Volatility and Spillover: Facts of Market Life

A Preprint

William G. Foote *
Department of Business Analytics
Manhattan College
Riverdale, NY 10471
wfoote01@manhattan.edu

Brian Wholey
Manhattan College
Riverdale, NY 10471
bwholey01@manhattan.edu

June 3, 2022

Abstract

Volatility and the interaction of markets continue to beguile traders, managers, investors and regulators. In this preliminary draft we use a working example from the renewable energy industry to develop three work flows for financial time series: univariate empirical characterizations; quantile regression spillover analysis; and bayesian multi-level hierarchical generation of a stratified industry risk structure. This latter flow deploys a Pareto-smoothed importance sampling with leave-one-out cross-validation to investigate uncertainty and variability of market events, especially so-called outliers.

***K*eywords** market pillover · volatility clustering · quantile regression · bayesian data analysis · multi-level hierarchical model · pareto-smoothed importance sampling · leave-one-out cross validation

1 What are the stylized facts?

A tool available to financial risk managers is the analysis of market spillover and volatility clustering. When one market becomes entangled with another through the normal course of trade, is it possible for the volatility of one to affect the volatility of another? Theoretically, of course it is possible. However, the only way to gauge the impact of one market on another another is through some sort of association analysis, Here we will use simple linear regression techniques to detect the existence and degree of market spillover.[2]

The experience of risk management is that spillover and other market observations are so common, across multiple market regimes, as to confer the status of fact. The supposition of such an analysis would be that managers would ignore such regularly occuring observations to their detriment. Here is a short compendium of the usual suspects for univariate series.[3]

- Risk factor changes exhibit little or no memory, evidenced by significant but less impactful dependency on univariate autocorrelations at any lag.
- The conditional expectation of returns is nearly always zero.
- On the other hand, volatility of the same series exhibits very slowly decaying autocorrelations with power law tail thicknesses.

*Thanks to numerous colleagues and financial engineering students at Manhattan College and Syracuse University for the organic development of the ideas in this paper.

[2]Theory and practice are themselves deeply entangled in a dialectical circle. We suppose we are naive investors with an account at our favorite online brokerage with three assets. We know nothing at the outset but profit and loss. We aim to learn a bit more about the three assets starting with their history. Our characterization of this history will inform us of the deeper reasons why, and how, our profit and loss statements evolve.

[3]Cont (2001) is generally recognized as the first cross asset survey of co-called stylized facts of financial markets. McNeil, Frey, and Embrechts (2015) devote Chapter 3 to a discussion about the existence of market factors for univariate and multivariate financial time series.

- Return series are left skewed, while volatility series are naturally right skewed but witb persistently thick tails.
- Return series tend to have thick tails with clusters of volatility.

This last observation about volatility in practice means that the conditional standard deviations of returns are themselves volatile. Volatile volatility is indeed the definition of highly leptokurtotic returns.[4] That the returns cluster reveals a story of the persistence of high volatility as well as the persistence of low volatility environments.

On the multivariate front, that is, the portfolio side of the world of markets we observe these hard to write off facts.[5]

- Returns exhibit little or no non-contemporaneous cross-correlation.
- Volatilities of returns (e.g., absolute values of returns) exhibit strong and weakly decaying, thus persistent, cross-correlations.
- Correlations, like volatilities, vary across time.
- Extreme returns on one market invariable relate to extreme returns in other markets.

It is this very last observation that deserves our attention in this paper; the experience of market spillover.

2 An initial work flow

We will use the following steps to analyze market spillover.

1. We build and visualize time series objects. We will also convert a time series object to a data frame for further processing. Inside of this object we will `summarise()` the data using descriptive statistics, estimate and visualize the lagged relationship of current and past realizations of the time series, and interpret autocorrelation and cross autocorrelation as indications of the styled facts of various financial markets.
2. We will derive monthly time series of correlations and volatilities from daily series.
3. We then visualize to analyze the sensitivity of entangled (as measured with correlation) markets as an endogenous system of risk.
4. We are then in the position to analyze and visualize the impact of one market's volatility on another through correlational entanglements between markets.

We will also use a probabilistic inference framework, more frequently called a Bayesian approach, to develop our provisional conclusions. What we will be able to pull out of the data is the full joint distribution of cross-market impacts as well as the marginal impacts within each market, especially as these markets communicate information among themselves.[6]

3 A working example

We suppose that we work with the CFO of an aluminum recycling company. Here is a story we can use to make practical our insights into connected markets.

The aluminum recycling company just bought a renewable energy manufacturing and services company to expand its earnings opportunities. The renewables devision dabbles in wind, clean energy technologies (very similar to the aluminum recycling clean technologies), and solar, a very new field for the company. The CFO

[4]That is, and much more colloquially, slender tails, to translate the original Greek term. But these tails extend to extreme values and thus thicken the possibility of occurring far from central locations of the data. Taleb (2018) details the use of a very different approach to including thick tailed analysis into the management of risk.

[5]McNeil, Frey, and Embrechts (2015) devote the last half of Chapter 3 to a discussion about the existence of market factors for multivariate financial time series. Their chapter 7 details three fallacies in the development of appropriate, attainable, measures of dependency, of which cross-correlation is but one.

[6]We use the generative stochastic models described as Bayesian statistical models in McElreath (2020) coded in R and Stan (Carpenter (2017); Team (2020b)) and fit using R (R Core Team) and rstan (Team (2020a)). The complete workflow will be maintained on GitHub at `https://github.com/wgfoote/market-risk`.

would like to measure the impact of this new market on the earnings of the proposed renewables division. To do this she commissions a project team.

The CFO has three questions for the team:

1. How do we characterize renewables variability?
2. Does one market affect another?
3. What stylized facts of these markets matter to the generation of earnings in this new division?

For the renewables sector we select exchange traded funds (ETF) fromm the global renewables sector: TAN for solar, ICLN for clean technologies, and PBW for wind.

These funds act as indices to effectively summarize the inputs, process, management, decisions, and outputs of various aspects of the renewables sector. Examining and analyzing these series will go a long way to helping the CFO understand the riskiness of these markets.

