Long Memory and Roughness in *Stochastic Volatility Models*

Alexandra Chronopoulou

Industrial & Enterprise Systems Engineering
University of Illinois at Urbana–Champaign

1^{st} Eastern Conference on Mathematical Finance

March 20^{th}, 2016

Classical Stochastic Volatility Model

- Log-Returns: $\{Y_t, t \in [0, T]\}$

$$dY_t = \left(r - \frac{{\sigma_t}^2}{2}\right)\, dt + \sigma_t\, dW_t.$$

where $\sigma_t = \sigma(X_t)$ and X_t is a non-observable diffusion, such as:

  - *Log-Normal*: $dX_t = c_1\, X_t\, dt + c_2\, X_t\, dZ_t$.
  - *Mean Reverting OU*: $dX_t = \alpha\, (m - X_t)\, dt + \beta\, dZ_t$.
  - *Feller or Cox–Ingersoll–Ross (CIR)*: $dX_t = k\, (\nu - X_t)\, dt + v\sqrt{X_t}\, dZ_t$.

Fractional Stochastic Volatility Model

Log-Returns: $\{Y_t, t \in [0, T]\}$

$$dY_t = \left(r - \frac{{\sigma_t}^2}{2}\right) dt + \sigma_t \, dW_t.$$

where $\sigma_t = \sigma(X_t)$ and X_t is described by:

$$dX_t = \alpha \, (m - X_t) \, dt + \beta \, dB_t^H$$

- B_t^H is a fractional Brownian motion with Hurst parameter $H \in (0, 1)$.
- X_t is a fractional Ornstein-Uhlenbeck process (fOU).

Fractional Stochastic Volatility Model

$$\begin{cases} dY_t &= \left(r - \frac{\sigma^2(X_t)}{2}\right)\, dt + \sigma(X_t)\, dW_t, \\ dX_t &= \alpha\,(m - X_t)\, dt + \beta\, dB_t^H \end{cases}$$

Some Properties

- *Hölder continuity*
  The fBm is Hölder continuous of order γ, for all $\gamma < H$.
  This property is inherited by the fOU.

- *Self-similarity*
  The fBm is self-similar in the sense that

$$\{B_{ct}^H; t \in \mathbb{R}\} \sim^{\mathcal{D}} \{c^H B_t^H; t \in \mathbb{R}\}, \quad \forall c \in \mathbb{R}$$

  This property is approximately inherited by the fOU, for scales smaller than $1/\alpha$.

Fractional Stochastic Volatility Model

Hurst Index	Model
$H > 1/2$	Long-Range Dependence: *persistence* ACF Decay $\sim dt^{2H-2}$
$H < 1/2$	Short-Range Dependence: *anti-persistence* ACF Decay $\sim dt^{H}$
$H = 1/2$	Classical SV

Long Memory Stochastic Volatility Models

- Comte and Renault (1998) : LMSV with volatility described by a fractional Ornstein-Uhlenbeck process.
- Comte, Coutin and Renault (2010) : Affine fractional models.
- C. and Viens (2010) : Pricing under the LMSV models .
- C. and Viens (2012) : Calibration of discrete and continuous time LMSV models.
- Gulisashvili, Viens and Zhang (2015) : Small time asymptotcis for LMSV models.
- Guennoun, Jacquier, and Roome (2015) : Asymptotic behavior of the LMSV model.

Rough Stochastic Volatility Models

- Gatheral, Jaisson, and Rosenbaum (2014): Rough volatility models.
- Bayer, Friz, and Gatheral (2015): Rough fractional stochastic volatility models.

Related Literature

- Willinger, Taqqu, Teverovsky (1999): LRD in the stock market
- Bayraktar, Poor, Sircar (2003): Estimation of fractal dimension of S&P 500.
- Björk, Hult (2005): Fractional Black-Scholes market.
- Cheriditio (2003): Arbitrage in fractional BS market.
- Guasoni (2006): No arbitral under transaction cost with fBm.

Inference for FSV Models

$$\begin{cases} dY_t &= \left(r - \frac{\sigma^2(X_t)}{2}\right) dt + \sigma(X_t)\, dW_t, \\ dX_t &= \alpha\,(m - X_t)\, dt + \beta\, dB_t^H \end{cases}$$

- Parameters to estimate: α, m, β, μ (for hedging) and H.

