# Dynamic Semiparametric Models for Expected Shortfall (and Value-at-Risk)

Andrew J. Patton     Johanna F. Ziegel     Rui Chen

*Duke University*     *University of Bern*     *Duke University*

March 2018

# Measures of market risk

- The simplest and most widely-used measure of risk is **variance**:

$$\sigma_t^2 \equiv \mathbb{E}_{t-1}\left[\left(Y_t - \mu_t\right)^2\right]$$

- In the 1990s, in part prompted by Basel I and II, attention in risk management moved to **Value-at-Risk:**

$$VaR_t \equiv F_t^{-1}\left(\alpha\right) \Rightarrow \Pr_{t-1}\left[Y_t \leq VaR_t\right] = \alpha$$

- The Basel III accord pushes banks to move from Value-at-Risk towards **Expected Shortfall:**

$$ES_t \equiv \mathbb{E}_{t-1}\left[Y_t | Y_t \leq VaR_t\right]$$

# Why the move from VaR to ES?

- Academic work has highlighted some problems with VaR (see McNeil, *et al.* 2015 for a summary):

- Value-at-Risk has some positive attributes:
    - Focuses on the **left tail** of returns, so more relevant for risk mgmt
    - **Easy** to interpret ("the loss that is only exceeded on 5% of days")
    - Is well-defined even for **fat-tailed** distributions; is a **robust** statistic

- But VaR suffers from important drawbacks (Artzner et al. 1999, *MathFin*):
    - Not "sub-additive:" diversification may make VaR look **worse**
    - **No information** about losses beyond the VaR

- Expected Shortfall addresses both of these drawbacks
    - But it is not a robust statistic, and does require moment assumptions

# Why aren't there more models for Expected Shortfall?

- To answer this, consider how we estimate and model Value-at-Risk.

- For a given sample $\{Y_t\}_{t=1}^T$, VaR can be obtained as

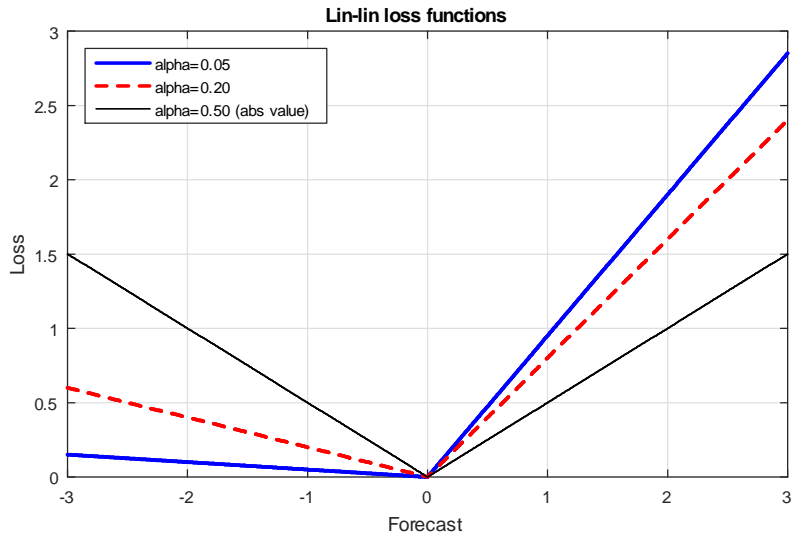$$\widehat{VaR}_T \;=\; \arg\min_v \; \frac{1}{T} \sum_{t=1}^T L\left(Y_t, v; \alpha\right)$$

$$\text{where} \;\; L\left(y, v; \alpha\right) \;=\; \left(\mathbf{1}\left\{y \leq v\right\} - \alpha\right)\left(v - y\right)$$

- The loss function here is the "tick" or "lin-lin" loss function

- Given this loss function, it is possible to consider models like "CAViaR" (Engle and Manganelli, 2004, *JBES*):

$$\hat{\boldsymbol{\theta}}_T \;=\; \arg\min_{\boldsymbol{\theta}} \; \frac{1}{T} \sum_{t=1}^T L\left(Y_t, v\left(\mathbf{Z}_{t-1}; \boldsymbol{\theta}\right); \alpha\right)$$

$$\text{and} \;\; VaR_t \;=\; v\left(\mathbf{Z}_{t-1}; \boldsymbol{\theta}\right)$$

# Why aren't there more models for Expected Shortfall?

- Given an estimator of VaR, sample Expected Shortfall can be computed as:

$$\widehat{ES}_T = \frac{1}{\alpha T} \sum\nolimits_{t=1}^{T} Y_t \mathbf{1} \left\{ Y_t \leq VaR_t \right\}$$

# Why aren't there more models for Expected Shortfall?

- Given an estimator of VaR, sample Expected Shortfall can be computed as:

$$\widehat{ES}_T = \frac{1}{\alpha T} \sum\nolimits_{t=1}^{T} Y_t \mathbf{1} \{Y_t \leq VaR_t\}$$

- But there *does not exist* an objective function such that ES is the solution:

$$\nexists \, L^* \quad \text{s.t.} \quad \widehat{ES}_T = \arg\min_{e} \, \frac{1}{T} \sum\nolimits_{t=1}^{T} L^* (Y_t, e; \alpha)$$

# Why aren't there more models for Expected Shortfall?

- Given an estimator of VaR, sample Expected Shortfall can be computed as:

$$\widehat{ES}_T = \frac{1}{\alpha T} \sum\nolimits_{t=1}^{T} Y_t \mathbf{1} \left\{ Y_t \leq VaR_t \right\}$$

- But there *does not exist* an objective function such that ES is the solution:

$$\nexists \, L^* \quad \text{s.t.} \quad \widehat{ES}_T = \arg\min_{e} \, \frac{1}{T} \sum\nolimits_{t=1}^{T} L^* \left( Y_t, e; \alpha \right)$$

- Expected Shortfall is "**non-elicitable**" (Gneiting 2011, *JASA*). This explains, perhaps, the lack of models for Expected Shortfall:

# Why aren't there more models for Expected Shortfall?

- Given an estimator of VaR, sample Expected Shortfall can be computed as:

$$\widehat{ES}_T = \frac{1}{\alpha T} \sum_{t=1}^{T} Y_t \mathbf{1} \left\{ Y_t \leq VaR_t \right\}$$

- But there *does not exist* an objective function such that ES is the solution:

$$\nexists \, L^* \quad \text{s.t.} \quad \widehat{ES}_T = \arg \min_e \ \frac{1}{T} \sum_{t=1}^{T} L^* \left( Y_t, e; \alpha \right)$$

- Expected Shortfall is "**non-elicitable**" (Gneiting 2011, *JASA*). This explains, perhaps, the lack of models for Expected Shortfall:

★ We exploit recent results in statistics and decision theory which shows that while ES is not elicitable, it is **jointly elicitable** with Value-at-Risk.

# Related literature

- A lot of work has been done on models for risk management, mostly VaR:

    - McNeil, Frey and Embrechts (2015, *Quantitative Risk Mgmt*)
    - Daníelsson (2011, *Financial Risk Forecasting*)
    - Komunjer (2010, *Handbook of Economic Forecasting*)

- This paper is closest to Engle and Manganelli (2004, *JBES*) who propose time series models for conditional quantiles, and establish conditions for estimation and inference

    - We extend their paper to consider ES (jointly with VaR)

- We draw on two distinct recent advances in the literature:

    - Statistical decision theory: Fissler and Ziegel (2016, AoS)
    - Parameter-driven time series models: Creal, Koopman & Lucas (2013, *JAE*), Harvey (2013, *book*).

# Outline

- Fissler and Ziegel (2016, AoS) show that while ES is not elicitable, it is **jointly elicitable** with VaR, using the following class of loss functions:

$$
\begin{aligned}
L(Y, v, e; \alpha) \;=\; & (\mathbf{1}\{Y \le v\} - \alpha)\left(G_1(v) - G_1(Y) + \frac{1}{\alpha}G_2(e)\,v\right) \\
& - G_2(e)\left(\frac{1}{\alpha}\mathbf{1}\{Y \le v\}\,Y - e\right) - \mathcal{G}_2(e)
\end{aligned}
$$

where

- $G_1$ is weakly increasing
- $G_2$ is strictly positive and strictly increasing, and $\mathcal{G}_2' = G_2$.