Our objective is to review the historical record for volatility and relationships among three repesentative markets. We load historical data on the market value of three ETFs, transform prices into returns, and then further transform the returns into within-month correlations and standard deviations.

4 Getting some data

We access daily market prices using the `tidyquant` package and transform market prices into daily returns and intra-monthly correlations and standard deviations.

```
#
options(digits = 4, scipen = 999999)
library(ggplot2)
library(GGally)
library(lubridate)
library(dplyr)
library(tidyverse)
library(quantreg)
library(forecast)
library(tidyquant)
library(timetk)
library(quantmod)
library(matrixStats)
#
symbols <- c("TAN", "ICLN", "PBW") #c("ENE", "REP", "")
price_tbl <- tq_get(symbols) %>%
  select(date, symbol, price = adjusted)
# long format ("TIDY") price tibble for possible other work
return_tbl <- price_tbl %>%
  group_by(symbol) %>%
  tq_transmute(mutate_fun = periodReturn, period = "daily", type = "log", col_rename = "daily_return") %
  mutate(abs_return = abs(daily_return))
#str(return_tbl)
r_2 <- return_tbl %>%
  select(symbol, date, daily_return) %>% spread(symbol, daily_return)
r_2 <- xts(r_2, r_2$date)[-1, ]
storage.mode(r_2) <- "numeric"
r_2 <- r_2[, -1]
r_corr <- apply.monthly(r_2, FUN = cor)[,c(2, 3, 6)]
colnames(r_corr) <- c("TAN_ICLN", "TAN_PBW", "ICLN_PBW")
r_vols <- apply.monthly(r_2, FUN = colSds)
#
corr_tbl <- r_corr %>%
  as_tibble() %>%
  mutate(date = index(r_corr)) %>%
```

```
  gather(key = assets, value = corr, -date)

vols_tbl <- r_vols %>%
  as_tibble() %>%
  mutate(date = index(r_vols)) %>%
  gather(key = assets, value = vols, -date)
#
corr_vols <- merge(r_corr, r_vols)
corr_vols_tbl <- corr_vols %>%
  as_tibble() %>%
  mutate(date = index(corr_vols))
#
```

4.1 Some simple summaries

We use tabular and graphical depictions of the shapes of each of the correlationa and volatility series. Here is the first routine to generate a tabular summary of the shape of within-month correlations.

```
corr_tbl %>% group_by(assets) %>%
  summarise(mean = mean(corr),
            sd = sd(corr), skew = skewness(corr),
            kurt = kurtosis(corr),
            min = min(corr),
            q_25 = quantile(corr, 0.25),
            q_50 = quantile(corr, 0.50),
            q_75 = quantile(corr, 0.75),
            max = max(corr),
            iqr = quantile(corr, 0.75) - quantile(corr, 0.25)
            )
```

```
## # A tibble: 3 x 11
##   assets     mean     sd   skew  kurt   min  q_25  q_50  q_75   max    iqr
##   <chr>     <dbl>  <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  <dbl>
## 1 ICLN_PBW 0.842 0.0973 -1.66   3.60  0.431 0.813 0.857 0.906     1 0.0927
## 2 TAN_ICLN 0.725 0.181  -0.855 0.369 0.126 0.598 0.779 0.863     1 0.264
## 3 TAN_PBW  0.795 0.135  -0.797 0.145 0.373 0.707 0.822 0.903     1 0.196
```

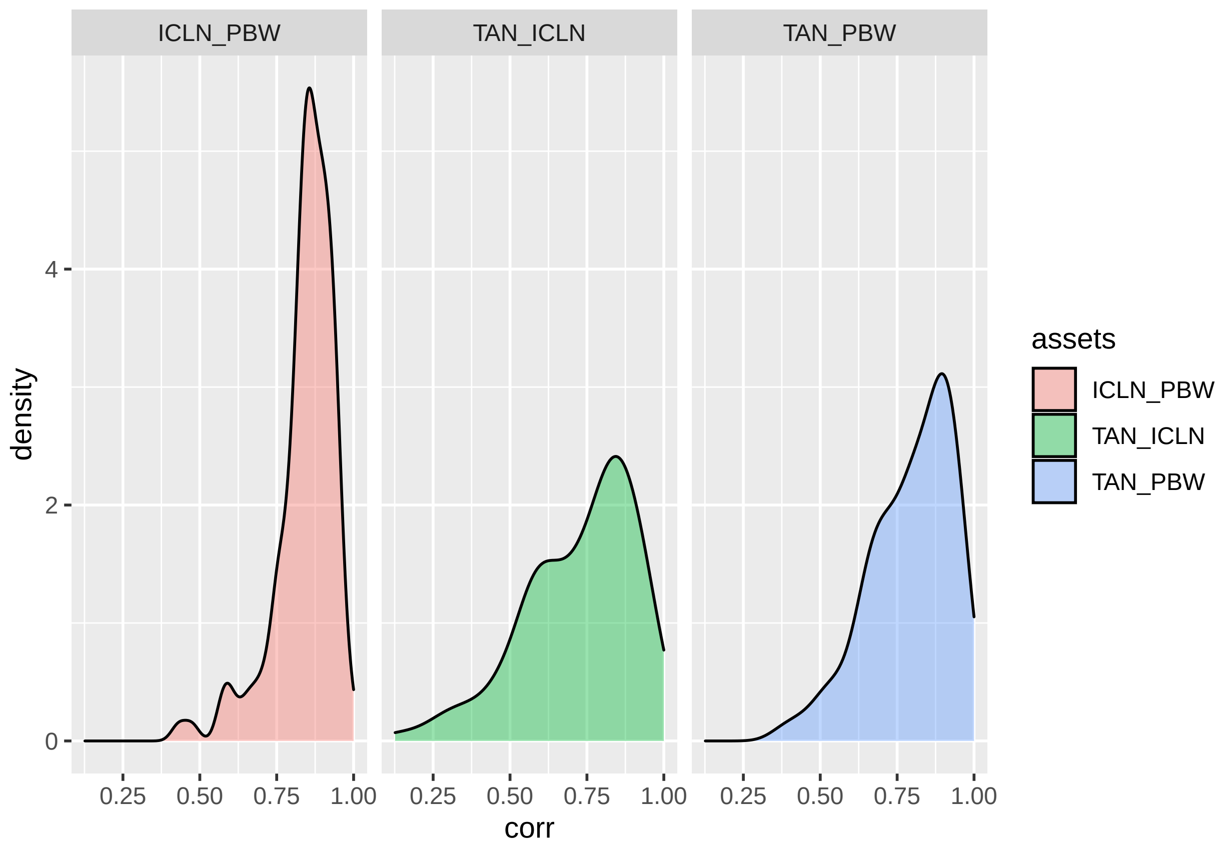
We just reuse the code above and substitute `vols` for `corr` to review the volatility data.

```
vols_tbl %>% group_by(assets) %>%
  summarise(mean = mean(vols),
            sd = sd(vols),
            skew = skewness(vols),
            kurt = kurtosis(vols),
            min = min(vols),
            q_25 = quantile(vols, 0.25),
            q_50 = quantile(vols, 0.50),
            q_75 = quantile(vols, 0.75),
            max = max(vols),
            iqr = quantile(vols, 0.75) - quantile(vols, 0.25)
            )
```

