

# Mean-variance portfolio optimization: do historical correlations help or hinder risk control in a crisis?

Tony Plate  
Plate Consulting, LLC  
Santa Fe, NM

[tplate@acm.org](mailto:tplate@acm.org)

(\*some of this research performed at  
Black Mesa Capital)

Expanded version of presentation with additional slides and explanations, 2010/05/29.

# Mean-Variance optimization

- Introduced by Markovitz, 1952 & 1959
- Widely used, in various forms (from several to thousands of instruments)
- Requires statistical models of two aspects of returns:
  - Mean – i.e., predictions of future returns
  - Covariance – focus on factor models for equities
- Optimize this objective function over a portfolio  $x$   
$$\{expected\ return\} - \lambda * \{expected\ volatility\} \quad s.t. \quad sum(x)=1$$
- In practical use, the objective function is often augmented with transaction cost penalties, and constraints are imposed

# Quotes

From “The Emperor has no Clothes: Limits to Risk Modeling” (Jon Danielsson, 2001)

- “The fundamental assumption in most statistical risk modeling is that the *basic* statistical properties of financial data during stable periods remain (almost) the same as during a crisis. This of course is not correct.”
- [Unlike forecasting the weather] “forecasting risk does change the nature of risk”
- “Correlations typically underestimate the joint risk of two or more assets”

*Note: Danielsson’s paper mostly focused on the use VaR, but his comments were also directed at the use of correlations*

# Quotes

From “Nickels and Dimes: Izzy Englander's Growth Strategy for Millennium” (Stephen Taub, Institutional Investor Magazine, 20 May 2009)

“Millennium Management founder Izzy Englander delivers consistently good returns by making low-risk bets that other hedge fund managers often ignore...”

“Millennium does not put much emphasis on the correlation between different investment strategies when deciding how to allocate capital. “That [approach] would have broken down last year,” points out Larkin, 34, who has day-to-day oversight responsibility for Millennium’s equities operation. “In a stress environment correlation goes to 1,” he says.

# Quotes

From “Time-varying Stock Market Correlations and Correlation Breakdown” (David Michael Rey, 2000)

[Abstract]

“... recent experience has highlighted the fact that international equity correlations can rise quickly and dramatically. These so-called correlation breakdowns call into question the usefulness of diversifying and hedging operations based on correlations estimated from historical data, since they may be inaccurate precisely when they are most desired, namely in periods of high volatility or worse, extreme negative price movements...”

# Literature Background

- Research on “correlation breakdown” shows that in a crisis:
  - Variance increases
  - Correlations between instruments increase
  - Correlations at extreme returns can be different than at normal returns
- Most research on “correlation breakdown” addresses international equity markets, i.e., international indices are the instruments
- In literature on correlation breakdown there is ubiquitous idea that a single “market factor” drives the correlations, but few mentions of factor models

# Question for this talk

- Can a factor model capture some of the dynamic statistics of variance and correlations of US equities during a crisis?
- Is mean-variance optimization (MVO) potentially useful or harmful during a crisis?
- Does the use of correlations in MVO hurt, help, or not matter during a crisis?
- Keep everything very simple and focus solely on risk control
- Recognize that the statistical assumptions of MVO are not satisfied by markets, esp during a crisis

# Minimum Variance Portfolios

Classic “plain vanilla” mean-variance optimization objective function:

$$G(x) = r^T x - \lambda x^T V x \text{ such that } e^T x = 1$$

- $x$  is the vector of positions
- $e$  is the column vector of 1's
- $r$  is the vector of forecasted returns
- $V$  is the co-variance matrix
- $\lambda$  is the risk aversion

(NB: in any real application, would use a version with constraints!)



# Closed-form solution

Solution:

$$h(\lambda) = \frac{V^{-1}r}{2\lambda} + \left(1 - \frac{e^T V^{-1}r}{2\lambda}\right) \frac{V^{-1}e}{e^T V^{-1}e}$$

For very large  $\lambda$  (infinite), we get the minimum risk portfolio:

$$h_c = \frac{V^{-1}e}{e^T V^{-1}e}$$

*Note: This very simple formulation of mean-variance optimization has an easily computable closed-form solution*

*For maximum risk aversion, the dependence on  $r$  drops out and the minimum risk portfolio is a function of just the covariance matrix*

R code:

```
> iV <- solve(V)
> h_c <- rowSums(iV) / sum(iV)
```

More efficient:

```
> iv1 <- solve(V, 1)
> h_c <- iv1 / sum(iv1)
```

(\*beware rank-deficient V !)

# Focus on minimum-variance portfolio

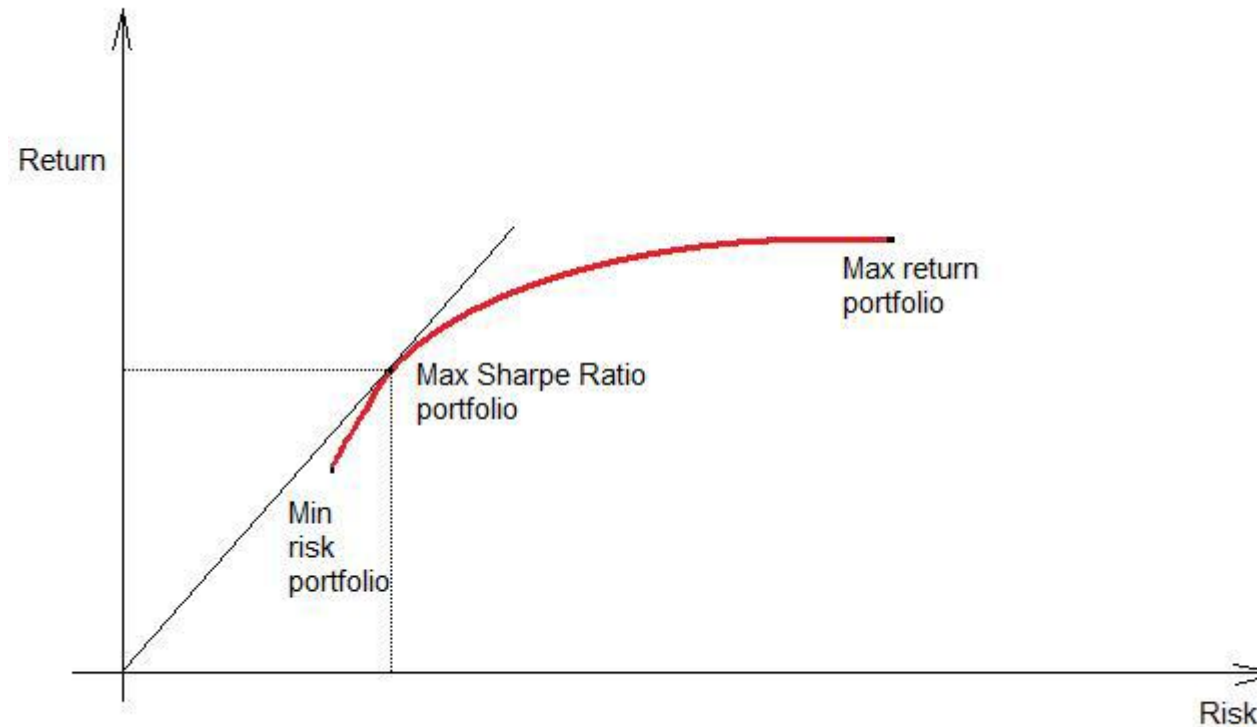
- Full solution is linear blend of  $V^{-1}r$  and minimum-variance portfolio

$$h(\lambda) = \frac{V^{-1}r}{2\lambda} + \left(1 - \frac{e^T V^{-1}r}{2\lambda}\right) \frac{V^{-1}e}{e^T V^{-1}e}$$

- Full solution includes Sharpe-ratio optimal portfolio (by choice of  $\lambda$ )
- By studying the minimum variance portfolio, focus on the risk-control properties of mean-variance optimization
- Big caveats: lack of constraints, & robustness!

*Note: The rationale for just focusing on minimum-variance portfolios is that the full solution is a linear blend of a function of  $r$  and  $V$ , and the minimum variance portfolio*

# Frontier of optimal risk-return portfolios



From “Maximizing the Sharpe Ratio and Information Ratio in the Barra Optimizer” Leonid Kopman and Scott Liu, 2009

# Assessment of risk models

Possible performance measures for risk models:

- Log-likelihood of data under model (either in- or out-of-sample)
- Bias statistics (how many times realized risk is greater than a threshold based on estimated risk)

MVO depends on assumptions that are not true in financial data (i.e., returns are iid from a Gaussian)

Treat MVO as a heuristic, and measure aspects of performance that matter, e.g.,

- Realized std dev of returns of optimized portfolios
- Realized drawdown/max-deviation in a moving window of cumulative returns of optimized portfolios

# Factor models for risk

Decompose risk into three components:

- $L$  : Factor loadings ( $n \times k$  matrix)
- $F$  : Factor covariance ( $k \times k$  matrix)
- $S$  : Specific risk ( $n$  element vector)

*E.g., 2000x60 factor model has approx 125,000 free parameters, compared to 2000x2000 variance matrix with > 2,000,000 free parameters*

$$V = LV_f L^T + \text{diag}(S^2)$$

$$\begin{array}{c}
 \begin{bmatrix} V_{11} & V_{12} & & V_{1n} \\ V_{21} & V_{22} & & \\ & & \ddots & \\ V_{n1} & & & V_{nn} \end{bmatrix} \\
 V
 \end{array}
 =
 \begin{array}{c}
 \begin{bmatrix} L_{11} & L_{12} \\ L_{21} & L_{22} \\ \vdots & \vdots \\ L_{n1} & L_{n2} \end{bmatrix} \\
 L
 \end{array}
 \begin{array}{c}
 \begin{bmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{bmatrix} \\
 F
 \end{array}
 \begin{array}{c}
 \begin{bmatrix} L_{11} & L_{21} & \cdots & L_{n1} \\ L_{12} & L_{22} & \cdots & L_{n2} \end{bmatrix} \\
 L^T
 \end{array}
 +
 \begin{array}{c}
 \begin{bmatrix} S_1^2 & 0 & & 0 \\ 0 & S_2^2 & \ddots & \\ & \ddots & \ddots & 0 \\ 0 & & 0 & S_n^2 \end{bmatrix} \\
 S^2
 \end{array}
 \end{array}$$