- Minimizing this loss function yields VaR and ES:

$$
[VaR_t, ES_t] = \arg\min_{(v,e)} \; \mathbb{E}_{t-1}\left[L(Y_t, v, e; \alpha)\right]
$$

# Joint estimation of VaR and Expected Shortfall: FZ0 loss

- We will use a **homogeneous of degree zero** FZ loss function

- This loss function yields loss function *differences* (between two competing sets of VaR and ES forecasts) thare homogeneous of degree zero.

  - This property has been found to have good finite-sample features in volatility applications, see Patton and Sheppard (2009).

- For the values of $\alpha$ of interest we know $ES_t < 0$, and we show that such an FZ loss is **unique** (up to location/scale constants):

$$L_{FZ0}\left(Y, v, e; \alpha\right) = -\frac{1}{\alpha e}\mathbf{1}\left\{Y \leq v\right\}\left(v - Y\right) - \frac{1}{e}\left(e - v\right) + \log\left(-e\right)$$

  - where $Y$ is the (future) return, $v$ is the VaR forecast, and $e$ is the ES forecast.

# The FZ0 loss function

The implied VaR loss is the familiar "tick" loss function; the implied ES loss resembles "QLIKE"
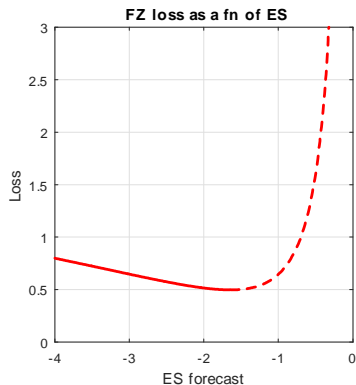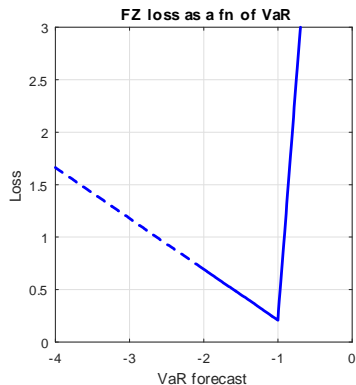
# The expected FZ0 loss function

for a N(0,1) target variable. Contours are convex.



Expected FZ0 loss for a standard Normal variable
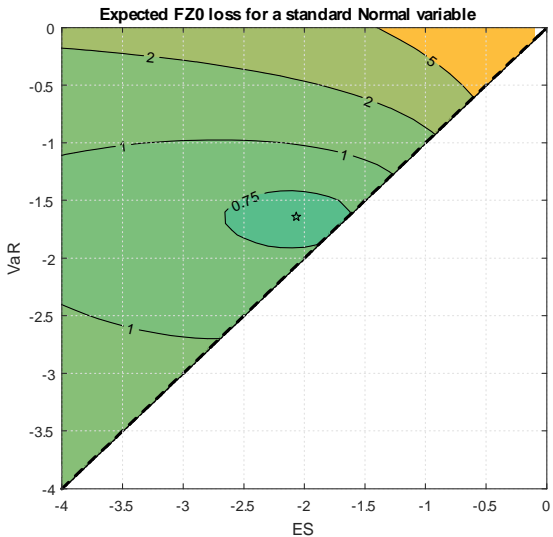
# Dynamic models for ES and VaR

- With a loss function available, it is possible to consider dynamic models for ES and VaR:

$$
\begin{aligned}
VaR_t &= v\left(\mathbf{Z}_{t-1}; \boldsymbol{\theta}\right) \\
ES_t &= e\left(\mathbf{Z}_{t-1}; \boldsymbol{\theta}\right)
\end{aligned}
$$

- The parameters of this model can then be obtained as:

$$
\hat{\boldsymbol{\theta}}_T = \arg\min_{\boldsymbol{\theta}} \frac{1}{T} \sum_{t=1}^{T} L\left(Y_t, v\left(\mathbf{Z}_{t-1}; \boldsymbol{\theta}\right), e\left(\mathbf{Z}_{t-1}; \boldsymbol{\theta}\right)\right)
$$

- We propose some new models for ES (and VaR), drawing on recent research, and then provide theory for estimation and inference for these models.

- Creal et al. (2013, *JAE*) proposed "generalized autoregressive score" models for time-varying density models:

$$Y_t | \mathcal{F}_{t-1} \quad \sim \quad F(\boldsymbol{\theta}_t)$$

$$\boldsymbol{\theta}_t \quad = \quad \mathbf{w} + \mathbf{B} \cdot \boldsymbol{\theta}_{t-1} + \mathbf{A} \cdot \mathbf{S}_{t-1} \frac{\partial \log f(y_{t-1}; \boldsymbol{\theta}_{t-1})}{\partial \boldsymbol{\theta}}$$

- This model was independently proposed as a "dynamic conditional score" (DCS) model in Harvey (2013, *book*).

- Using the score ($\partial \log f / \partial \boldsymbol{\theta}$) as the "forcing variable" enables them to nest many existing models, including ARMA and GARCH models.

   - The "scale" matrix, $\mathbf{S}_{t-1}$, is often set to the inverse Hessian.

- This choice of forcing variable can be motivated as the Newton-Raphson step in a numerical optimization algorithm.

# GAS models for dynamic ES and VaR II

- We adopt this modeling approach, and apply it to our M-estimation problem.

- Consider the following GAS(1,1) specification for VaR and ES:

$$
\left[ \begin{array}{c} v_{t+1} \\ e_{t+1} \end{array} \right] = \mathbf{w} + \mathbf{B} \left[ \begin{array}{c} v_t \\ e_t \end{array} \right] + \mathbf{A} \left( \frac{\partial^2 \mathbb{E}_{t-1} \left[ L\left( Y_t, v_t, e_t \right) \right]}{\partial \left( \mathbf{ve} \right) \partial \left( \mathbf{ve} \right)'} \right)^{-1} \frac{\partial L\left( Y_t, v_t, e_t \right)}{\partial \left( \mathbf{ve} \right)}
$$

$$
= \mathbf{w} + \mathbf{B} \left[ \begin{array}{c} v_{t-1} \\ e_{t-1} \end{array} \right] + \mathbf{A} \left[ \begin{array}{c} \lambda_{v,t-1} \\ \lambda_{e,t-1} \end{array} \right]
$$

where the "forcing variables" are given by

$$
\lambda_{v,t} = v_t \left( \mathbf{1} \left\{ Y_t \leq v_t \right\} - \alpha \right)
$$

$$
\lambda_{e,t} = \frac{1}{\alpha} \mathbf{1} \left\{ Y_t \leq v_t \right\} Y_t - e_t
$$

- While there are relatively few dynamic models for ES, there are some. We consider the following models as competition:

**1** **Rolling window**:

$$\widehat{VaR}_t = \widehat{Quantile}\left\{Y_s\right\}_{s=t-m+1}^{t}$$

$$\widehat{ES}_t = \frac{1}{\alpha m}\sum_{s=t-m+1}^{t} Y_s \mathbf{1}\left\{Y_s \leq \widehat{VaR}_s\right\}$$

- $m \in \{125, 250, 500\}$

**2** **ARMA-GARCH models**

$$Y_t = \mu_t + \sigma_t \eta_t$$
$$\mu_t \sim ARMA(p,q) \quad , \quad \sigma_t^2 \sim GARCH(p,q)$$

a. $\quad \eta_t \sim iid\ N(0,1)$

b. $\quad \eta_t \sim iid\ Skew\ t(0,1,\nu,\lambda)$

c. $\quad \eta_t \sim iid\ F(0,1)$ (estimated by the EDF)

- Model 2(c) is also known as "filtered historical simulation," and is probably the best existing model for ES (see survey by Engle and Manganelli (2004, *book*)).

