(j)} = \sum_{h=0}^{+\infty} \frac{\alpha_h}{\sqrt{2^j}} \varepsilon_{t-h}^{(j)}.$$