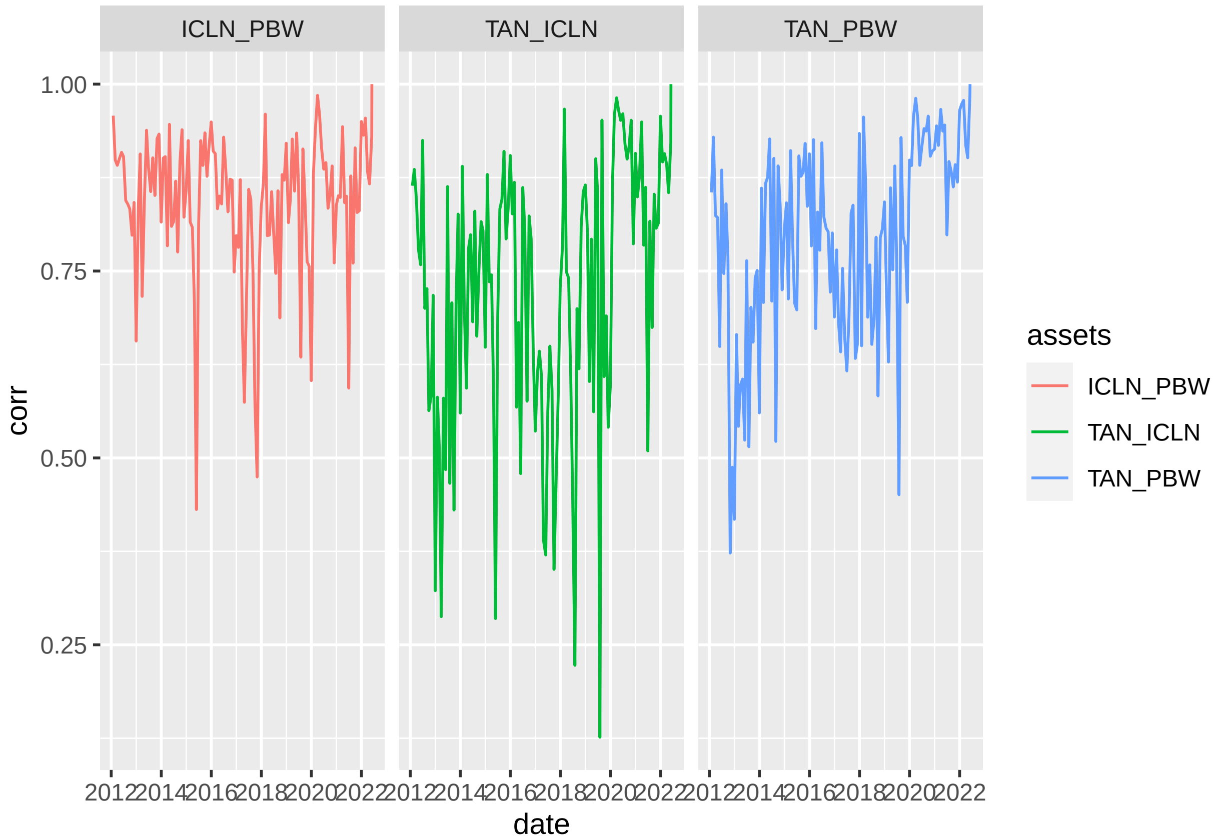
```
## # A tibble: 3 x 11
##   assets   mean      sd  skew  kurt     min   q_25   q_50   q_75    max     iqr
##   <chr>   <dbl>   <dbl> <dbl> <dbl>   <dbl>  <dbl>  <dbl>  <dbl>  <dbl>   <dbl>
## 1 ICLN   0.0146 0.00781  3.08 15.6  0.00519 0.0101 0.0128 0.0169 0.0671 0.00684
## 2 PBW    0.0187 0.0105   2.50  8.90 0.00803 0.0118 0.0162 0.0210 0.0763 0.00922
## 3 TAN    0.0227 0.00995  1.57  5.38 0.00769 0.0152 0.0216 0.0274 0.0758 0.0122
```

We view densities and line plots of historical correlations with these routines.

```
corr_tbl %>% ggplot(aes(x = corr, fill = assets)) +
  geom_density(alpha = 0.4) +
  facet_wrap(~assets)
```




```
#
corr_tbl %>% ggplot(aes(x = date, y = corr, color = assets)) +
  geom_line() +
  facet_wrap(~assets)
```