# Pros and cons of directly modeling ES and VaR

- Consider a generic model:

$$
\begin{aligned}
VaR_t &= v\left(\mathbf{Z}_{t-1}; \boldsymbol{\theta}\right) \\
ES_t &= e\left(\mathbf{Z}_{t-1}; \boldsymbol{\theta}\right)
\end{aligned}
$$

# Pros and cons of directly modeling ES and VaR

- Consider a generic model:

$$
\begin{aligned}
VaR_t &= v\left(\mathbf{Z}_{t-1}; \boldsymbol{\theta}\right) \\
ES_t &= e\left(\mathbf{Z}_{t-1}; \boldsymbol{\theta}\right)
\end{aligned}
$$

- Such a model is a *semiparametric* model for returns:

  - We assume **parametric dynamics** for ES and VaR

  - We make **no assumptions about the distribution** of returns (beyond regularity conditions required for estimation and inference)

# Pros and cons of directly modeling ES and VaR

- Consider a generic model:

$$\begin{aligned} VaR_t &= v\left(\mathbf{Z}_{t-1}; \boldsymbol{\theta}\right) \\ ES_t &= e\left(\mathbf{Z}_{t-1}; \boldsymbol{\theta}\right) \end{aligned}$$

- Such a model is a *semiparametric* model for returns:

    - We assume **parametric dynamics** for ES and VaR

    - We make **no assumptions about the distribution** of returns (beyond regularity conditions required for estimation and inference)

- By eliminating the need for assumptions about the distribution of returns, we hopefully obtain a more robust model. But:

    - There may be **efficiency losses**. We will study this carefully in our OOS forecasting analysis.

    - This is not a **complete probability model**: further assumptions are needed to draw simulations, for example.

# A one-factor model

- Consider a case where there is only one latent factor driving VaR and ES:

$$
\begin{aligned}
v_t &= a \exp\{\kappa_t\} \\
e_t &= b \exp\{\kappa_t\}, \text{ where } b < a < 0 \\
\text{where } \kappa_t &= \omega + \beta\kappa_{t-1} + \gamma H_{t-1}^{-1} s_{t-1}
\end{aligned}
$$

- If we derive the GAS dynamics for $\kappa_t$ we find

$$
H_{t-1}^{-1} s_{t-1} = \frac{1}{e_{t-1}} \left( \frac{1}{\alpha} \mathbf{1}\left\{ Y_{t-1} \le v_{t-1} \right\} Y_{t-1} - e_{t-1} \right) \equiv \frac{\lambda_{e,t-1}}{e_{t-1}}
$$

  - The intercept, $\omega$, is not identified here so we fix it at zero.

# GARCH with FZ estimation

- Next consider GARCH dynamics for the latent factor, but estimate using the FZ0 loss function rather than QML:

$$
\begin{aligned}
Y_t &= \kappa_t \eta_t, \quad \eta_t \sim iid\ F_\eta \\
\text{so}\quad v_t &= a \cdot \kappa_t \\
e_t &= b \cdot \kappa_t, \ \text{with}\ b < a < 0 \\
\text{and}\quad \kappa_t^2 &= \omega + \beta \kappa_{t-1}^2 + \gamma Y_{t-1}^2
\end{aligned}
$$

  - As above, the intercept, $\omega$, is not identified here and we fix it at one.

- If the GARCH model is **correct**, this is consistent but almost certainly less efficient than QML

- If the model is **misspecified**, estimating this way yields the parameters that lead to the best possible VaR and ES forecasts**.**

# A hybrid GAS+GARCH model

- Finally, consider a "hybrid" model, where as before we have:

$$
\begin{aligned}
Y_t &= \exp\{\kappa_t\}\eta_t, \quad \eta_t \sim iid\ F_\eta \\
\text{so} \quad v_t &= a\exp\{\kappa_t\} \\
e_t &= b\exp\{\kappa_t\}, \quad \text{with } b < a < 0
\end{aligned}
$$

- We augment the GAS dynamics for $\kappa_t$ with a "GARCH" term:

$$
\kappa_t = \omega + \beta\kappa_{t-1} + \gamma\frac{\lambda_{e,t-1}}{e_{t-1}} + \delta\log|Y_{t-1}|
$$

# Outline

# Statistical inference on models for ES and VaR

- We are in the general framework of M-estimation for time series models:

$$\hat{\boldsymbol{\theta}}_T = \arg\min_{\boldsymbol{\theta}} \ \frac{1}{T} \sum_{t=1}^{T} L_{FZ0}\left(Y_t, v\left(\mathbf{Z}_{t-1}; \boldsymbol{\theta}\right), e\left(\mathbf{Z}_{t-1}; \boldsymbol{\theta}\right); \alpha\right)$$

- Our loss function is **non-differentiable**, but if we assume that $Y_t$ is continuously distributed, this is easily handled.

- Under some regularity conditions, we obtain:

$$\sqrt{T}\left(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}^*\right) \xrightarrow{d} N\left(0, \mathbf{A}^{-1}\mathbf{D}\mathbf{A}^{-1}\right)$$

  - **D** is the usual covariance matrix of the scores (easy to estimate)

  - **A** is the Hessian, which is a bit trickier to obtain (details in the paper)

- The proof builds on Huber (1967), Weiss (1991), Engle-Manganelli (2004).

- **Assumption 1**: See paper for details. Key parts of this assumption:

  - Need finite first moments (unlike VaR estimation)

  - Need unique $\alpha$-quantiles (see Zwingmann and Holzmann (2016) for results when this condition is violated).

- **Theorem 1:** Under Assumption 1, $\hat{\boldsymbol{\theta}}_T \xrightarrow{p} \boldsymbol{\theta}^0$ as $T \to \infty$.

- Proof is straightforward given Theorem 2.1 of Newey and McFadden (1994) and Corollary 5.5 of Fissler and Ziegel (2016).

# Asymptotic normality

- **Assumption 2:** See paper for details. Key parts of this assumption:

    - Need $2 + \delta$ moments of returns

- **Theorem 2:** Under Assumptions 1 and 2, we have

$$\sqrt{T}\mathbf{A}^{-1/2}\mathbf{D}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}^0) \xrightarrow{d} N(0, I) \text{ as } T \to \infty$$

where

$$
\begin{aligned}
\mathbf{A} &= \mathbb{E}\left[g_t(\boldsymbol{\theta}^0)g_t(\boldsymbol{\theta}^0)'\right], \quad g_t(\boldsymbol{\theta}^0) = \frac{\partial L\left(y_t, v_t\left(\boldsymbol{\theta}^0\right), e_t\left(\boldsymbol{\theta}^0\right); \alpha\right)}{\partial \boldsymbol{\theta}} \\
\mathbf{D} &= \mathbb{E}\left[\frac{f_t\left(v_t(\boldsymbol{\theta}^0)\right)}{-e_t(\boldsymbol{\theta}^0)\alpha}\nabla' v_t(\boldsymbol{\theta}^0)\nabla v_t(\boldsymbol{\theta}^0) + \frac{\nabla' e_t(\boldsymbol{\theta}^0)\nabla e_t(\boldsymbol{\theta}^0)}{e_t(\boldsymbol{\theta}^0)^2}\right]
\end{aligned}
$$

- The proof builds on Huber (1967), Weiss (1991), Engle-Manganelli (2004).

# Estimation of the asymptotic covariance matrix

- **Assumption 3:** See paper for details. Key parts of this assumption:

  - Bandwidth $(c_T)$ satisfies $c_T \to 0$ and $c_T \sqrt{T} \to \infty$.