Two scenarios

Observation

The estimation of H is decoupled from the estimation of the drift components, but not from the estimation of the "diffusion" terms.

- *Scenario 1*: Assume the Hurst Index is known.
- *Scenario 2*: Assume the Hurst index is not known.

Goal

- Given H, how do we estimate the remaining parameters of the model?

Framework

- Observations:
  Historical stock prices $\rightarrow$ Discrete, even when in high-frequency.
- Unobserved State:
  Stochastic Volatility with non-Markovian structure is hidden.

Employ a Sequential Monte Carlo (SMC) method to estimate the unobserved state along with the unknown parameters.

Non-Markovian dynamic models

Denote θ the vector of all parameters, except for H.

- Observation equation
$$f(Y_t|X_t;\theta)$$
- State equation
$$f(X_t|X_{t-1},\ldots,X_1;\theta)$$
- Initial distribution
$$f(X_0|\theta)$$

Filtering for states & parameters

Non-Markovian Case

$$
\begin{aligned}
& f(X_{1:t};\theta|Y_{1:t}) \\
\propto\ & f(X_{1:t}, Y_{1:t}, \theta) \\
\propto\ & f(X_1)\cdot f(X_2|X_1;\theta)\cdot\ldots\cdot f(X_n|X_{n-1},\ldots,X_1;\theta)\cdot\prod_{i=1}^{t} f(Y_i|X_i;\theta)\cdot f(\theta|Y_t),
\end{aligned}
$$

where $f(\theta|Y_t)$ is a prior density for the parameter vector θ.

- If the parameter is known, then the density is degenerate.
- Otherwise, we need to compute or approximate the theoretical density function $f(\theta|Y_t)$.

Learning θ offline

Two step strategy:

1. Approximate $f(\theta|Y)$ by

$$f^N(\theta|Y) = \frac{f^N(Y|\theta)f(\theta)}{p(y)} \propto f^N(Y|\theta)f(\theta)$$

   where $f^N(Y|\theta)$ is a SMC approximation to $f(Y|\theta)$.

2. Sample θ via an MCMC scheme or an SIR scheme.

Disadvantages

1. SMC looses its appealing sequential nature.
2. Overall sampling scheme is sensitive to $p^N(Y|\theta)$

Learning θ sequentially

Filtering for states and parameter(s): *Learning X_t and θ sequentially*.

$$\text{Posterior at } t : p(X_t|\theta, Y_t)\; p(\theta|Y_t)$$

$$\Downarrow$$

$$\text{Prior at } t+1 : p(X_{t+1}|\theta, Y_t)\; p(\theta|Y_t)$$

$$\Downarrow$$

$$\text{Posterior at } t+1 : p(X_{t+1}|\theta, Y_{t+1})\; p(\theta|Y_{t+1})$$

Advantages

1. Sequential updates of $f(\theta|Y_t)$, $f(X_t|Y_t)$ and $f(\theta, X_t|Y_t)$.
2. Sequential h-step ahead forecasts $f(Y_{t+h}|Y_t)$
3. Sequential approximations for $f(Y_t|Y_{t-1})$.

Artificial Evolution of θ

- Gordon et al (1993)

$$\theta_{t+1} = \theta_t + \zeta_{t+1}$$
$$\zeta_{t+1} \sim \mathcal{N}(0, W_{t+1})$$

- West et al. (1993, 1998)
  - $\forall t$ update $f(\theta|Y_t)$
  - Compute a Monte Carlo approximation of $f(\theta|Y_t)$, by using samples $\theta_t^{(j)}$ and weights $w_t^{(j)}$.
  - Smooth kernel density approximation

$$f(\theta|Y_t) \approx \sum_{j=1}^{N} \omega_t^{(j)} \mathcal{N}(\theta|m_t^{(j)}, h^2 V_t)$$

Parameter Learning Algorithm

- At time $t = 1$
  - Sample $\tilde{X}_{1,1}^{(i)} \sim\ q_1(\cdot)$ and $\theta_{(1)}^{(i)} \sim p_1(\cdot)$.
  - Compute the weights $w_1\left(\tilde{X}_{1,1}^{(i)};\ \theta_{(1)}^{(i)}\right) \propto \frac{p\left(\theta_{(1)}^{(i)}\right)\mu\left(X_{1,1}^{(i)}|\theta_{(1)}^{(i)}\right)\ p\left(y_1\middle|\tilde{X}_{1,1}^{(i)}\right)}{q_1\left(\tilde{X}_{1,1}^{(i)}\right)}$ andnormalize $W_1^{(i)} = \frac{w_1^{(i)}}{\sum_{i=1}^{N} w_1^{(i)}}$.
  - Resample to obtain $X_{1,1}^{(i)} \sim \sum_{j=1}^{N} W_1^{(j)}\delta_{\tilde{X}_{(1,1)}^{(j)}}\ (dx_1)$.