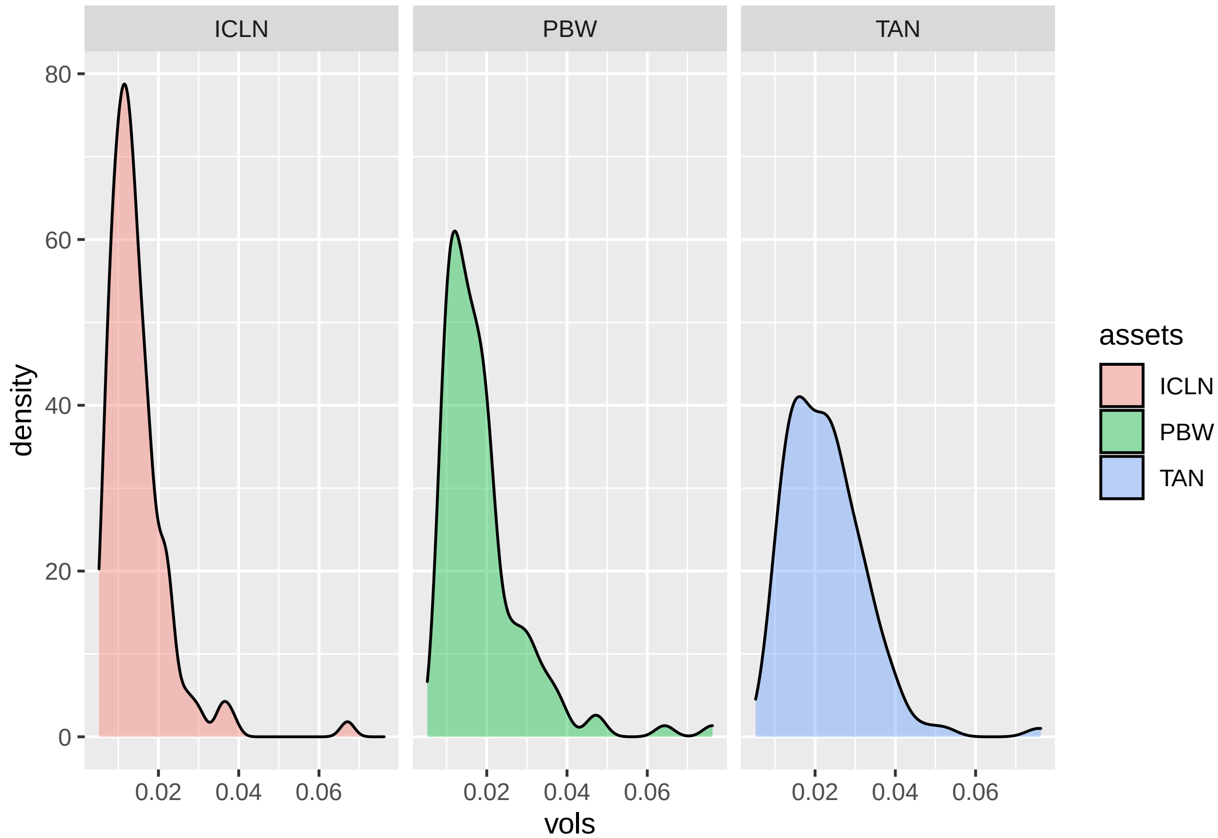



```
#
```

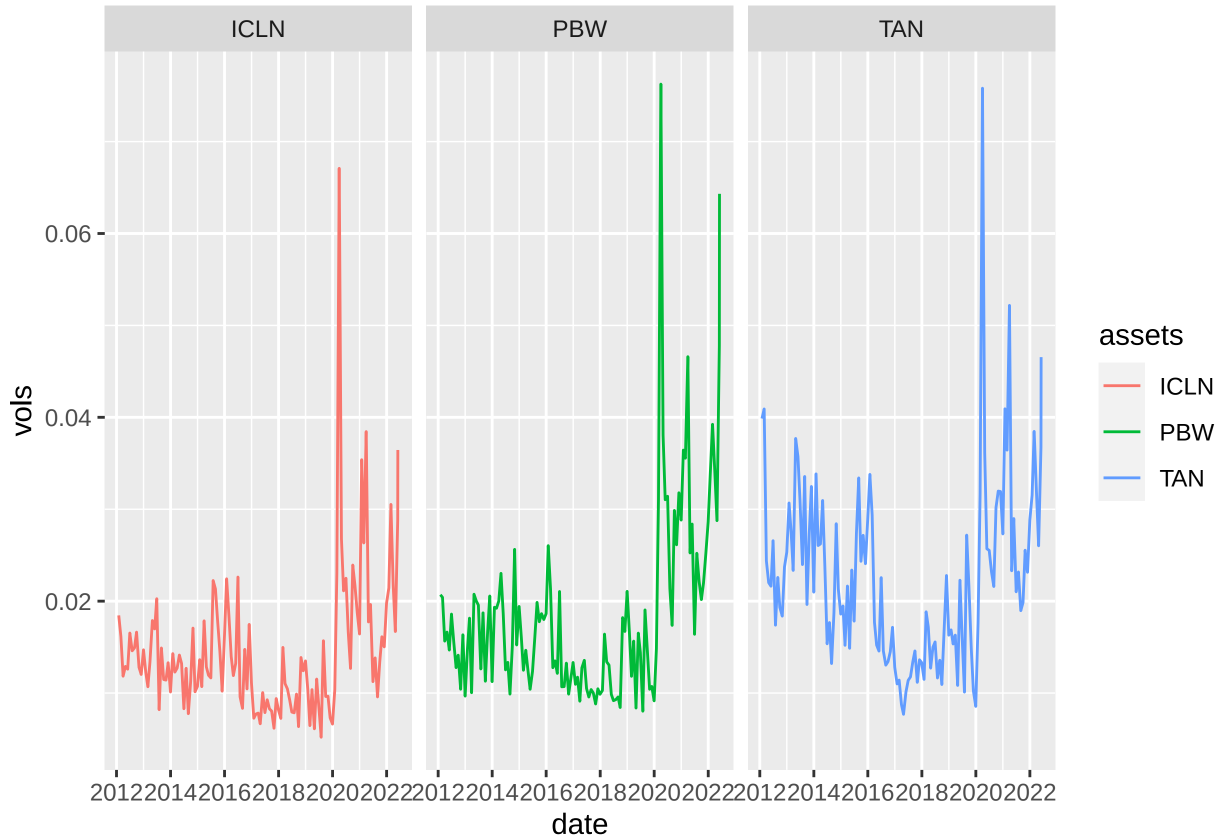
These plots not only support the summary statistics, but they also illustrate the phenomenon of volatility clustering effectively.