- **Theorem 3:** Under Ass'ns 1–3, $\hat{\mathbf{A}}_T - \mathbf{A} \xrightarrow{p} \mathbf{0}$ and $\hat{\mathbf{D}}_T - \mathbf{D} \xrightarrow{p} \mathbf{0}$, where

$$\hat{\mathbf{A}}_T = T^{-1} \sum_{t=1}^{T} g_t(\hat{\boldsymbol{\theta}}_T) g_t(\hat{\boldsymbol{\theta}}_T)'$$

$$\hat{\mathbf{D}}_T = T^{-1} \sum_{t=1}^{T} \left\{ \frac{1}{2\hat{c}_T} \mathbf{1}\left\{ \left| y_t - v_t\left(\hat{\boldsymbol{\theta}}_T\right)\right| < \hat{c}_T \right\} \frac{\nabla' v_t\left(\hat{\boldsymbol{\theta}}_T\right) \nabla v_t\left(\hat{\boldsymbol{\theta}}_T\right)}{-e_t\left(\hat{\boldsymbol{\theta}}_T\right)\alpha} \right.$$

$$\left. + \frac{\nabla' e_t\left(\hat{\boldsymbol{\theta}}_T\right) \nabla e_t\left(\hat{\boldsymbol{\theta}}_T\right)}{e_t\left(\hat{\boldsymbol{\theta}}_T\right)^2} \right\}$$

- This extends Engle and Manganelli (2004) from dynamic VaR models to dynamic joint models for VaR and ES.

# Simulation study

- For comparability with the existing literature, we simulate a GARCH process:

$$
\begin{aligned}
Y_t &= \sigma_t \eta_t \\
\eta_t &\sim iid\ F_\eta\left(0,1\right) \\
\sigma_t^2 &= \omega + \beta \sigma_{t-1}^2 + \gamma Y_{t-1}^2 \\
[v_t, e_t] &= [a, b]\, \sigma_t
\end{aligned}
$$

- $[\omega, \beta, \gamma] = [0.05, 0.9, 0.05]$.

- $F_\eta \in \{\ N\left(0,1\right)\ ,\ Skewt\left(5, -0.5\right)\ \}$.

- $\alpha \in \{\ 0.01\ ,\ 0.025\ ,\ 0.05\ ,\ 0.1\ , 0.2\}$.

- For std errors, we use $c_T = T^{-1/3}$.

- $T \in \{\ 2500\ ,\ 5000\ \}$, and $reps = 1000$.

- $\theta = [\beta, \gamma, b, c]$ where $c \equiv a/b$.

# Finite-sample properties: varying T

Estimator is approximately unbiased, and 95% confidence intervals have reasonable coverage

**Normal innovations, $\alpha = 0.05$**

|  | $T = 2500$ | | | | $T = 5000$ | | | |
|---|---|---|---|---|---|---|---|---|
|  | $\beta$ | $\gamma$ | $b_\alpha$ | $c_\alpha$ | $\beta$ | $\gamma$ | $b_\alpha$ | $c_\alpha$ |
| True | 0.900 | 0.050 | -2.063 | 0.797 | 0.900 | 0.050 | -2.063 | 0.797 |
| Median | 0.901 | 0.048 | -2.051 | 0.800 | 0.899 | 0.049 | -2.094 | 0.799 |
| Bias | -0.013 | 0.005 | -0.097 | 0.002 | -0.008 | 0.002 | -0.081 | 0.001 |
| St dev | 0.062 | 0.046 | 0.707 | 0.015 | 0.041 | 0.021 | 0.511 | 0.010 |
| Cov'age | 0.913 | 0.874 | 0.916 | 0.947 | 0.923 | 0.907 | 0.927 | 0.948 |

# Finite-sample properties: varying distribution

Similar results for Skew t innovations as for Normal case

**T=2500, $\alpha = 0.05$**

|  | Normal | | | | Skew t | | | |
|---|---|---|---|---|---|---|---|---|
|  | $\beta$ | $\gamma$ | $b_\alpha$ | $c_\alpha$ | $\beta$ | $\gamma$ | $b_\alpha$ | $c_\alpha$ |
| True | 0.900 | 0.050 | -2.063 | 0.797 | 0.900 | 0.050 | -2.767 | 0.651 |
| Median | 0.901 | 0.048 | -2.051 | 0.800 | 0.898 | 0.048 | -2.795 | 0.654 |
| Bias | -0.013 | 0.005 | -0.097 | 0.002 | -0.011 | 0.003 | -0.114 | 0.003 |
| St dev | 0.062 | 0.046 | 0.707 | 0.015 | 0.053 | 0.025 | 0.782 | 0.017 |
| Cov'age | 0.913 | 0.874 | 0.916 | 0.947 | 0.916 | 0.904 | 0.922 | 0.951 |

# Finite-sample properties: varying alpha

Std dev higher for smaller alpha, and coverage worse for smaller alpha

**T=2500, Normal**

|  | $\alpha = 0.01$ | | | | $\alpha = 0.10$ | | | |
|---|---|---|---|---|---|---|---|---|
|  | $\beta$ | $\gamma$ | $b_\alpha$ | $c_\alpha$ | $\beta$ | $\gamma$ | $b_\alpha$ | $c_\alpha$ |
| True | 0.900 | 0.050 | -2.665 | 0.873 | 0.900 | 0.050 | -1.755 | 0.730 |
| Median | 0.901 | 0.049 | -2.615 | 0.882 | 0.900 | 0.048 | -1.769 | 0.730 |
| Bias | -0.017 | 0.015 | -0.108 | 0.008 | -0.015 | 0.006 | -0.103 | 0.000 |
| St dev | 0.077 | 0.076 | 1.095 | 0.022 | 0.065 | 0.052 | 0.623 | 0.013 |
| Cov'age | 0.868 | 0.827 | 0.875 | 0.919 | 0.917 | 0.883 | 0.925 | 0.954 |

# Estimation of VaR and ES: Normal-GARCH DGP

FZ estimation beats CAViaR, but MLE (naturally) performs best here

**Normal innovations, T $= 2500$**

| | **VaR** | | | | **ES** | | |
|---|---|---|---|---|---|---|---|
| | *MAE* | *MAE ratio* | | | *MAE* | *MAE ratio* | |
| $\alpha$ | ML | CAViaR | FZ | | ML | CAViaR | FZ |
| 0.01 | 0.069 | 1.368 | 1.369 | | 0.084 | 1.487 | 1.345 |
| 0.025 | 0.055 | 1.305 | 1.288 | | 0.064 | 1.341 | 1.290 |
| 0.05 | 0.043 | 1.302 | 1.271 | | 0.051 | 1.332 | 1.289 |
| 0.10 | 0.034 | 1.322 | 1.253 | | 0.042 | 1.394 | 1.302 |

# Estimation of VaR and ES: Skew t-GARCH DGP

FZ estimation beats CAViaR, QMLE performs best here

**Skew t innovations, T = 2500**

| | VaR | | | ES | | |
|---|---|---|---|---|---|---|
| | *MAE* | *MAE ratio* | | *MAE* | *MAE ratio* | |
| $\alpha$ | QML | CAViaR | FZ | QML | CAViaR | FZ |
| 0.01 | 0.196 | 1.327 | 1.381 | 0.342 | 1.249 | 1.252 |
| 0.025 | 0.120 | 1.228 | 1.244 | 0.205 | 1.166 | 1.166 |
| 0.05 | 0.084 | 1.193 | 1.166 | 0.141 | 1.154 | 1.129 |
| 0.10 | 0.056 | 1.168 | 1.089 | 0.098 | 1.160 | 1.083 |

# Estimation of VaR and ES: Mixed Normal DGP

FZ estimation is approx as good as QMLE for smaller quantiles

**Mixed normal innovations, T = 2500**

| | **VaR** | | | | **ES** | | |
|---|---|---|---|---|---|---|---|
| | *MAE* | *MAE ratio* | | | *MAE* | *MAE ratio* | |
| $\alpha$ | QML | CAViaR | FZ | | QML | CAViaR | FZ |
| 0.01 | 0.124 | 1.015 | 1.032 | | 0.120 | 1.682 | 1.040 |
| 0.025 | 0.089 | 0.969 | 0.959 | | 0.093 | 1.494 | 0.957 |
| 0.05 | 0.092 | 1.127 | 1.064 | | 0.085 | 1.640 | 1.083 |
| 0.10 | 0.174 | 1.744 | 1.549 | | 0.083 | 2.748 | 2.097 |

# Estimation of VaR and ES: "Offsetting dynamics" DGP

FZ estimation strongly dominates when VaR/ES dynamics very different from volatility

**Offsetting dynamics, T = 2500**

| | **VaR** | | | **ES** | | |
|---|---|---|---|---|---|---|
| | *MAE* | *MAE ratio* | | *MAE* | *MAE ratio* | |
| $\alpha$ | QML | CAViaR | FZ | QML | CAViaR | FZ |
| 0.01 | 0.111 | 0.395 | 0.406 | 0.264 | 1.048 | 0.420 |
| 0.025 | 0.074 | 0.339 | 0.343 | 0.222 | 0.679 | 0.331 |
| 0.05 | 0.058 | 0.369 | 0.348 | 0.196 | 0.522 | 0.298 |
| 0.10 | 0.058 | 0.476 | 0.465 | 0.166 | 0.436 | 0.348 |

# Outline

# Data

- We study daily returns on four equity indices

    - S&P 500

    - Dow Jones Industrial Average

    - NIKKEI 225

    - FTSE 100.