Parameter Learning Algorithm

- At time $t, t \geq 2$ (step $t-1 \rightarrow t$)
  - Set

$$\tilde{X}^{(i)}_{t,1:t-1} = X^{(i)}_{t-1,1:t-1},$$
$$m_{t-1} = \alpha\theta^{(i)}_{(t-1)} + (1-\alpha)\bar{\theta}_{(t-1)}$$

    and sample

$$\theta^{(i)}_{(t)} \sim \mathcal{N}(m^{(i)}_{t-1} | h^2 V_{t-1}),$$
$$\tilde{X}^{(i)}_{t,t} \sim q_t\left(\cdot \middle| X^{(i)}_{1:t-1,t-1}; \theta^{(i)}_{(t)}\right)$$

  - Compute the weights $w^{(i)}_t = w_t\left(X^{(i)}_{1:t-1,t-1}, \tilde{X}^{(i)}_{t,t};\ \theta^{(i)}_{(t)}\right)$ and normalize .
  - Resample
    $X_{1:t,t} \sim \pi^N_t(dx_{1:t})$, where $\pi^N_t(dx_{1:t}) = \sum_{j=1}^N W^{(j)}_t \delta_{X^{(j)}_{1:t-1,t-1}, \tilde{X}^{(j)}_{t,t}}(dx_{1:t})$
    and set $\bar{\theta}_{(t)} = \sum_{i=1}^N W^{(i)}_t \theta^{(i)}_{(t)}$.

Parameter Learning Algorithm

Output The filtering distribution $p(dx_{1:t}|y_{1:t})$ is approximated by

$$\pi^N(dx_{1:t}) = \sum_{j=1}^{N} W_t^{(j)} \delta_{X^{(j)}_{1:t-1,t-1}, \tilde{X}^{(j)}_{t,t}}(dx_{1:t}), \quad \text{or} \quad \tilde{\pi}^N(dx_{1:t}) = \frac{1}{N}\sum_{j=1}^{N} \delta_{X^{(j)}_{1:t,t}}(dx_{1:t}).$$

and the estimator for θ is

$$\bar{\theta}_{(t)} = \sum_{i=1}^{N} W_t^{(i)} \theta_{(t)}^{(i)}.$$

Filter convergence

Let $\phi : \mathcal{X} \mapsto \mathbb{R}$ be an appropriate test function and assume that we want estimate

$$\bar{\phi}_t = \int \phi_t(x_{1:t}) p(x_{1:t}, \theta_{(t)} | Y_{1:t}) dx_{1:t} d\theta_{(t)}.$$

The SISR algorithm provides us with the estimator

$$\hat{\phi}_t^N = \int \phi_t(x_{1:t}) \pi^N(dx_{1:t}) = \sum_{i=1}^{N} W_t^i \phi_t \left(X_{1:t-1,t-1}^{(i)}, \tilde{X}_{t,t}^{(i)} \right)$$

CLT for the filter (C. and Spiliopoulos)

$$\sqrt{N} \left(\hat{\phi}_t^N - \bar{\phi}_t \right) \Rightarrow \mathcal{N} \left(0, \sigma^2(\phi_t) \right)$$

as $N \to \infty$.