We use the right column of `vols_tbl`, namely, `vols`.

```
#
vols_tbl %>% ggplot(aes(x = vols, fill = assets)) +
  geom_density(alpha = 0.4) +
  facet_wrap(~assets)
```




```
#
vols_tbl %>% ggplot(aes(x = date, y = vols, color = assets)) +
  geom_line() +
  facet_wrap(~assets)
```



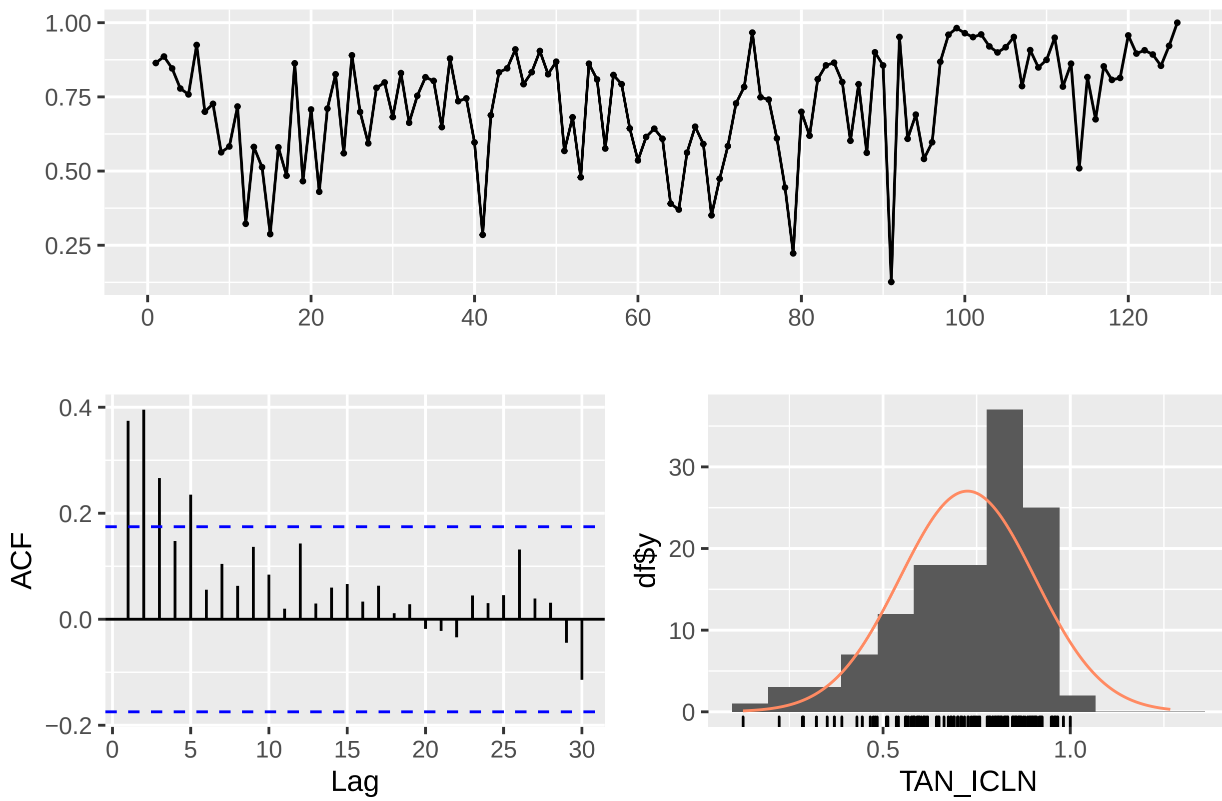

```
#
```

These initial forays into exploring the data clearly indicate the highly volatile nature both of correlation and volatility. The shape of the data shows prominent right skews and potentially thick tails as well. All of these point to the existing stylized facts of financial market returns.

4.2 Do volatility and correlation persist?

with the `TAN_ICLN` interactions and using the ggplot2 function `ggtsdisplay()` from the `forecast` package, we get all of this at one stop on the way.

```
TAN_ICLN <- r_corr$TAN_ICLN
forecast::ggtsdisplay(TAN_ICLN, lag.max=30, plot.type = "histogram")
```



The verdict on correlation persistence?

- There is some monthly persistence to 5 lags.
- The variability of correlation varies only in the positive direction.
- The distribution seems skewed to the left and non-normal.

We check the `TAN` volatilities next.

```
TAN <- r_vols$TAN
forecast::ggtsdisplay(TAN,lag.max=30, plot.type = "scatter")
```

What is the verdict on volatility persistence?

- Strong and persistent lags over 10 months shows slow decay. Is this some evidence of market memory of risk?
- Perhaps, but it also indicates in this monthly time interval influences of outliers in the third scatter panel and variability seen in the first time series panel.

And, perhaps, the term *verdict* is too strong. But we do seem some recurring patterns in line with previously identified stylizations of market data.