Good way of capturing risk behavior and avoiding having to estimate too many parameters with many instruments

# Decomposition of factor-covariance matrix

Factor covariance matrix further decomposed:

- $\sigma_i^2$  : Factor variances ( $k$  element vector)
- $\rho_{ij}$  : Factor correlations ( $k \times k$  matrix)

$$F_{ij} = \sigma_i \sigma_j \rho_{ij}$$

Useful because it allows correlations and variances to be estimated with different techniques

# Factor models & rapidly changing correlations

- Factor models can model rapidly changing correlations if variance components can adapt quickly (cf Garch estimators)
- Consider single factor model with unit loadings, i.e.,  $L=[1,1,\dots,1]^T$  and factor variance  $\sigma^2$
- Then, the instrument variance matrix will be:

$$\begin{matrix}
 S_1^2 + \sigma^2 & \sigma^2 & & \sigma^2 \\
 \sigma^2 & S_2^2 + \sigma^2 & \ddots & \\
 & & \ddots & \sigma^2 \\
 \sigma^2 & \ddots & \sigma^2 & S_n^2 + \sigma^2
 \end{matrix}$$

*Note: In the literature on “correlation breakdown” there is the ubiquitous idea of a single market factor whose behavior becomes dominant in a crisis, and consequently drives all correlations to 1.*

- If  $\sigma^2$  increases compared to  $S_i^2$ , all correlations go up

# Risk model construction: Loadings

- For loadings, use a style + industry scheme:
- Five style loadings, transformed by rank into  $[-0.5, 0.5]$ 
  1. Beta: coef of time-series regression of instrument returns to uniformly-weighted market returns over last 126 days
  2. Size: Market Cap (in dollars)
  3. SD63: std deviation of last 63 days of returns
  4. BetaSD:  $\text{Beta} * \text{SD63}$
  5. Mom: 252 day cumulative return
- Industry loadings are 0/1 to indicate membership one of approx 60 GICS industry codes
- Gives an  $n \times 65$  loadings matrix



# Risk model construction: universe

- Risk model is estimated based on observed price changes
- Restrict estimation universe to set of liquid stocks:
  - Price > \$5
  - Median share volume over last 15 days > 20,000
  - Median dollar volume over last 15 days > \$500,000
  - Above conditions have to have been satisfied for past 63 days to “pass” the liquidity filter

# Risk model construction: Factor returns

- Factor returns computed daily by linear regression (cross-sectional) of stock returns on loadings

$$r_{ti} = R_{t1}L_{ti1} + R_{t2}L_{ti2} + \dots + R_{tk}L_{tik} + \varepsilon_{ti}$$

- $r_{ti}$  is the return of stock  $i$  on day  $t$
- $R_{tk}$  is the return of factor  $k$  on day  $t$
- $L_{tik}$  is the loading for factor  $k$  of stock  $i$  on day  $t$
- $\varepsilon_{ti}$  is the residual (return) for stock  $i$  on day  $t$
- Variation: include an intercept, but as is redundant, set to:  $M_t = \frac{1}{n} \sum_i r_{ti}$  (cf Multilevel modeling)
- And then regress  $(r_{ti} - M_t)$  on loadings as above

# Risk model construction: factor correlations

- Computing correlations for financial data can be tricky because often have incomplete time-series
  - Stocks/instruments come into and out-of-existence and/or may not trade some days
  - Factors (industries) get added and dropped
- Very important to have a valid (positive semi-definite correlation/covariance matrix), non-PSD causes problems in the optimization
- Standard methods of getting valid correlation matrix completely drop days with incomplete obs – unacceptable here!
- Use complete pairwise observations `cor(..., use="pairwise.complete.obs")` – produces non-PSD result
- “Fix” resulting correlation matrix using algorithm of Higham 2002 – finds PSD correlation matrix that is closest in L2 norm

# Algorithm for “fixing” a correlation matrix

```
makeCorMatrixPSD <- function(A, tol = 1e-007, maxiter = 10 * ncol(A)) {
  # Based on the algorithm of Higham 2002 which uses alternating projections:
  #   repeat:
  #     compute eigenvalues, set negative eigenvalues to zero
  #     reconstruct matrix from corrected eigen decomposition, set diagonals to one
  dS <- 0
  Y <- A
  k <- 0
  while( k < maxiter ) {
    # Dykstra's correction for R:
    R <- Y - dS
    # Project R onto non-negative eigenvalues:
    e <- eigen(R, symmetric = T)
    X <- e$vectors %*% diag(pmax(e$values, 0)) %*% t(e$vectors)
    dS <- X - R
    # Project onto unit diagonal:
    Y <- X
    Y[cbind(seq(len = nrow(Y)), seq(len = nrow(Y)))] <- 1
    k <- k + 1
    if ( mean(abs(diag(Y) - diag(X))) < tol )
      break
  }
  return(Y)
}
#*! All error and sanity checking removed from above version! See end pages for
complete version
```

# Full code for makeCorMatrixPSD()

```
makeCorMatrixPSD <- function(A, tol=1e-7, maxiter=10*ncol(A), verbose=FALSE) {
  # Make a symmetric matrix into correlation matrix by making it
  # positive semi-definite using an iterative approach.
  # Return the closest PSD correlation matrix to A in terms of the Frobenius norm
  # Based on the algorithm of Higham 2002 which uses alternating projections:
  #   repeat:
  #     compute eigenvalues, set negative eigenvalues to zero
  #     reconstruct matrix from corrected eigen decomposition, set diagonals to one
  #
  A.attr <- attributes(A)
  if (length(dim(A))!=2 || nrow(A) != ncol(A))
    stop("A must be a square matrix")
  # Replace NA values with their corresponding transpose element
  if (any(A.na <- is.na(A)))
    A[] <- ifelse(!A.na, A, t(A))
  # Replace NA diagonals with 1 and NA off-diagonals with zero
  ii <- which(is.na(A), arr=TRUE)
  if (nrow(ii)) {
    on.diag <- ii[,1] == ii[,2]
    if (any(on.diag)) A[ii[on.diag,,drop=FALSE]] <- 1
    if (any(!on.diag)) A[ii[!on.diag,,drop=FALSE]] <- 0
  }
  # Make sure the matrix is symmetric -- otherwise can get complex eigen values
  tA <- t(A)
  if (any(abs(tA - A) > 0)) {
    # only warn if lack of symmetry > tol (use abs tol val here because we are working
    # with a correlation matrix which has a natural scale of -1/1
    if (any(abs(tA - A) > tol))
      warning("making non-symmetric correlation matrix symmetric")
    A <- (A + tA) / 2
  }
  dS <- 0
  Y <- A
  k <- 0
  while (k==0 || k < maxiter) {
    # Dykstra's correction for R:
    R <- Y - dS
    # Project R onto non-negative eigenvalues:
    e <- eigen(R, symmetric=TRUE)
    X <- e$vectors %*% diag(pmax(e$values, 0)) %*% t(e$vectors)
    dS <- X - R
    # Project onto unit diagonal:
    Y <- X
    Y[cbind(seq(len=nrow(Y)), seq(len=nrow(Y)))] <- 1
    k <- k+1
    if (verbose)
      catn("iter", k, "closeness to diag=", format(mean(abs(diag(Y) - diag(X))))
    conv <- mean(abs(diag(Y) - diag(X)))
    if (conv < tol)
      break
  }
  if (conv >= tol)
    warning("failed to achieve convergence tol=", tol, " in ", k, " iterations; reached ",
format(conv))
  e <- eigen(Y, symmetric=TRUE, only.values=TRUE)

  attributes(Y) <- A.attr
  structure(Y, psdinfo=list(convergence=(conv < tol), delta=conv, dist=sqrt(sum((A-Y)^2)),
rank=sum(abs(e$values)>tol*10), iters=k))
}
```

```
makeCovMatrixPSD <- function(A, tol=1e-7, maxiter=10*ncol(A), verbose=FALSE) {
  dd <- diag(A)
  if (any(i <- is.na(dd) | dd<=0)) dd[i] <- 1
  scale <- sqrt(dd)
  C <- A * outer(1/scale, 1/scale, FUN="*")
  X <- makeCorMatrixPSD(C, tol=tol, maxiter=maxiter, verbose=verbose)
  return(X * outer(scale, scale, FUN="*"))
}
```

# Risk model construction: factor variances

- Many possibilities, e.g.,
  - Moving window variance (used 63-day window)
  - Exponentially weighted variance (weight recent observations more heavily) (used 63-day half-life)
  - GARCH
  - Predictors that take into account other measures or predictors of volatility, e.g., VIX
- Results described here use moving window and exp weighted variance measures

# Risk model construction: Specific risk

- “Specific risk” is the idiosyncratic risk of stocks, after factor risk is removed
- Calculate from residual returns  $\varepsilon_{ti}$
- Has a lot of persistence & quarterly periodicity

- Use two part model
- $$S_{ti} = 100\sqrt{252} \frac{\sigma_{ti}(252)}{\text{med}_j(\sigma_{tj}(252))} \left( 0.32 \text{med}_j(\sigma_{tj}(63)) + 0.68 \text{med}_j(\sigma_{tj}(5)) \right)$$