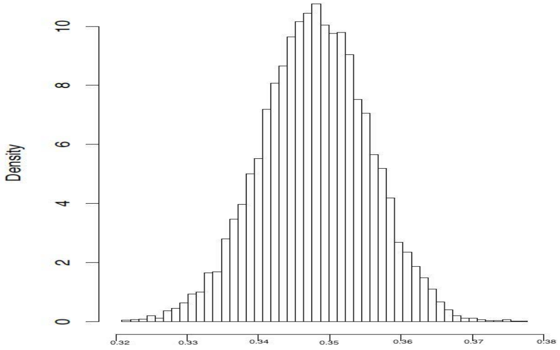
Real Data Example

- S& P 500 Data: 252 observations, starting in January 2010 until December 2010
- Model: Fractional ARIMA$(1, d, 1)$ model

$$\begin{cases} Y_t = \sigma\left(\frac{X_t}{2}\right)\ \epsilon_t \\ (1-\varphi B)\,(1-B)^d\ X_t = \vartheta\ \eta_{t-1} + \eta_t, \end{cases}$$

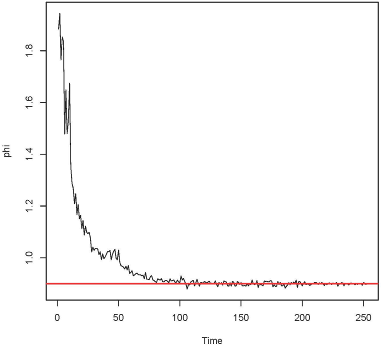
- The long-memory parameter d for the particular data set is estimated to be 0.2 using the GPH (Geweke and Porter-Hudak) method.
- We apply the SISR algorithm to estimate:
  1. the unobserved volatility distribution, and
  2. the remaining unknown parameters of the model φ and ϑ

Output 1: *Volatility Particle Filter*

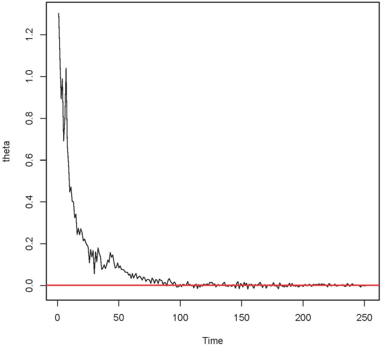


(*) As a reference, the implied volatility is computed to be around 0.345.

Output 2: *Parameter Estimators*

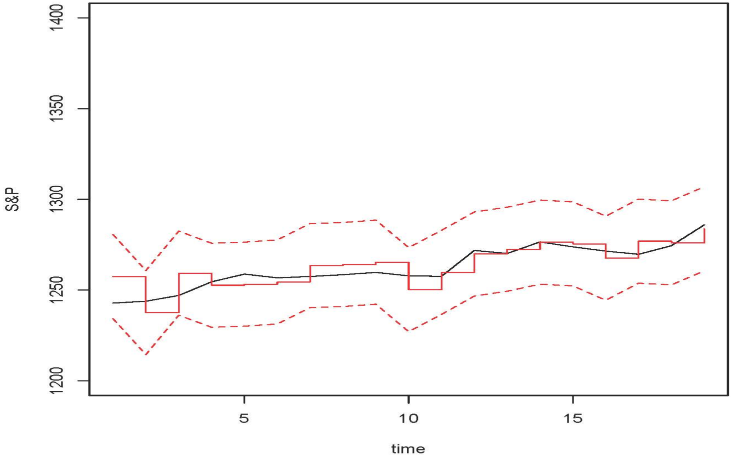


(a) Estimator of φ

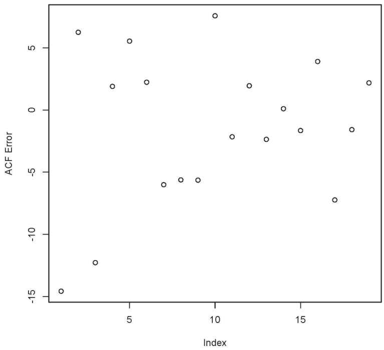


(b) Estimator of ϑ

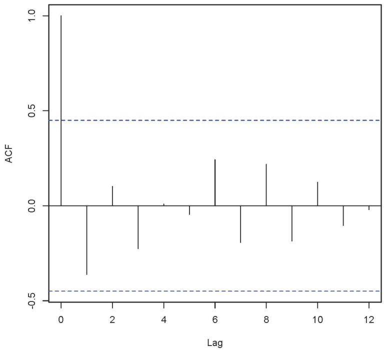
Model Validation: *1-Step Ahead Prediction*



Model Validation: *Residuals*



(a) Residuals



(b) ACF of Residuals

Continuous time model

Challenge

- Computational complexity increases significantly.