5 Do markets spill into one another?

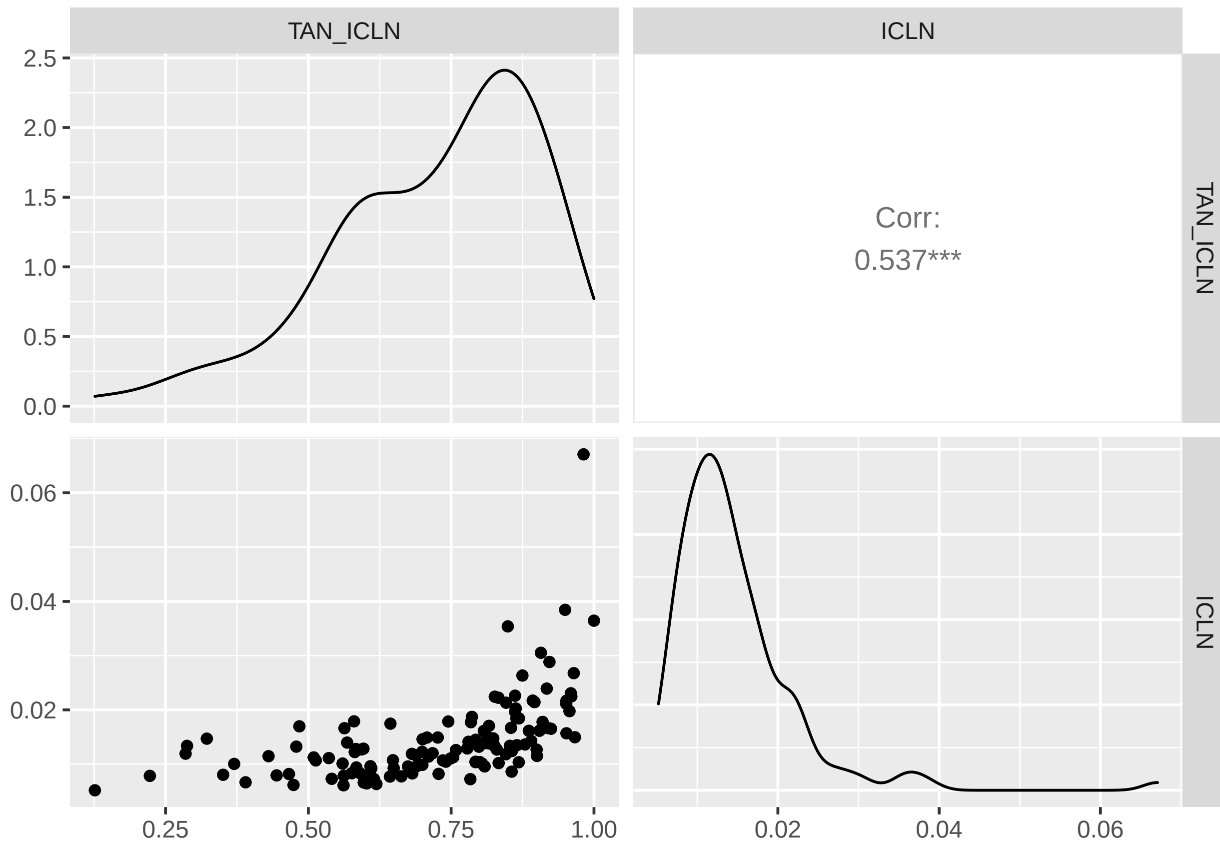
Market spillover occurs when the volatility of one market, through entanglement[7], affects the volatility of another market. We have three markets here: TAN, ICLN, and PBW all interacting with one another. We are

[7]This is a term from quantum mechanics when the state of the system (the market here) is indeterminate but nonetheless components are correlated. See David Orrell (2020) to peer into these ideas.

not asking why, just the question of whether we observe spillover. If ICLN is volatile, will TAN be affected? If so, unanticipated competitive movements in one sector (ICLN) will cause unanticipated movements in another (TAN), here coupled through correlational structures, perhaps modeled with copulae.

Let's examine this idea with a simple scatter matrix.

```
corr_vols <- merge(r_corr, r_vols)
corr_vols_tbl <- corr_vols %>% as_tibble() %>%
  mutate(date = index(corr_vols))
ggpairs(corr_vols_tbl[, c("TAN_ICLN", "ICLN")])
```



What do we observe?

1. Are they apparently normally distributed? Not at all, apparently. We observe a negative skew in correlation and the characteristically positive skew in volatility.
2. What do the outliers look like in a potential relationship between correlations and volatility? The scatter plot indicates potential outliers in very high and very low correlation market environments.
3. Are there potentially multiple regions of outliers? Yes, in very high and low correlation environments. The body of the relation appears to have a positive impact in line with a fairly high correlation of 0.513.

5.1 Quantile regression thoughts

With the existence of outliers in multiple regions we might consider a technique that respects this situation. That technique is quantile regression using the `quantreg` package. Quantile regression (Koenker (2005)) can help us measure the impact of high stress episodes on markets, modeled as high and low quantiles.[8]

[8]Here is a tutorial on quantile regression that is helpful for the formulation of models and the interpretation of results.

- Just like `lm()`for Ordinary Leasat Squares (OLS), we set up `rq()` with left-hand side (correlations) and right hand side variables (volatilities).
- We also specify the quantiles of the left-hand side to identify outliers and the median of the relationship using the `taus` vector. Each value of `tau` will run a separate regression.