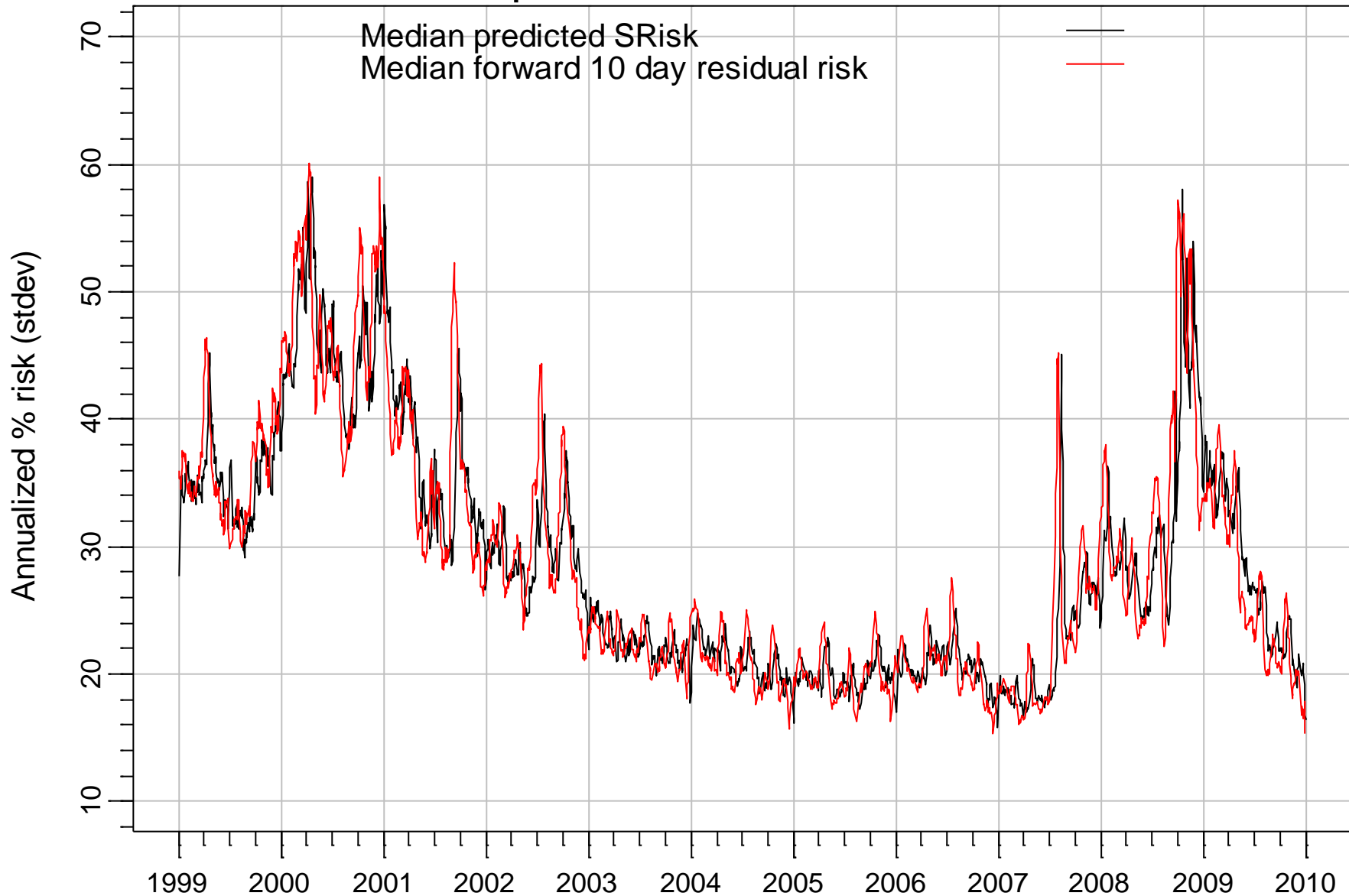
Relative risk for stock i

Market level of specific risk

- $\text{med}_j(\sigma_{tj}(n))$  is cross-sectional median (at time  $t$ ) of trailing  $n$ -period standard deviation of residual return over stocks  $j$  passing liquidity filter
- $\sigma_{ti}(252)$  is the trailing 252-day standard deviation of residual return of stock  $i$  (at time  $t$ )
- The  $100 \sqrt{252}$  factor converts to annualized percentage, the 0.32 and 0.68 factors come from empirical modeling

# Specific risk behavior

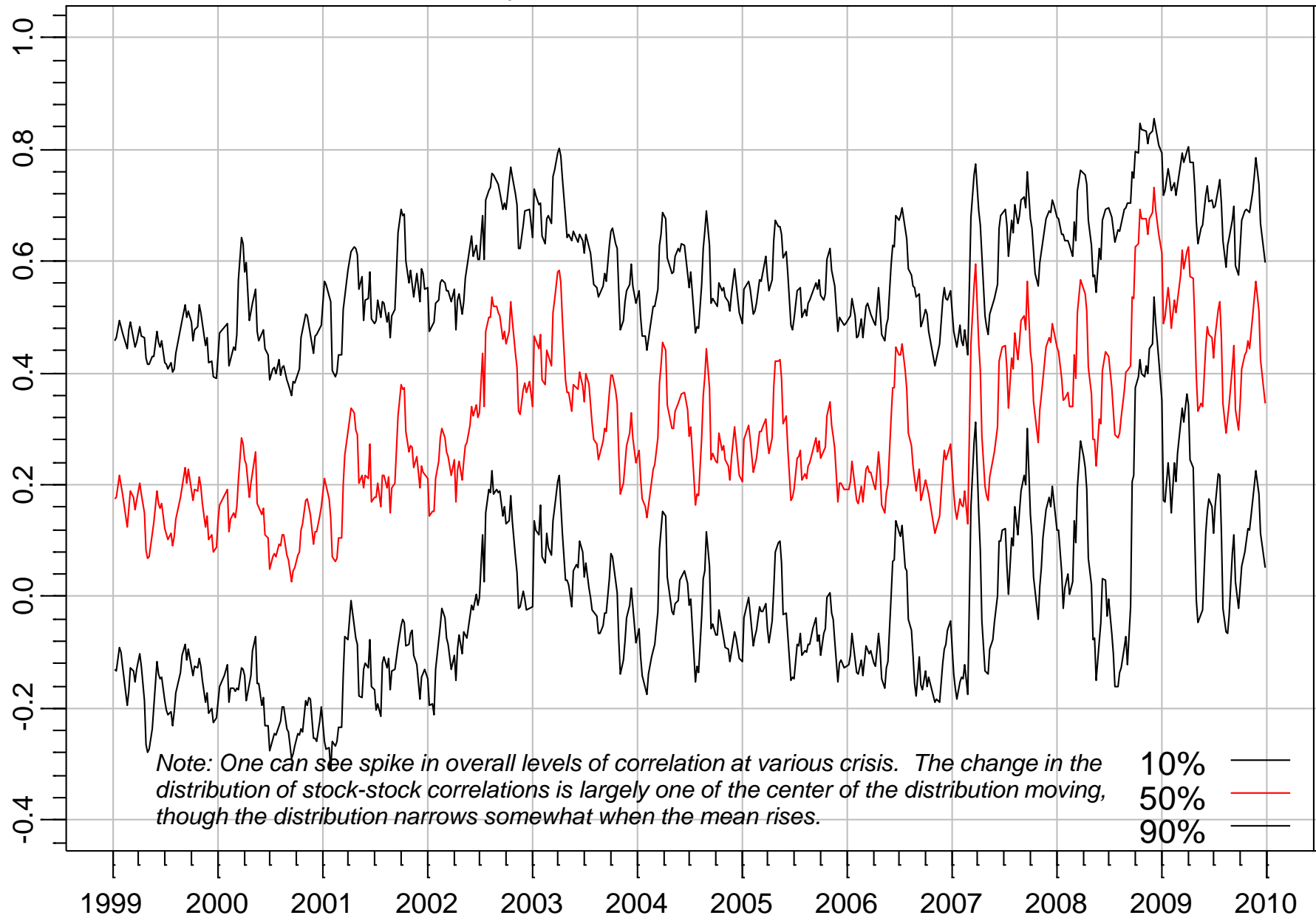
## Forward and predicted SRisk for GIBRM2 model