Suggestion

- Take advantage of the exact solution of the fractional OU

$$X_t = \beta \int_{-\infty}^{t} e^{-\alpha m(t-u)} dB_u^H$$

and the fact that it is a Gaussian process with known mean and variance. So, we can directly sample from its distribution, and not recursively simulate it using a discretization scheme.

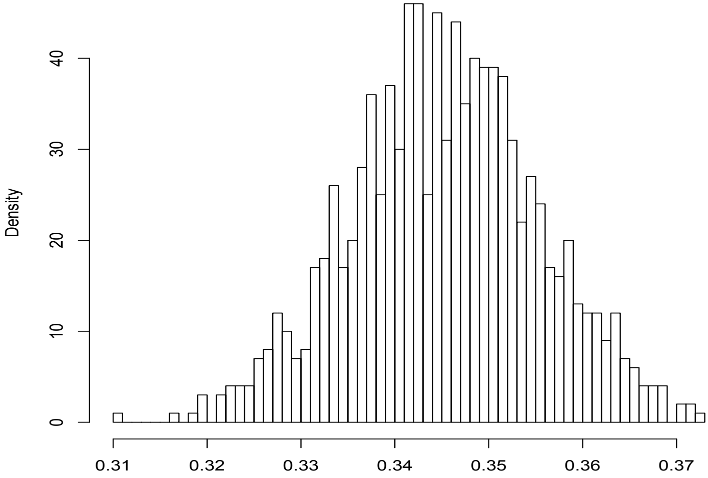
Real Data Example

- S& P 500 Data: 252 observations, starting in January 2010 until December 2010
- Fractional Ornstein-Uhlenbeck Model

$$\begin{cases} dY_t &= \left(\mu - \frac{\sigma^2(X_t)}{2}\right)\, dt + \sigma(X_t)\, dW_t, \\ dX_t &= \alpha\,(m - X_t)\, dt + \beta\, dB_t^H \end{cases}$$

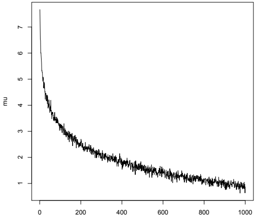
- The long-memory parameter H for the particular data set is estimated to be 0.55 using the GPH (Geweke and Porter-Hudak) method.
- We apply the SISR algorithm to estimate:
  1. the unobserved volatility distribution, and
  2. the remaining unknown parameters of the model, μ, α, m, β.

Output 1: *Volatility Particle Filter*

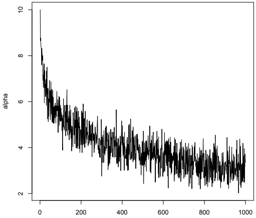


(*) As a reference, the implied volatility is computed to be around 0.345

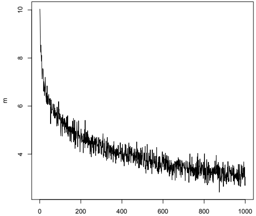
Output 2: *Parameter Estimators*



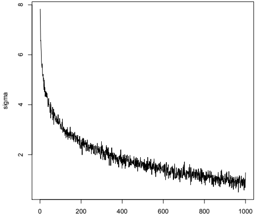
(a) Estimator of μ



(b) Estimator of α



(c) Estimator of *m*



(d) Estimator of β

How do we “learn” H?

Model Calibration

1. Use GPH estimator for an initial estimator for H; $\hat{H}_{GPH}$.
2. Choose a range of values of H around $\hat{H}_{GPH}$ and repeat:
   (a) Employ a SMC methods to estimate the model parameters.
   (b) Use a tree-based method to price a liquid option for a range of K's.
   (c) Choose the value of H that minimizes the distance of the computed option prices in (2b) from the center of the bid-ask spread; $\hat{H}_{imp}$.
3. Use $\hat{H}_{impl}$ for the SMC algorithm.