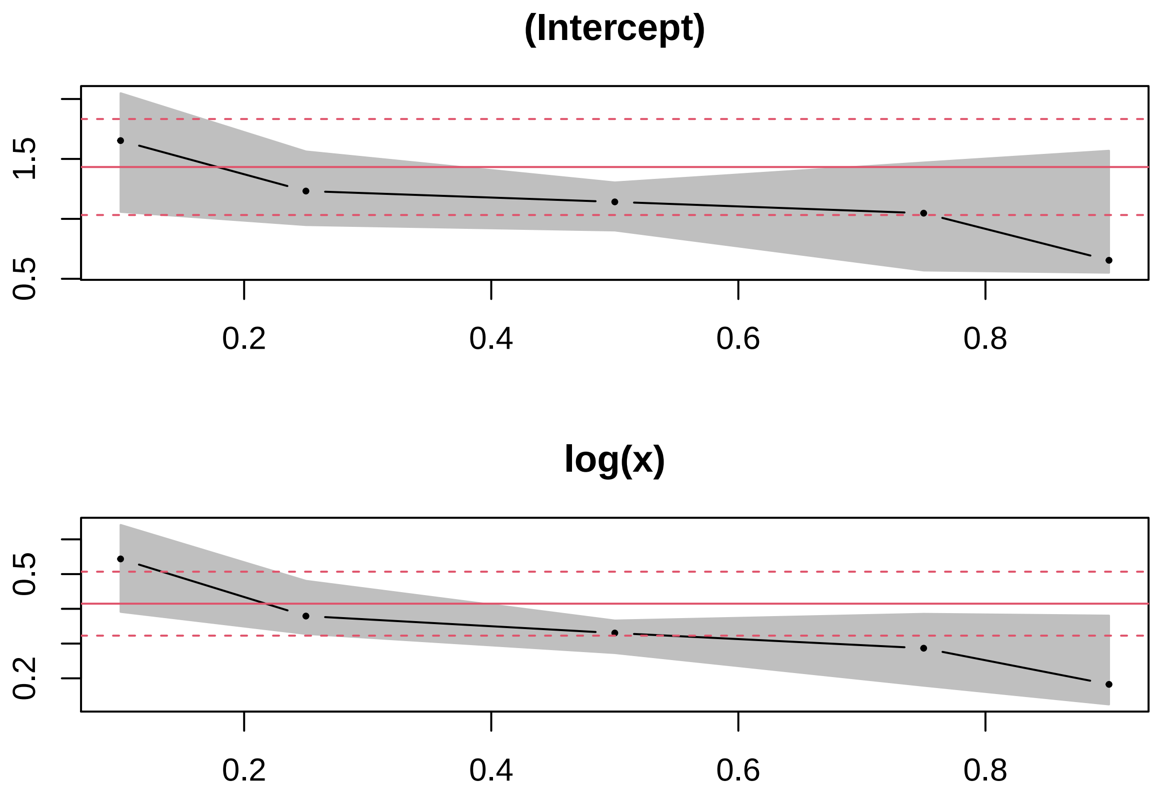
We run this code for one combination of correlations and volatilities. We can modify `y` and `x` for other combinations, and thus, other markets. A log-log transformation can help us understand the relationship between markets as the elasticity of correlation with respect to volatility.

```
library(quantreg)
taus <- c(0.10, 0.25, 0.50, 0.75, 0.90) # quantiles of y for a 95% confidence interval
y <- corr_vols_tbl$TAN_ICLN; x <- corr_vols_tbl$ICLN
fit_corr_vols <- rq(log(y) ~ log(x), tau = taus)
fit_summary <- summary(fit_corr_vols)
fit_summary
```

```
##
## Call: rq(formula = log(y) ~ log(x), tau = taus)
##
## tau: [1] 0.1
##
## Coefficients:
##             coefficients lower bd upper bd
## (Intercept) 1.6530       1.0586   2.0482
## log(x)      0.5436       0.3914   0.6414
##
## Call: rq(formula = log(y) ~ log(x), tau = taus)
##
## tau: [1] 0.25
##
## Coefficients:
##             coefficients lower bd upper bd
## (Intercept) 1.2320       0.9486   1.5622
## log(x)      0.3792       0.3280   0.4805
##
## Call: rq(formula = log(y) ~ log(x), tau = taus)
##
## tau: [1] 0.5
##
## Coefficients:
##             coefficients lower bd upper bd
## (Intercept) 1.1416       0.9033   1.3037
## log(x)      0.3303       0.2728   0.3663
##
## Call: rq(formula = log(y) ~ log(x), tau = taus)
##
## tau: [1] 0.75
##
## Coefficients:
##             coefficients lower bd upper bd
## (Intercept) 1.0475       0.5695   1.4699
## log(x)      0.2867       0.1787   0.3855
##
## Call: rq(formula = log(y) ~ log(x), tau = taus)
##
## tau: [1] 0.9
##
## Coefficients:
##             coefficients lower bd upper bd
```

```
## (Intercept) 0.6545       0.5511   1.5681
## log(x)      0.1825       0.1249   0.3804
```

```
plot(fit_summary)
```



The plot depicts the parameter estimate (intercept and slope) on the vertical axis and the quantile of correlation on the horizontal axis. The gray range is the 95% confidence interval of the parameter estimates. The dashed red lines depict the ordinary least squares regression confidence intervals.

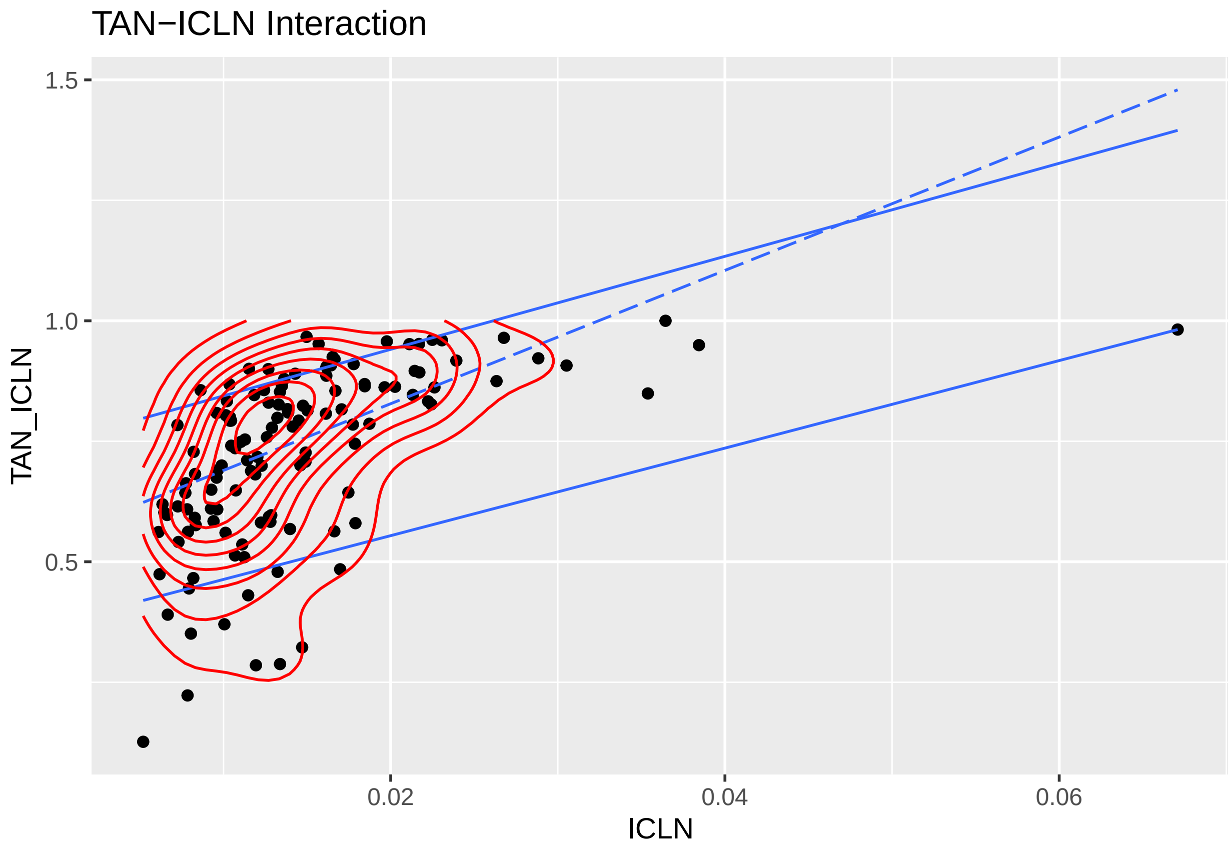
We might ask further questions of this analysis.