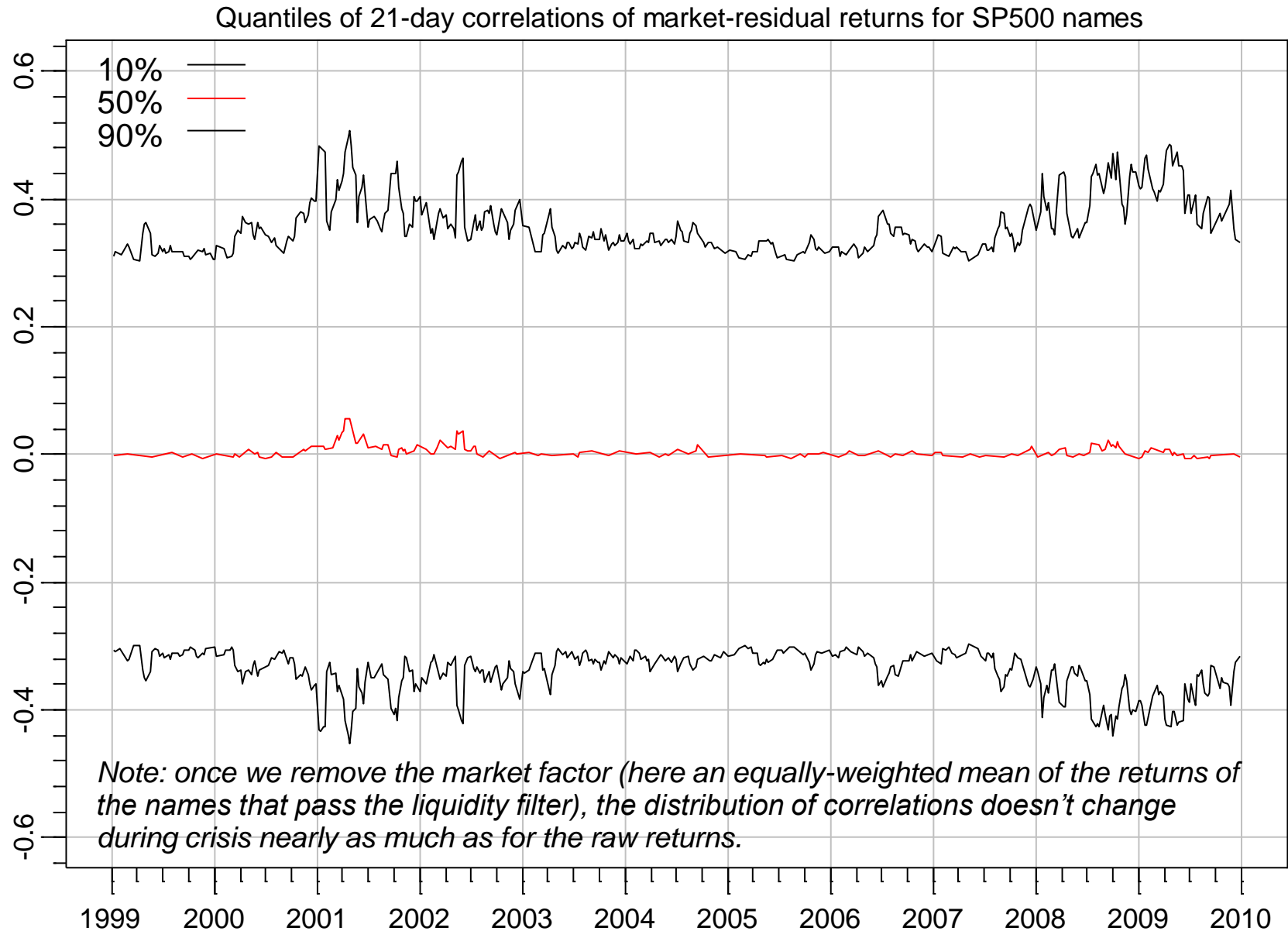


# Empirical correlations of returns b/w SP500 names

Quantiles of 21-day correlations of returns for SP500 names

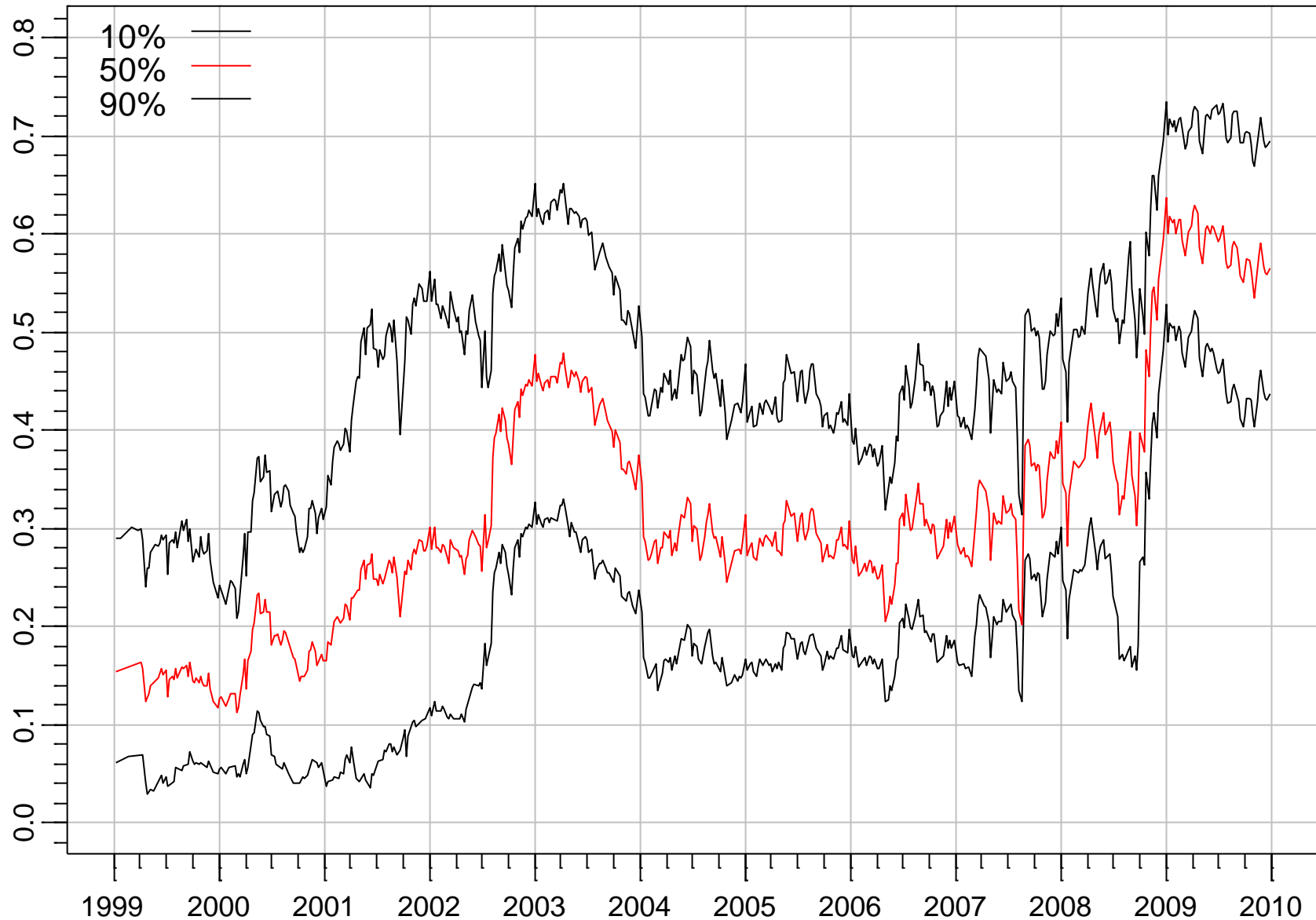


# Empirical correlations of market-residual returns for SP500 names



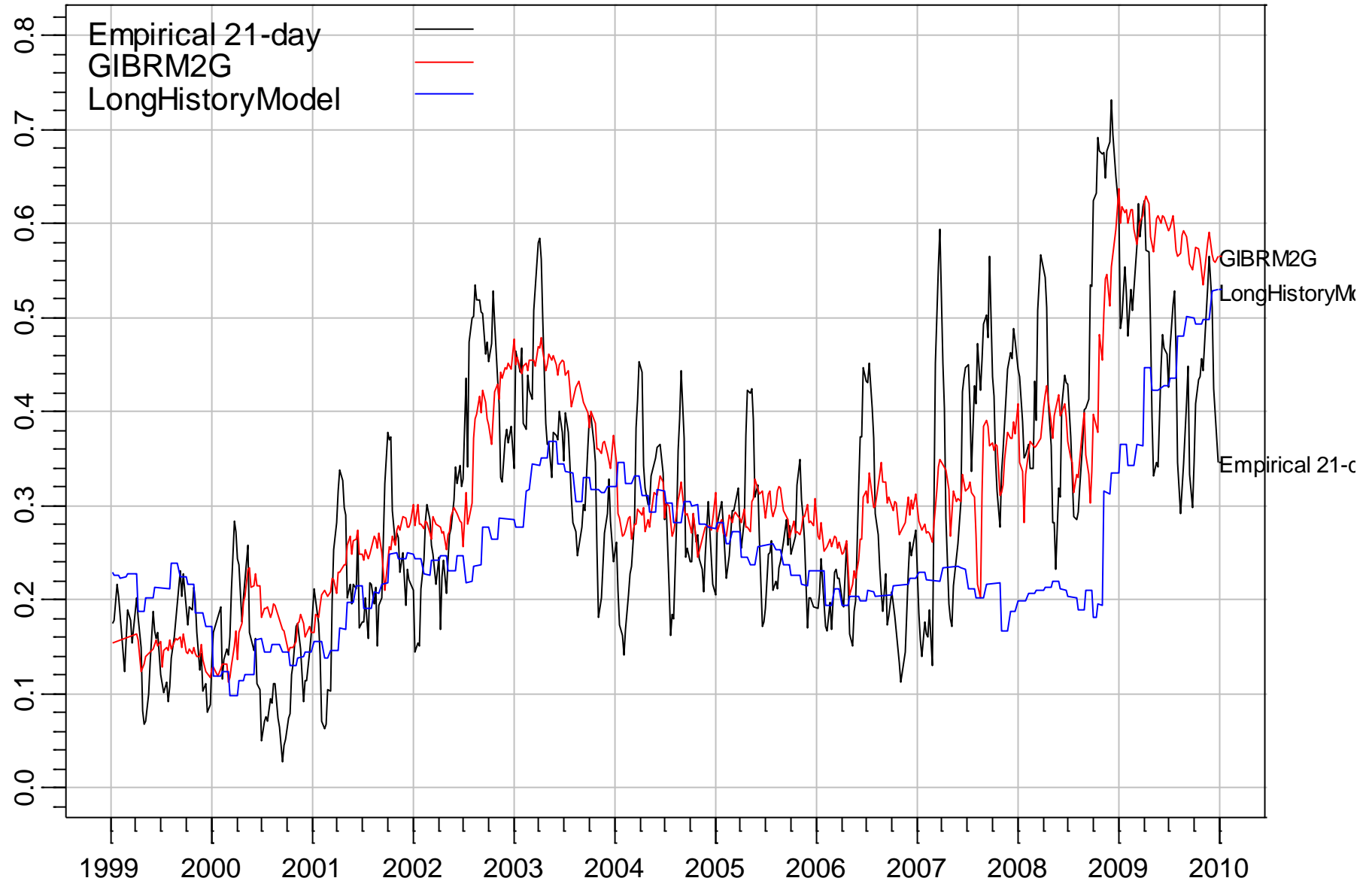
# Modeled correlations for medium-history model