S&P 500: Implied H

Strike	Bid	Ask	H=0.5	H=0.51	H=0.58	H=0.6	H=0.62	H=0.7
750	97.5	100	92.88	92.61	**93.53**	92.60	92.79	92.41
760	89.9	92.4	85.95	85.68	**85.70**	85.67	85.86	85.49
770	82.6	85	79.32	79.05	**79.99**	79.05	79.23	78.88
780	75.5	77.8	73.02	72.74	**73.69**	72.74	72.92	72.58
790	68.6	71	67.02	66.75	**66.70**	66.76	66.93	66.60
800	61.9	64.4	61.35	61.09	**61.05**	61.10	61.26	60.95
810	55.6	58.2	56.01	55.75	**55.71**	55.76	55.92	55.62
820	49.5	52.3	50.98	50.74	**50.71**	50.75	50.90	50.62
830	43.8	46.7	46.28	46.04	**46.02**	46.06	46.20	45.94
840	38.6	41.4	41.89	41.66	**41.65**	41.68	41.82	41.57
850	33.5	36.4	37.82	37.60	**37.59**	37.61	37.74	37.51
860	28.9	31.8	34.04	33.83	**33.82**	33.85	33.97	33.76

$$\hat{H}_{implied} = 0.58$$

Discussion

- Of course, we cannot assume that the volatility is constant over a long period of time.
- To determine how much data we need to properly calibrate the model wrt H, we use the following rule:

  If trying to price an option one month ahead, one should essentially look at data over the last month. If trying to price an option one year ahead, one should essentially look at data over the last year.

How about Leverage Effects

- Fractional Ornstein-Uhlenbeck Model

$$\begin{cases} dY_t &= \left(\mu - \frac{\sigma^2(X_t)}{2}\right)\, dt + \sigma(X_t)\, dW_t, \\ dX_t &= \alpha\,(m - X_t)\, dt + \beta\, dB_t^H \end{cases}$$

where

$$Corr(B_t^H, W_t) = \rho$$

Question: How to estimate ρ?

Leverage Effects in a LMSV

Definition

Consider a function F twice differentiable and monotone on $(0, \infty)$. The leverage effect in the LMSV model is defined as the quadratic co-variation between Y_t and $F(X_t)$:

$$\left\langle Y, F(X_t^2) \right\rangle = 2\, \rho\, \beta \int_0^T 2F'(X_t^2)\, X_t^2\, (dt)^{H+\frac{1}{2}}$$

Variations-based Estimator for ρ

Estimator $\hat{\rho}_n$

A variations-based estimator of the correlation ρ is defined as

$$\hat{\rho}_n = \frac{n^{-(H+\frac{1}{2})} \sum_{i=1}^{n} v_i}{2\beta \, T^{H+\frac{1}{2}}},$$

where

$$v_i := \sum_{i=1}^{n} \Big\{ [Y_{i+1} - Y_i] \cdot \big[F(X_{i+1}^2) - F(X_i^2)\big] \Big\}$$

Consistency of $\hat{\rho}_n$

Theorem (C. and Viens)

The variations-based estimator for the correlation coefficient ρ is strongly consistent, that is

$$\lim_{n\to\infty} \hat{\rho}_n = \rho, \quad \text{a.s.}$$

Idea of Proof

- Define

$$V_T^n = \sum_{i=1}^{n} \Big\{ [Y(t_{i+1}) - Y(t_i)] \cdot \big[F(X^2(t_{i+1})) - F(X^2(t_i)) \big] \Big\}.$$

- Using “Itô” for fractional SDEs, show that

$$\lim_{n \to \infty} n^{(H+\frac{1}{2})} V_T^n = 2\beta\rho\, \sigma_t^2\, F'(\sigma_t^2) \int_0^T (dt)^{H+\frac{1}{2}} = \langle Y, F(\sigma_t^2) \rangle$$

almost surely.

Central Limit Theorem

CLT for $\hat{\rho}_n$ (C. and F. Viens)

Let $Z_n = n^{H-\frac{1}{2}} \frac{\hat{\rho}_n - \rho}{\sqrt{Var(\hat{\rho}_n)}}$ and compute $G_{Z_n} = \langle DZ_n, -D(L^{-1}Z_n) \rangle_{L^2}$.
If $E[|G_{Z_n} - 1|] \to 0$, then

$$n^{H-\frac{1}{2}} \frac{\hat{\rho}_n - \rho}{\sqrt{Var(\hat{\rho}_n)}} \to^D N(0, 1)$$

as $n \to \infty$.

Conclusion

- We discussed a SMC methodology for estimating the model parameters in a SV with long-range or short-range dependent volatility process.
- We developed a calibration methodology to determine the value of H.
- We developed a variations-based estimator for the leverage effect ρ, given H.

Thank you!