- Sample period is January 1990 to December 2016

    - Number of observations ($T$) is 6630 to 6805.

    - We use the first 10 years ($R \approx 2500$) for estimation, and the last 17 years ($P \approx 4250$) for out-of-sample forecast comparison.

# Daily returns on the S&P 500 index
## Rolling window estimates of the 5% VaR and ES



**5% VaR forecasts for S&P 500 daily returns**

Legend:
- Return
- 5% VaR
- 5% ES
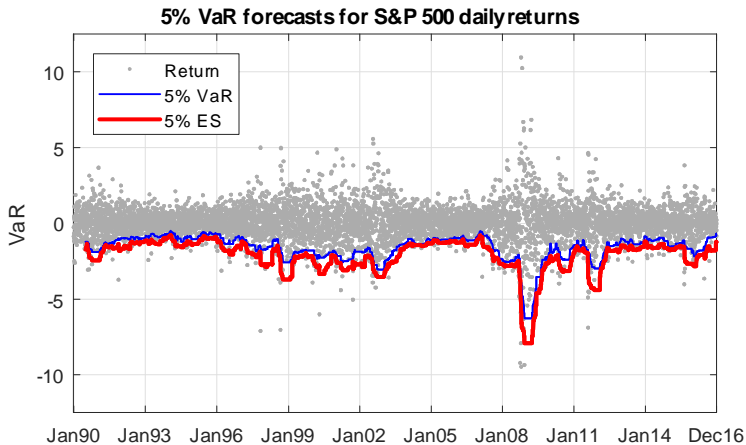
# Summary statistics

|  | S&P 500 | DJIA | NIKKEI | FTSE |
|---|---|---|---|---|
| Mean (Annualized) | 6.776 | 7.238 | -2.682 | 3.987 |
| Std dev (Annualized) | 17.879 | 17.042 | 24.667 | 17.730 |
| Skewness | -0.244 | -0.163 | -0.114 | -0.126 |
| Kurtosis | 11.673 | 11.116 | 8.580 | 8.912 |
| VaR-0.01 | -3.128 | -3.034 | -4.110 | -3.098 |
| VaR-0.025 | -2.324 | -2.188 | -3.151 | -2.346 |
| VaR-0.05 | -1.731 | -1.640 | -2.451 | -1.709 |
| ES-0.01 | -4.528 | -4.280 | -5.783 | -4.230 |
| ES-0.025 | -3.405 | -3.215 | -4.449 | -3.295 |
| ES-0.05 | -2.697 | -2.553 | -3.603 | -2.643 |

# ARMA-GARCH-Skew t models for these returns

|            | S&P 500   | DJIA   | NIKKEI | FTSE   |
|------------|-----------|--------|--------|--------|
| Mean       | ARMA(1,1) | AR(2)  | AR(0)  | AR(4)  |
| $R^2$      | 0.006     | 0.004  | 0.000  | 0.009  |
|            |           |        |        |        |
| $\omega$   | 0.014     | 0.017  | 0.066  | 0.016  |
| $\beta$    | 0.905     | 0.897  | 0.863  | 0.893  |
| $\alpha$   | 0.082     | 0.088  | 0.113  | 0.094  |
|            |           |        |        |        |
| $\nu$      | 6.934     | 7.062  | 7.806  | 11.800 |
| $\lambda$  | -0.115    | -0.100 | -0.066 | -0.102 |

# One-factor models for ES and VaR

SP500, alpha=0.05. Preferred model is the "hybrid" model

|   | GAS-1F | GARCH-FZ | Hybrid |
|---|---|---|---|
| $\beta$ | 0.990 (0.004) | 0.908 (0.072) | 0.968 (0.015) |
| $\gamma$ | 0.010 (0.002) | 0.030 (0.010) | 0.011 (0.002) |
| $\delta$ | – | – | 0.018 (0.009) |
| $a$ | −1.490 (0.346) | −2.659 (0.492) | −2.443 (0.473) |
| $b$ | −2.089 (0.487) | −3.761 (0.747) | −3.389 (0.664) |
| Avg Loss | 0.750 | 0.762 | 0.745 |

# Dynamic Expected Shortfall: 1990-2016

ES ranges from around -1.5% in mid 90s, to -10% in financial crisis
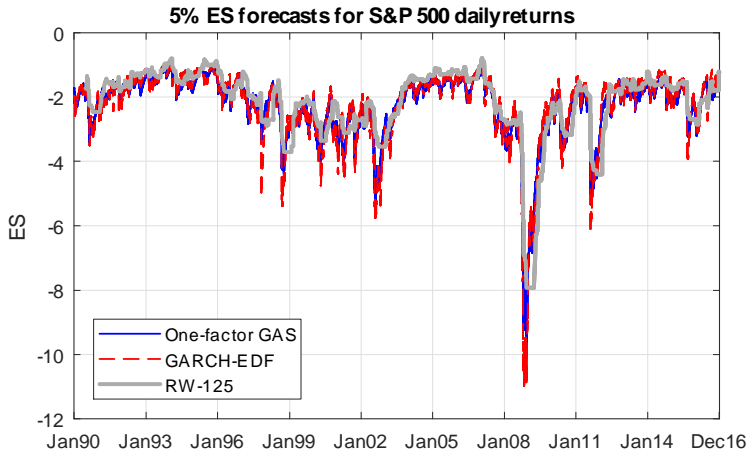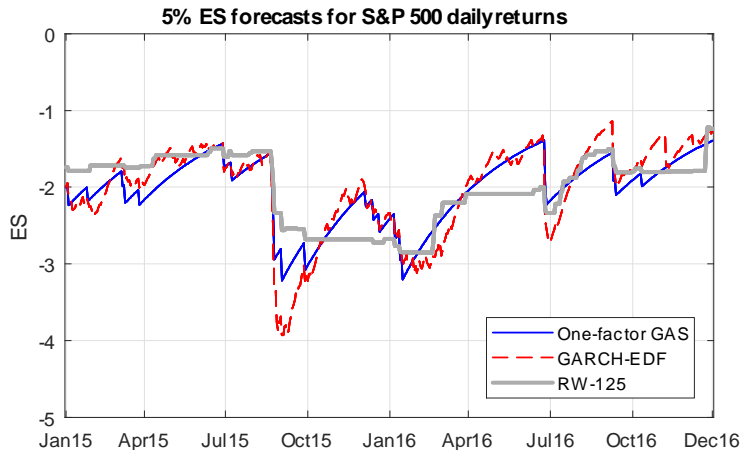
# Dynamic Expected Shortfall: 2015-2016

The difference between the GAS and GARCH forcing variables is apparent here



5% ES forecasts for S&P 500 daily returns

# The models used for in OOS forecast comparison

- ⊙ Rolling Window, with $m \in \{125, 250, 500\}$

- ⊙ GARCH(1,1) with Normal, Skew t, or EDF for the residuals

- ★ GAS(1,1) dynamics, 2 factors

- ★ GAS(1,1) dynamics, 1 factor

- ★ GARCH-FZ: estimating the GARCH model using the FZ loss function

- ★ Hybrid model: one-factor GAS model, with GARCH forcing variable included

# Evaluating and comparing out-of-sample forecasts

- We estimate the models using data only up until Dec 1999

    - $R \approx 2500$, $P \approx 4250$

- Forecasts of VaR and ES are then produced for each day in the OOS period

    - No look-ahead bias in the forecasts

- We compare the forecasts using the FZ loss function:

    1. Rankings by average loss in the OOS period(s)

    2. Diebold-Mariano tests on average losses from these forecasts

    3. Goodness-of-fit tests

# OOS forecast comparison results: Average loss

SP500, alpha=0.05. 1-factor GAS model, w/wo "hybrid" forcing variable, is best.

|  | **SP500** | **DJIA** | **NIKKEI** | **FTSE** |
|---|---|---|---|---|
| RW-125 | 0.914 | 0.864 | 1.290 | 0.959 |
| RW-250 | 0.959 | 0.909 | 1.294 | 1.002 |
| RW-500 | 1.023 | 0.976 | 1.318 | 1.056 |
| GARCH-N | 0.876 | 0.808 | 1.170 | 0.871 |
| GARCH-Skt | 0.866 | 0.796 | 1.168 | *0.863* |
| GARCH-EDF | 0.862 | *0.796* | *1.166* | 0.867 |
| FZ-2F | *0.856* | 0.798 | 1.206 | 1.098 |
| FZ-1F | **0.853** | **0.784** | 1.191 | 0.867 |
| GARCH-FZ | 0.862 | 0.797 | 1.167 | 0.866 |
| Hybrid | 0.869 | 0.797 | **1.165** | **0.862** |

# OOS forecast comparison results : Diebold-Mariano t-stats

SP500, alpha=0.05. FZ-1F beats all. Not signif better than GARCH-EDF/Skew t

■ A positive entry indicates the Column model is better than the Row model

|  | RW125 | G-EDF | FZ-2F | FZ-1F | G-FZ | Hybrid |
|---|---|---|---|---|---|---|
| RW125 | – | 2.900 | 2.978 | 3.978 | 3.020 | 2.967 |
| RW250 | 2.580 | 3.730 | 3.799 | 4.701 | 3.921 | 4.110 |
| RW500 | 4.260 | 4.937 | 5.168 | 5.893 | 5.125 | 5.450 |
| G-N | -2.109 | 3.068 | 1.553 | 2.248 | 2.818 | 0.685 |
| G-Skt | -2.693 | 2.103 | 0.889 | 1.475 | 1.232 | -0.403 |
| G-EDF | -2.900 | – | 0.599 | 1.157 | 0.024 | -0.769 |
| FZ-2F | -2.978 | -0.599 | – | 0.582 | -0.555 | -0.580 |
| FZ-1F | -3.912 | -1.198 | -0.582 | – | -1.266 | -1.978 |
| G-FZ | -3.020 | -0.024 | 0.555 | 1.266 | – | -0.914 |
| Hybrid | -3.276 | 0.045 | 0.580 | 1.978 | 0.914 | – |

■ Ranking models by OOS average loss, for different tail probabilities

|        | 0.01 | 0.025 | 0.05 | 0.10 |
|--------|------|-------|------|------|
| RW-125 | 8    | 7.75  | 7.75 | 8    |
| RW-250 | 8.25 | 8.25  | 8.75 | 9    |
| RW-500 | 9.5  | 9.5   | 9.75 | 10   |
| G-N    | 5.25 | 5     | 6.25 | 3.75 |
| G-Skt  | 3    | 2.5   | 3.5  | 4.75 |
| G-EDF  | 2.5  | 2.25  | 3.25 | 3.25 |
| FZ-2F  | 5.5  | 7.25  | 6.25 | 5.5  |
| FZ-1F  | 7    | 4.25  | 3    | 3    |
| G-FZ   | 2    | 2.25  | 3.5  | 5.75 |
| Hybrid | 4    | 6     | 3    | 2    |

# Goodness-of-fit tests for VaR and ES

- Recall the forcing variables in our GAS model:

$$\lambda_{v,t} = v_t \left( \mathbf{1}\left\{ Y_t \leq v_t \right\} - \alpha \right)$$

$$\lambda_{e,t} = \frac{1}{\alpha} \mathbf{1}\left\{ Y_t \leq v_t \right\} Y_t - e_t$$

- These variables can be considered as "**generalized forecast errors**" as under correct specification of the models for VaR and ES, we have

$$\mathbb{E}_{t-1} \left[ \begin{array}{cc} \lambda_{v,t} & \lambda_{e,t} \end{array} \right] = 0$$

- We adopt the "dynamic quantile" **regression-based test** of Engle and Manganelli (2004) for VaR, and propose its natural analog for ES:

$$\lambda_{v,t} = a_0 + a_1 \lambda_{v,t-1} + a_2 v_t + \varepsilon_{v,t}$$

$$\lambda_{e,t} = b_0 + b_1 \lambda_{e,t-1} + b_2 e_t + \varepsilon_{e,t}$$

# OOS goodness-of-fit tests: VaR and ES

alpha=0.05. Red=rejected. FZ-1F performs best. All models fail for FTSE

| | GoF p-values: VaR | | | | GoF p-values: ES | | | |
|---|---|---|---|---|---|---|---|---|
| | S&P | DJIA | NIK | FTSE | S&P | DJIA | NIK | FTSE |
| RW-125 | 0.021 | 0.013 | 0.000 | 0.000 | 0.029 | 0.018 | 0.006 | 0.000 |
| RW-250 | 0.001 | 0.001 | 0.007 | 0.000 | 0.043 | 0.014 | 0.018 | 0.002 |
| RW-500 | 0.001 | 0.001 | 0.000 | 0.000 | 0.012 | 0.011 | 0.001 | 0.000 |
| GCH-N | 0.031 | 0.139 | 0.532 | 0.000 | 0.001 | 0.006 | 0.187 | 0.000 |
| GCH-Skt | 0.003 | 0.085 | 0.114 | 0.000 | 0.003 | 0.085 | 0.282 | 0.000 |
| GCH-EDF | 0.003 | 0.029 | 0.583 | 0.000 | 0.014 | 0.098 | 0.527 | 0.000 |
| FZ-2F | 0.000 | 0.000 | 0.258 | 0.000 | 0.061 | 0.195 | 0.247 | 0.000 |
| FZ-1F | 0.242 | 0.248 | 0.317 | 0.019 | 0.313 | 0.130 | 0.612 | 0.003 |
| GCH-FZ | 0.005 | 0.001 | 0.331 | 0.000 | 0.018 | 0.011 | 0.389 | 0.000 |
| Hybrid | 0.001 | 0.069 | 0.326 | 0.000 | 0.010 | 0.159 | 0.518 | 0.000 |

# Summary and conclusions

- The new Basel Accord will generate demand for models for Expected Shortfall

  - Existing models for volatility and VaR do not seem to do well for ES

- We exploit a recent result from decision theory that shows that ES is jointly elicitable with VaR

  - The "Fissler-Ziegel" loss function

- We propose new models and adaptations of old models, for forecasting ES

  - For $\alpha = 0.01$ and 0.025, the best models are GARCH estimated via FZ loss minimization and GARCH with nonparametric residuals.

  - For $\alpha = 0.05$ and 0.10, the best models are the one-factor GAS model, and the hybrid one-factor GAS/GARCH model.

# Appendix

- "The financial crisis exposed material weaknesses in the overall design of the framework for capitalising trading activities."

- "A number of weaknesses have been identified with using Value-at-Risk for determining regulatory capital requirements, including its inability to capture 'tail risk.' For this reason, the Committee proposed in May 2012 to replace Value-at-Risk with Expected Shortfall."

- "Risk reporting: the desk must produce, at least once a week... risk measure reports, including desk VaR/ES, desk VaR/ES sensitivities to risk factors, backtesting and p-value."

$\Rightarrow$ Expected shortfall is going to become an important part of risk management, complementing past emphasis on VaR.