Quantiles of correlations for SP500 names from GIBRM2G model

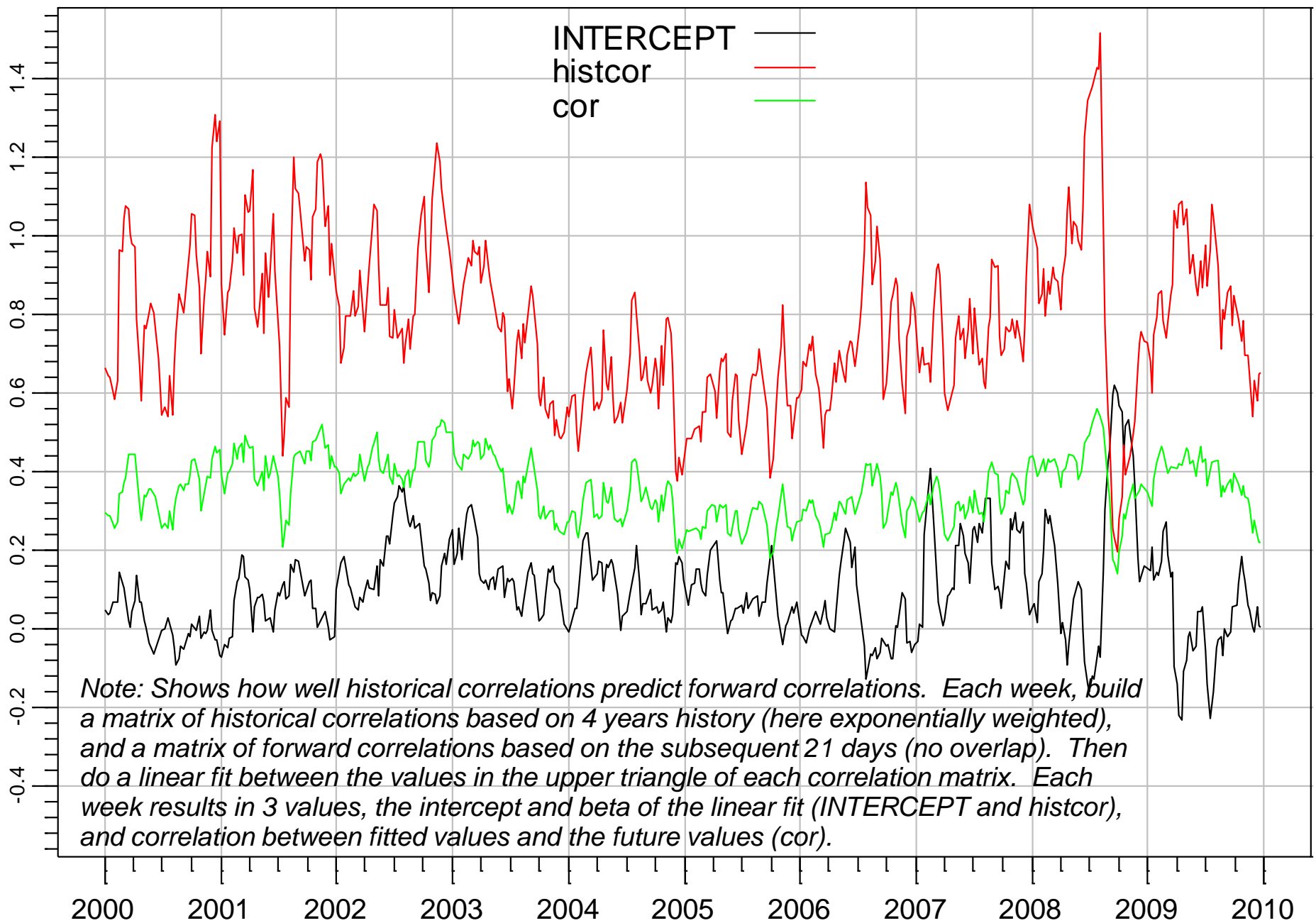


# Comparing median raw correlations

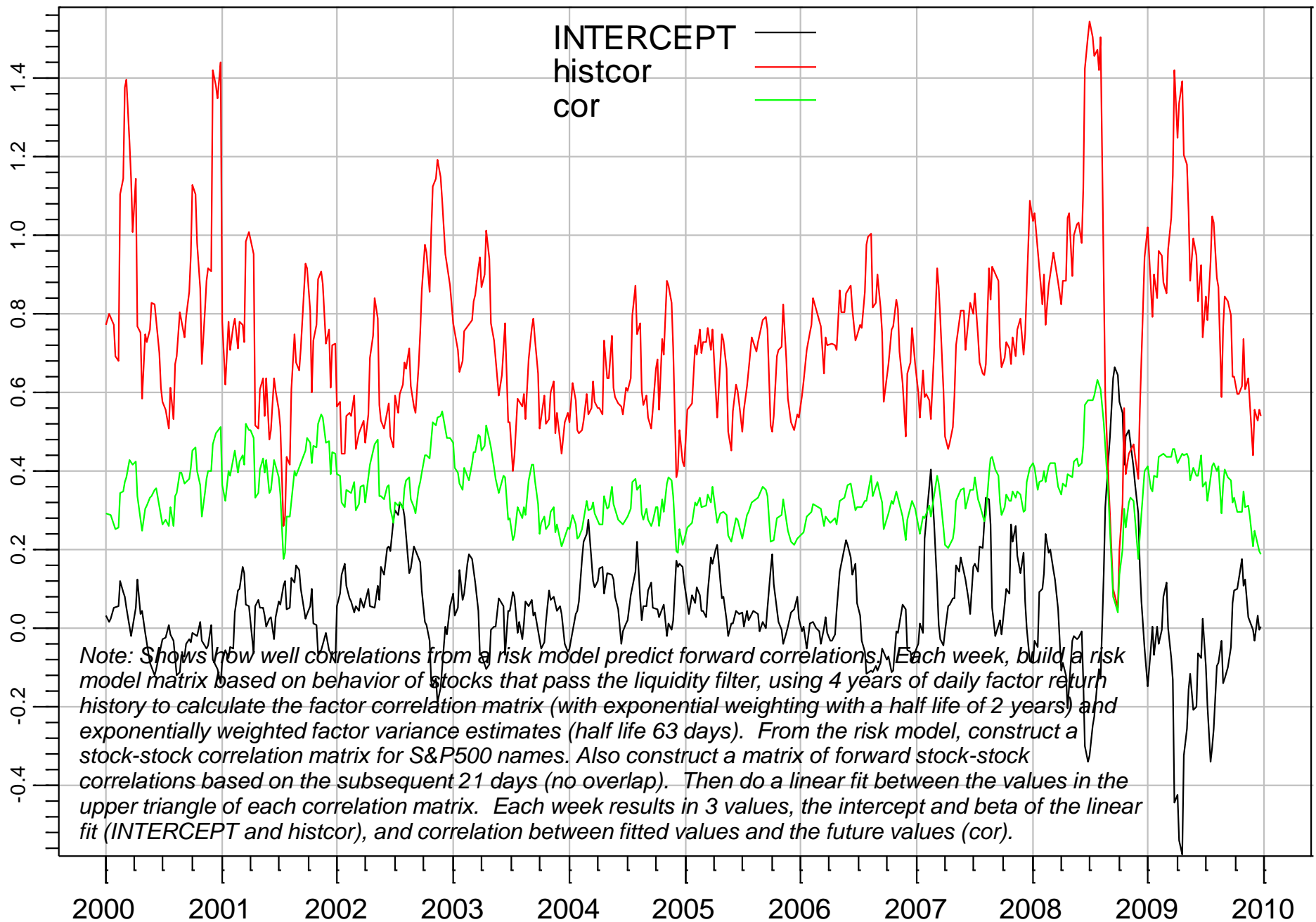
## Medians of correlations for SP500 names



# Linear fit between hist exp-wght corr (2yr halflife) and 21-day forw corr for S&P500 names



# Linear fit between GIBRM2G.MR.corxw2y.uvarxw63d corr and 21-day forw corr for S&P500 names



# Optimization experiments

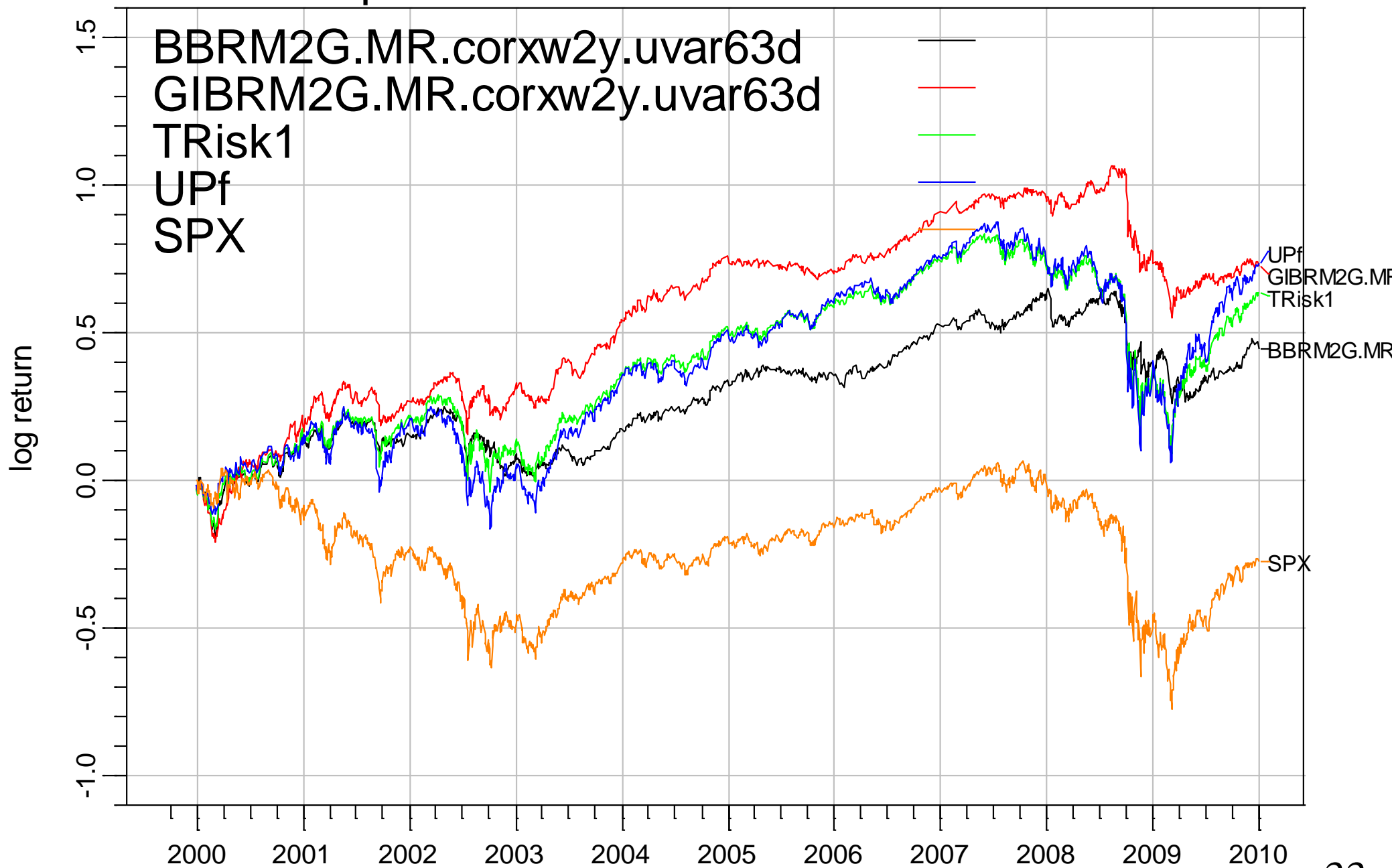
- Report optimization experiments using a variety of risk models
- Use period 2000-2009, S&P 500 stocks for optimization
- Rebalance each portfolio every 5 business days, calculate PnL every day
- Always use causally valid risk models – for optimizing portfolio at any time, risk model is built from data available strictly before then
- Risk model estimated on a universe of the most liquid US stocks (exact number varies over time, typically between 1500 and 2000 names pass the liquidity filter)
- Look at performance in “normal” (2003 thru 08) periods vs the crisis of 2008/09 (Sept 08 thru Mar 09)
- To understand whether performance differences are meaningful, use random subset portfolios
  - Evaluate PnL for an optimized portfolio constructed from each of 200 random subsets (2/3) of the S&P500

# Variations of risk models

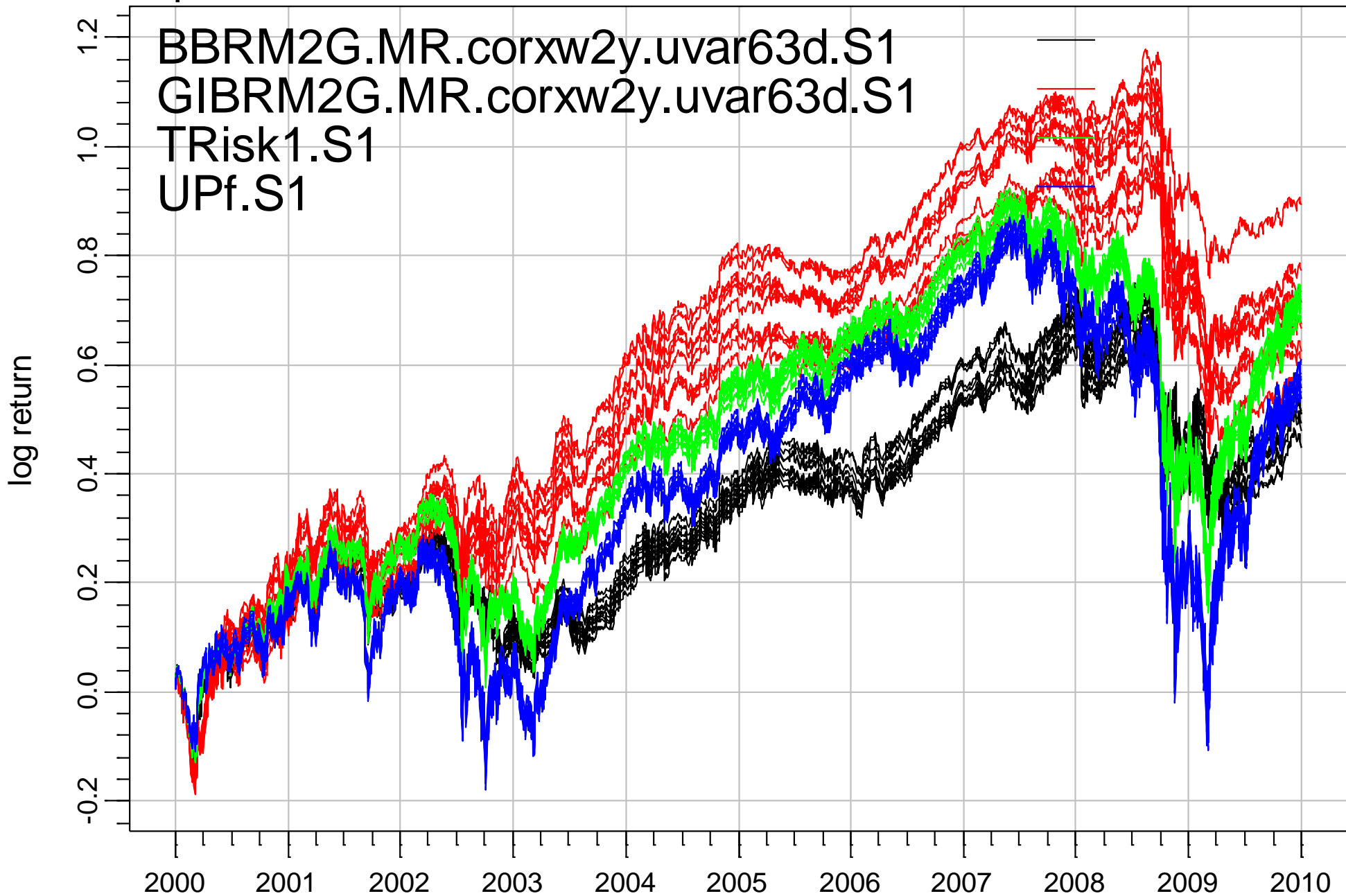
- Simple schemes:
  - UPf: uniform weight portfolio
  - TRisk1: weight of a name is prop to inverse of total risk (variance)
- Loadings:
  - MKTRMG: Intercept only
  - BBRM2G: Beta factor + Intercept
  - GIBRM2G: 5 style factors, approx 60 industries
  - GIBRM2OG: like GIBRM2G, but factors mutually orthogonalized each week (first make industries zero-mean, then orthogonalize remaining factors to industries, and make them mutually orthogonal)
- Factor return fitting technique:
  - MR: use Intercept together with industry membership factors, set to mean return (MR="market residual")
  - VN: no Intercept (VN="Vanilla") (when used with industry membership factors)
- Correlation matrix technique:
  - cor: 2 year moving window
  - corxw: 4 year moving window with exponential weighting with 2-year half life
  - coraz: off-diagonal elements are zero (to remove dependence on historical correlations)
- Factor variance technique
  - uvar: 63-day moving window
  - uvarxw: 126-day moving window with exponential weighting with 63-day half-life