- For intuition, assume that returns follow a conditional location-scale model (eg, ARMA-GARCH)

$$Y_t = \mu_t + \sigma_t \eta_t, \quad \eta_t \sim iid \; F_\eta (0, 1)$$

- For intuition, assume that returns follow a conditional location-scale model (eg, ARMA-GARCH)

$$Y_t = \mu_t + \sigma_t \eta_t, \quad \eta_t \sim iid\ F_\eta(0,1)$$

- In this case, we have

$$
\begin{aligned}
VaR_t &= \mu_t + a\sigma_t, \quad \text{where } a = F_\eta^{-1}(\alpha) \\
ES_t &= \mu_t + b\sigma_t, \quad \text{where } b = \mathbb{E}[\eta_t | \eta_t \leq a]
\end{aligned}
$$

and we we can recover $(\mu_t, \sigma_t)$ from $(VaR_t, ES_t)$.

# Expected Shortfall and VaR in location-scale models

- For intuition, assume that returns follow a conditional location-scale model (eg, ARMA-GARCH)

$$Y_t = \mu_t + \sigma_t \eta_t, \quad \eta_t \sim iid \ F_\eta(0,1)$$

- In this case, we have

$$
\begin{aligned}
VaR_t &= \mu_t + a\sigma_t, \quad \text{where} \quad a = F_\eta^{-1}(\alpha) \\
ES_t &= \mu_t + b\sigma_t, \quad \text{where} \quad b = \mathbb{E}[\eta_t | \eta_t \leq a]
\end{aligned}
$$

and we we can recover $(\mu_t, \sigma_t)$ from $(VaR_t, ES_t)$.

- If $\sigma_t = \bar{\sigma} \ \forall \ t$, then $ES_t = c + VaR_t$, where $c = (b-a)\bar{\sigma}$

# Expected Shortfall and VaR in location-scale models

- For intuition, assume that returns follow a conditional location-scale model (eg, ARMA-GARCH)

$$Y_t = \mu_t + \sigma_t \eta_t, \quad \eta_t \sim iid \ F_\eta(0, 1)$$

- In this case, we have

$$
\begin{aligned}
VaR_t &= \mu_t + a\sigma_t, \quad \text{where} \quad a = F_\eta^{-1}(\alpha) \\
ES_t &= \mu_t + b\sigma_t, \quad \text{where} \quad b = \mathbb{E}[\eta_t | \eta_t \leq a]
\end{aligned}
$$

and we we can recover $(\mu_t, \sigma_t)$ from $(VaR_t, ES_t)$.

- If $\sigma_t = \bar{\sigma} \ \forall \ t$, then $ES_t = c + VaR_t$, where $c = (b - a)\bar{\sigma}$

- If $\mu_t = 0 \ \forall \ t$, then $ES_t = d \times VaR_t$, where $d = b/a$

- Baseline specification:

$$
\left[ \begin{array}{c} v_{t+1} \\ e_{t+1} \end{array} \right] = \mathbf{w} + \mathbf{B} \left[ \begin{array}{c} v_t \\ e_t \end{array} \right] + \mathbf{A} \left[ \begin{array}{c} \lambda_{v,t} \\ \lambda_{e,t} \end{array} \right]
$$

# Location-scale restrictions on the GAS model

- Baseline specification:

$$\left[ \begin{array}{c} v_{t+1} \\ e_{t+1} \end{array} \right] = \mathbf{w} + \mathbf{B} \left[ \begin{array}{c} v_t \\ e_t \end{array} \right] + \mathbf{A} \left[ \begin{array}{c} \lambda_{v,t} \\ \lambda_{e,t} \end{array} \right]$$

- Motivated by the familiarity of **location-scale models**, where $Y_t = \mu_t + \sigma_t \eta_t$, we consider the following versions of this model

# Location-scale restrictions on the GAS model

- Baseline specification:

$$\left[ \begin{array}{c} v_{t+1} \\ e_{t+1} \end{array} \right] = \mathbf{w} + \mathbf{B} \left[ \begin{array}{c} v_t \\ e_t \end{array} \right] + \mathbf{A} \left[ \begin{array}{c} \lambda_{v,t} \\ \lambda_{e,t} \end{array} \right]$$

- Motivated by the familiarity of **location-scale models**, where
$Y_t = \mu_t + \sigma_t \eta_t$, we consider the following versions of this model

**1** $\mu_t = 0 \ \forall \ t$. This implies:

$$H_0 : \frac{w_e}{w_v} = \frac{a_{ev}}{a_{vv}} = \frac{a_{ee}}{a_{ve}} \cap b_e = b_v$$

# Location-scale restrictions on the GAS model

- Baseline specification:

$$\left[ \begin{array}{c} v_{t+1} \\ e_{t+1} \end{array} \right] = \mathbf{w} + \mathbf{B} \left[ \begin{array}{c} v_t \\ e_t \end{array} \right] + \mathbf{A} \left[ \begin{array}{c} \lambda_{v,t} \\ \lambda_{e,t} \end{array} \right]$$

- Motivated by the familiarity of **location-scale models**, where $Y_t = \mu_t + \sigma_t \eta_t$, we consider the following versions of this model

1. $\mu_t = 0 \ \forall \ t$. This implies:

$$H_0 : \frac{w_e}{w_v} = \frac{a_{ev}}{a_{vv}} = \frac{a_{ee}}{a_{ve}} \cap b_e = b_v$$

2. $\mu_t = \bar{\mu} \ \forall \ t$. This implies:

$$H_0 : \frac{a_{ev}}{a_{vv}} = \frac{a_{ee}}{a_{ve}} \cap b_e = b_v$$

# Location-scale restrictions on the GAS model

- Baseline specification:

$$\left[ \begin{array}{c} v_{t+1} \\ e_{t+1} \end{array} \right] = \mathbf{w} + \mathbf{B} \left[ \begin{array}{c} v_t \\ e_t \end{array} \right] + \mathbf{A} \left[ \begin{array}{c} \lambda_{v,t} \\ \lambda_{e,t} \end{array} \right]$$

- Motivated by the familiarity of **location-scale models**, where $Y_t = \mu_t + \sigma_t \eta_t$, we consider the following versions of this model

1. $\mu_t = 0 \ \forall \ t$. This implies:

$$H_0 : \frac{w_e}{w_v} = \frac{a_{ev}}{a_{vv}} = \frac{a_{ee}}{a_{ve}} \cap b_e = b_v$$

2. $\mu_t = \bar{\mu} \ \forall \ t$. This implies:

$$H_0 : \frac{a_{ev}}{a_{vv}} = \frac{a_{ee}}{a_{ve}} \cap b_e = b_v$$

3. $\sigma_t = \bar{\sigma} \ \forall \ t$. This implies:

$$H_0 : a_{ev} = a_{vv} \cap a_{ee} = a_{ve} \cap b_e = b_v$$

■ A positive entry indicates the Column model is better than the Row model

| | RW125 | G-EDF | FZ-2F | FZ-1F | G-FZ | Hybrid |
|---|---|---|---|---|---|---|
| RW125 | – | 3.125 | 1.972 | 3.599 | 3.212 | 2.642 |
| RW250 | 2.035 | 3.472 | 2.637 | 4.240 | 3.613 | 3.447 |
| RW500 | 3.587 | 4.731 | 3.966 | 5.605 | 4.879 | 4.968 |
| G-N | -1.100 | 3.522 | 1.645 | 2.346 | 3.835 | 1.963 |
| G-Skt | -2.728 | 2.393 | 0.093 | 0.738 | 2.850 | -0.447 |
| G-EDF | -3.125 | – | -0.595 | -0.198 | 1.482 | -1.500 |
| FZ-2F | -1.972 | 0.595 | – | 0.348 | 1.111 | 0.368 |
| FZ-1F | -3.599 | 0.198 | -0.348 | – | 0.739 | -1.406 |
| G-FZ | -3.212 | -1.482 | -1.111 | -0.739 | – | -2.300 |
| Hybrid | -2.642 | 1.500 | -0.368 | 1.406 | 2.300 | – |

5% VaR forecasts for S&P 500 daily returns

5% VaR forecasts for S&P 500 daily returns
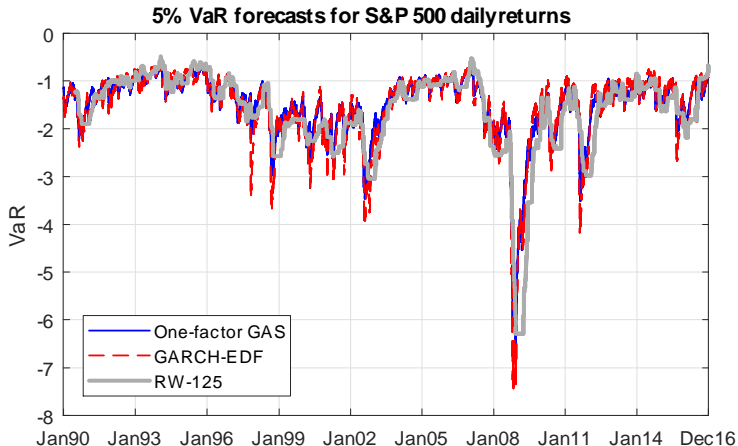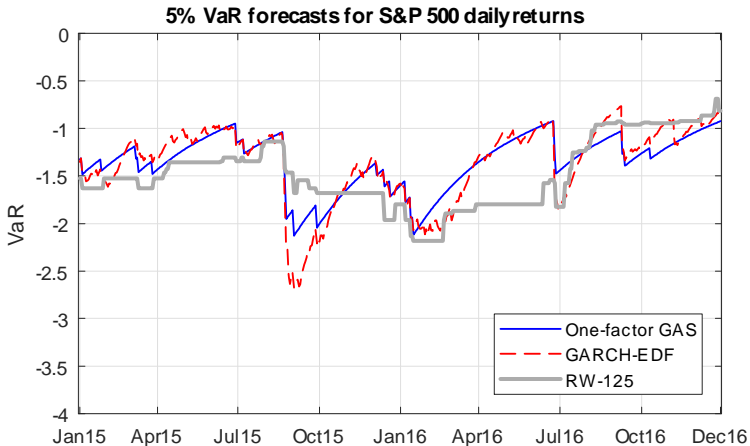
# OOS forecast rankings across various alphas: a=0.01
GARCH estimated by FZ loss is best on average

- Ranking models by OOS average loss, for different tail probabilities

|        | S&P | DJIA | NIK | FTSE | Avg |
|--------|-----|------|-----|------|-----|
| RW-125 | 7   | 8    | 10  | 7    | 8   |
| RW-250 | 8   | 9    | 8   | 8    | 8.25 |
| RW-500 | 10  | 10   | 9   | 9    | 9.5 |
| G-N    | 6   | 6    | 5   | 4    | 5.25 |
| G-Skt  | 5   | 3    | 2   | 2    | 3   |
| G-EDF  | 4   | 2    | 3   | 1    | 2.5 |
| FZ-2F  | 1   | 4    | 7   | 10   | 5.5 |
| FZ-1F  | 9   | 7    | 6   | 6    | 7   |
| G-FZ   | 3   | 1    | 1   | 3    | 2   |
| Hybrid | 2   | 5    | 4   | 5    | 4   |

# OOS forecast rankings across various alphas: a=0.025
## GARCH-EDF and GARCH-FZ are best on average

- Ranking models by OOS average loss, for different tail probabilities

| | S&P | DJIA | NIK | FTSE | Avg |
|---|---|---|---|---|---|
| RW-125 | 8 | 8 | 8 | 7 | 7.75 |
| RW-250 | 9 | 9 | 7 | 8 | 8.25 |
| RW-500 | 10 | 10 | 9 | 9 | 9.5 |
| G-N | 7 | 6 | 4 | 3 | 5 |
| G-Skt | 5 | 3 | **1** | **1** | *2.5* |
| G-EDF | *2* | *2* | 3 | *2* | **2.25** |
| FZ-2F | 4 | 5 | 10 | 10 | 7.25 |
| FZ-1F | 3 | 4 | 6 | 4 | 4.25 |
| G-FZ | **1** | **1** | *2* | 5 | **2.25** |
| Hybrid | 6 | 7 | 5 | 6 | 6 |

# OOS forecast rankings across various alphas: a=0.05
FZ-1F, with and without "hybrid" term, is best

- Ranking models by OOS average loss, for different tail probabilities

|        | S&P | DJIA | NIK | FTSE | Avg  |
|--------|-----|------|-----|------|------|
| RW-125 | 8   | 8    | 8   | 7    | 7.75 |
| RW-250 | 9   | 9    | 9   | 8    | 8.75 |
| RW-500 | 10  | 10   | 10  | 9    | 9.75 |
| G-N    | 7   | 7    | 5   | 6    | 6.25 |
| G-Skt  | 5   | 3    | 4   | *2*  | 3.5  |
| G-EDF  | 4   | *2*  | *2* | 5    | *3.25* |
| FZ-2F  | *2* | 6    | 7   | 10   | 6.25 |
| FZ-1F  | **1** | **1** | 6 | 4  | **3** |
| G-FZ   | 3   | 5    | 3   | 3    | 3.5  |
| Hybrid | 6   | 4    | **1** | **1** | **3** |

# OOS forecast rankings across various alphas: a=0.10
FZ-1F with "hybrid" term is best

- Ranking models by OOS average loss, for different tail probabilities

|         | S&P | DJIA | NIK | FTSE | Avg  |
|---------|-----|------|-----|------|------|
| RW-125  | 8   | 8    | 8   | 8    | 8    |
| RW-250  | 9   | 9    | 9   | 9    | 9    |
| RW-500  | 10  | 10   | 10  | 10   | 10   |
| G-N     | 3   | *2*  | 5   | 5    | 3.75 |
| G-Skt   | 7   | 4    | 4   | 4    | 4.75 |
| G-EDF   | 4   | 3    | 3   | 3    | 3.25 |
| FZ-2F   | *2* | 6    | 7   | 7    | 5.5  |
| FZ-1F   | **1** | 7  | *2* | *2*  | *3*  |
| G-FZ    | 6   | 5    | 6   | 6    | 5.75 |
| Hybrid  | 5   | **1** | **1** | **1** | **2** |

# OOS goodness-of-fit tests: VaR and ES

alpha=0.025. Red=rejected. GARCH-EDF and GARCH-FZ perform best. All models fail for FTSE.

| | GoF p-values: VaR | | | | GoF p-values: ES | | | |
|---|---|---|---|---|---|---|---|---|
| | S&P | DJIA | NIK | FTSE | S&P | DJIA | NIK | FTSE |
| RW-125 | 0.022 | 0.003 | 0.000 | 0.000 | 0.009 | 0.004 | 0.001 | 0.001 |
| RW-250 | 0.005 | 0.007 | 0.002 | 0.000 | 0.023 | 0.039 | 0.010 | 0.005 |
| RW-500 | 0.001 | 0.000 | 0.004 | 0.000 | 0.019 | 0.011 | 0.007 | 0.000 |
| GCH-N | 0.000 | 0.002 | 0.172 | 0.000 | 0.000 | 0.000 | 0.048 | 0.000 |
| GCH-Skt | 0.005 | 0.057 | 0.789 | 0.000 | 0.010 | 0.076 | 0.736 | 0.001 |
| GCH-EDF | 0.164 | 0.149 | 0.789 | 0.000 | 0.237 | 0.379 | 0.588 | 0.000 |
| FZ-2F | 0.000 | 0.117 | 0.000 | 0.000 | 0.001 | 0.341 | 0.000 | 0.000 |
| FZ-1F | 0.343 | 0.314 | 0.043 | 0.028 | 0.393 | 0.334 | 0.047 | 0.045 |
| GCH-FZ | 0.095 | 0.358 | 0.608 | 0.000 | 0.188 | 0.419 | 0.473 | 0.000 |
| Hybrid | 0.002 | 0.082 | 0.700 | 0.000 | 0.007 | 0.064 | 0.629 | 0.000 |