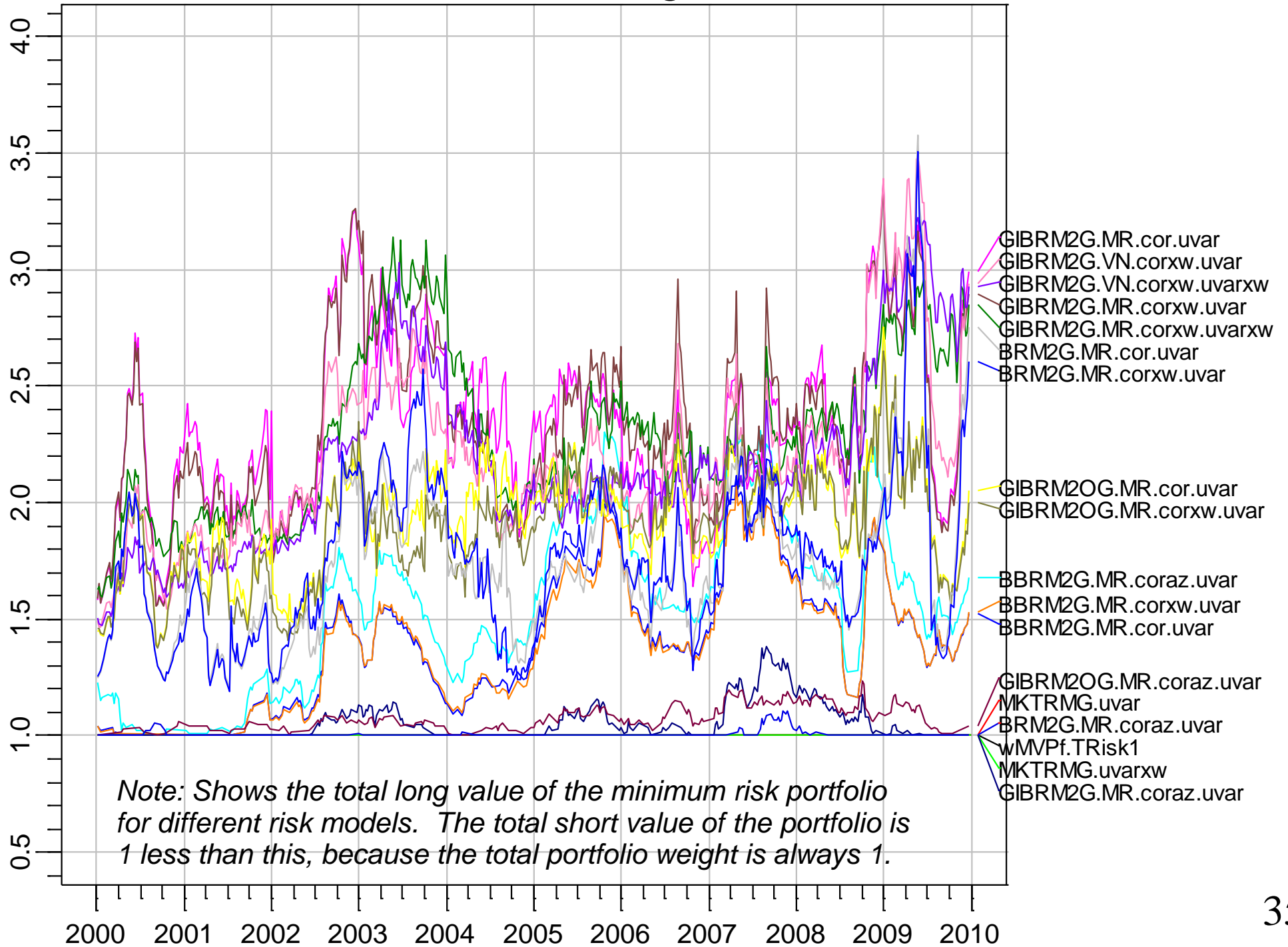
# Compounded PnL for full S&P500 universe

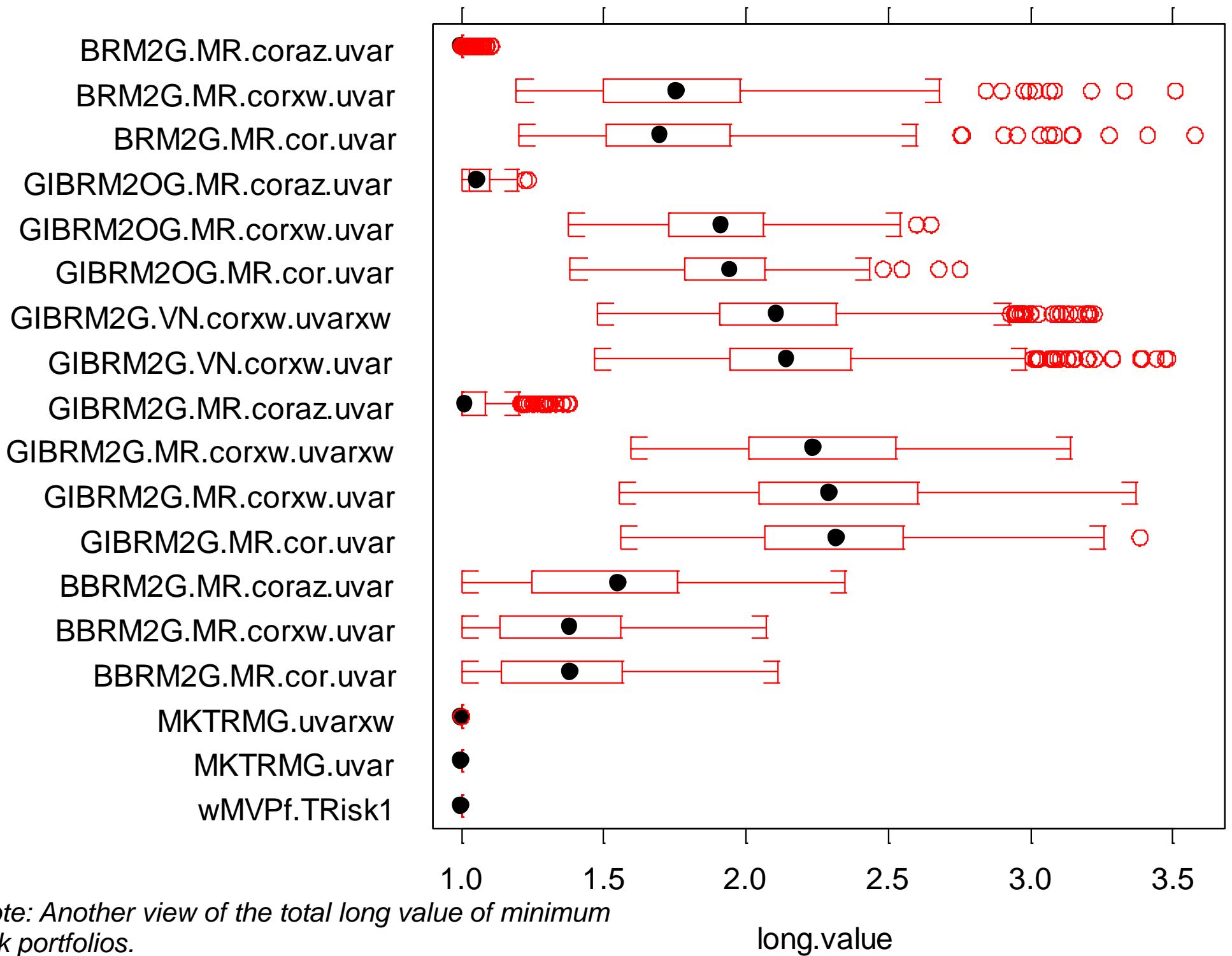


Compounded PnL for 10 random subsets of S&P500 for each scheme



# rs200.res.long





Note: Another view of the total long value of minimum risk portfolios.

# Portfolio performance results

Next three slides show performance of portfolios (200 replications for each portfolio strategy)

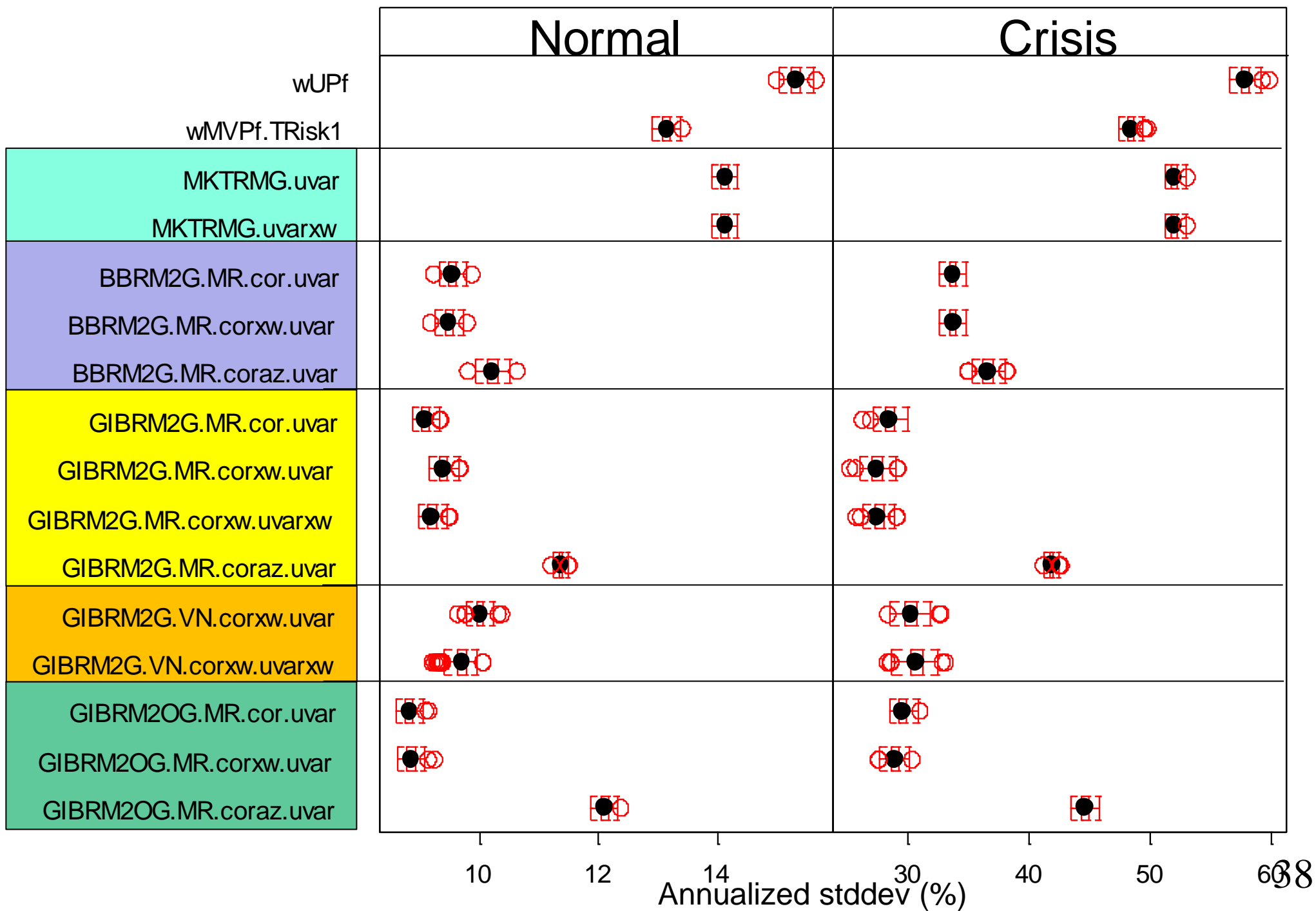
3 performance measures:

- Standard deviation (annualized)
- Max portfolio value range in 21 day period
- Max portfolio value range in 42 day period

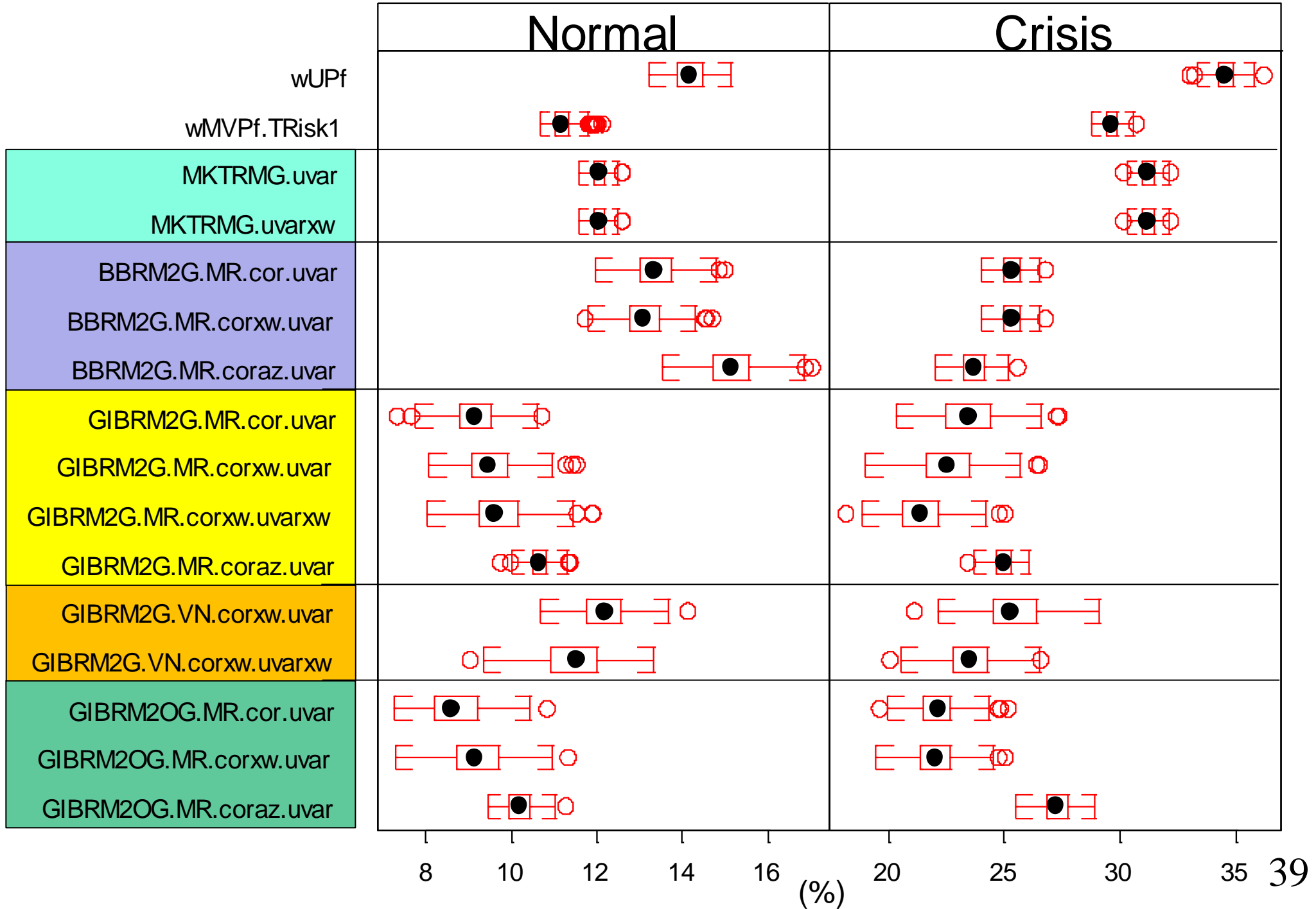
Observations:

- Full risk models (industry + style factors: GIBRM2.\*) do better than smaller models (MKTRMG.\* and BBRM2G.\*)
- Risk models with factor correlations set to zero (\*.coraz) do worse in a crisis
- A risk model with good overall performance is GIBRM2G.MR.corxw.uvarxw
  - Has overall market factor (MR) removed prior to multiple factor regression
  - Uses 5 styles factors + GICS industry factors
  - Uses exponentially weighted correlations with half life of 2 years
  - Uses exponentially weighted factor variances with half life of 63 days
- Some indication that the factor model with orthogonalized factors does well, esp on the performance measure of 42 day portfolio value range

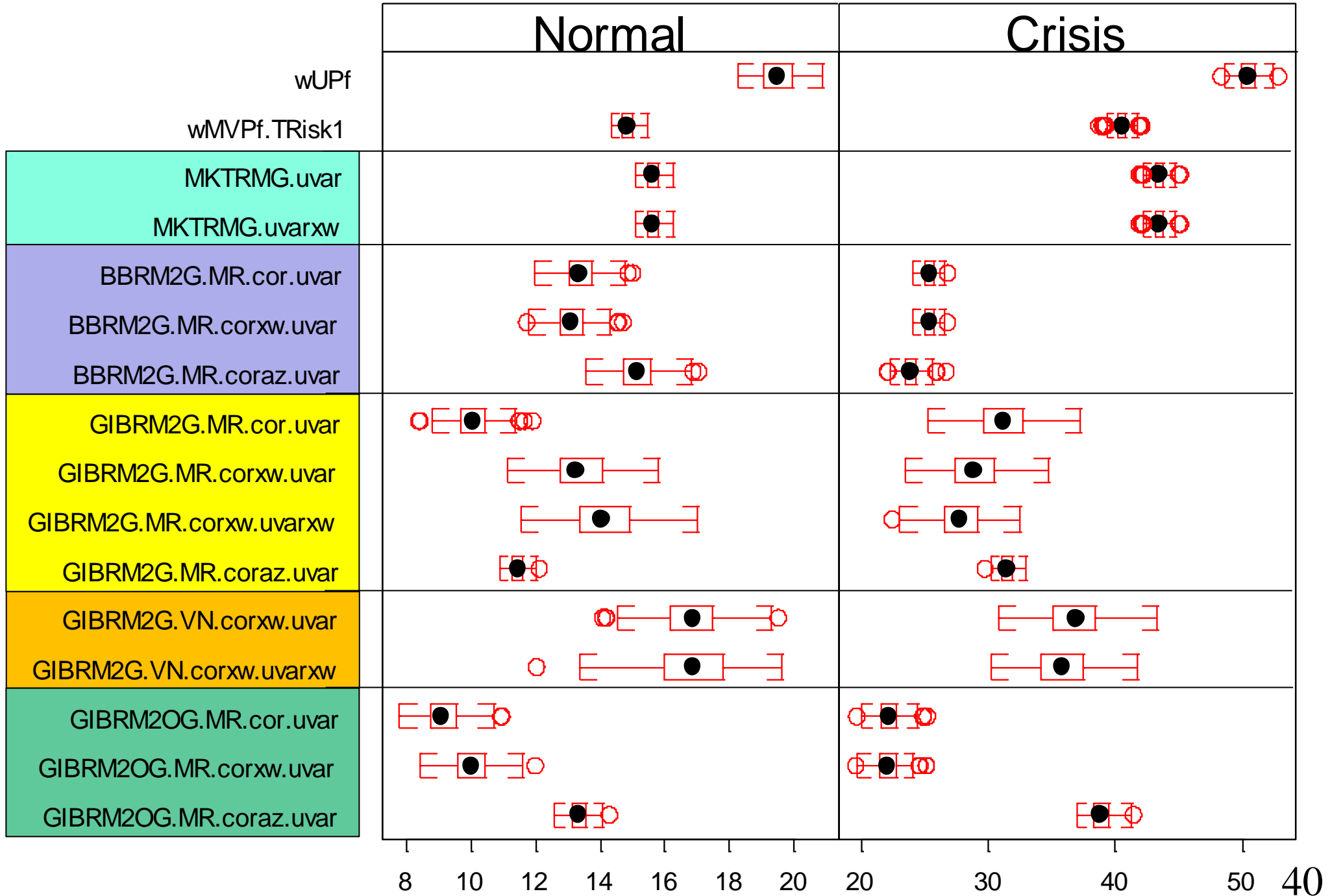
# Std dev of portfolio returns



# Max range of cumulative portfolio return in any 21 day subperiod



# Max range of cumulative portfolio return in any 42day subperiod





# Conclusions

- Initial goal was to show that correlations are so unstable in a crisis that it was better to ignore them all the time
- Minimum Variance Portfolios for S&P500 equities based on optimizing with a factor model showed:
  - Ignoring correlations doesn't help
  - A model based on historical correlations does better than simple alternatives
- Maybe the emperor is at least wearing underpants, after all