1. When is it likely for markets to spill over? Mostly across low to high correlation quantiles.
2. At what likelihood of correlations are market spillovers most uncertain? Again in very low and very high correlation quantile regions.
3. What about the other markets and their spillover effects?
4. What should the CFO glean from from these results?

The last two questions deserve further analysis, which means more regressions. The CFO can get an idea that preparing for market risk is a very high risk management priority.

One more plot to tie up the market spillover questions.

```
p <- ggplot(corr_vols_tbl,  aes(x = ICLN, y = TAN_ICLN)) +
    geom_point() +
    ggtitle("TAN-ICLN Interaction") +
    geom_quantile(quantiles = c(0.10, 0.90)) +
    geom_quantile(quantiles = 0.5, linetype = "longdash") +
    geom_density_2d(colour = "red")
p
```



To tailor this picture a bit, we can use `+ ylim(0.25, 1)` to specify the y-axis limits. The dashed line depicts the 50th quantile.

To what degree do our conclusions change when we perform similar analyses on the other market interactions? We just need to re-run the same script with the other market dyads.

5.2 Bayesian thoughts

Alternatively, we can examine the impact of the riskiness of one market on the other using a probabilistic model. Up to this point we have implemented a robust, perhaps in a frequentist mood, model of market interaction. This allows us to form a binary hypothesis: H_0 no spillover, and H_1 spillover. We might question the acceptance or rejection of hypotheses based on the probability emphasis of the null hypothesis. We might also observe overlap of the probability that either might occur.

This objection raises the issue of prior expectations about the hypotheses. If we assume that each is equally probable, perhaps *Beta* distributed then we could let the likelihood of each hypothesis directly impact our inference. If we were to update priors with posteriors, we might also be able to tune the inference further.

Inherently these are not complex models with only a single regressor acting on a dependent variable. Built into each variate are several assumptions about the variability and co-variability of returns. We might ask next what is the industry structure of spillover effects, at least as represented by these three segments of the renewables market.

We propose this generative model.

$$\rho_[i] \sim \text{Normal}(\mu_\rho, \sigma_\rho) \tag{1}$$
$$\mu_{\rho[i]} = \alpha_[i] + \beta_[i]\sigma_[i] \tag{2}$$
$$\alpha_[i] \sim \text{Normal}(0, 1) \tag{3}$$
$$\beta_[i] \sim \text{Normal}(0, 1) \tag{4}$$
$$\sigma_{\rho[i]} \sim \text{Exp}(1) \tag{5}$$
$$\tag{6}$$

The markets are indicated by $[i]$, with ρ and σ the within-month correlation and market volatility, μ_ρ and σ_ρ the expected value and volatility of the relationship between within-month correlation and market volatility. This formulation allows a mixed-random effects hierarchical model of the probably market risk infrastructure.

```
library(rethinking)
y <- corr_vols_tbl$TAN_ICLN; x <- corr_vols_tbl$ICLN
d_1 <- tibble(
  tan_icln = y,
  icln = x
)
m_1 <- quap(
  alist(
    tan_icln ~ dnorm( mu, sigma ),
    mu <- a + b*icln,
    a ~ dnorm( 0, 1 ),
    b ~ dnorm( 0, 1 ),
    sigma ~ dexp( 1 )
  ),
  data = d_1
)
precis( m_1 )
```

```
##         mean      sd   5.5%  94.5%
## a     0.6873 0.02029 0.6549 0.7197
## b     2.6048 0.92907 1.1199 4.0896
## sigma 0.1697 0.01116 0.1519 0.1875
```

Priors are proper to the generation of the dependent variable. They can also take on positive or negative signs. Spillover is much in evidence here through size of impact and direction. Let's repeat this for the TAN-PBW and PBW-ICLN markets separately.

```
#library(rethinking)
y <- corr_vols_tbl$TAN_PBW; x <- corr_vols_tbl$PBW
d_2 <- tibble(
  tan_pbw = y,
  pbw = x
)
m_2 <- quap(
  alist(
    tan_pbw ~ dnorm( mu, sigma ),
    mu <- a + b*pbw,
    a ~ dnorm( 0, 1 ),
    b ~ dnorm( 0, 1 ),
    sigma ~ dexp( 1 )
  ),
  data = d_2
)
precis( m_2 )
```

```
##          mean       sd   5.5%  94.5%
## a     0.7247 0.017371 0.6970 0.7525
## b     3.7469 0.747278 2.5526 4.9412
## sigma 0.1163 0.007758 0.1039 0.1287
```

A comparatively stronger influence is felt in this market interaction. Now for the last pair, ICLN_PBW.

```
#library(rethinking)
y <- corr_vols_tbl$ICLN_PBW; x <- corr_vols_tbl$PBW
d_3 <- tibble(
  icln_pbw = y,
  pbw = x
)
m_3 <- quap(
  alist(
    icln_pbw ~ dnorm( mu, sigma ),
    mu <- a + b*pbw,
    a ~ dnorm( 0, 1 ),
    b ~ dnorm( 0, 1 ),
    sigma ~ dexp( 1 )
  ),
  data = d_3
)
precis( m_3 )
```

```
##           mean       sd    5.5%   94.5%
## a     0.78796 0.013735 0.76600 0.80991
## b     2.88780 0.609095 1.91435 3.86125
## sigma 0.08656 0.005578 0.07765 0.09547
```

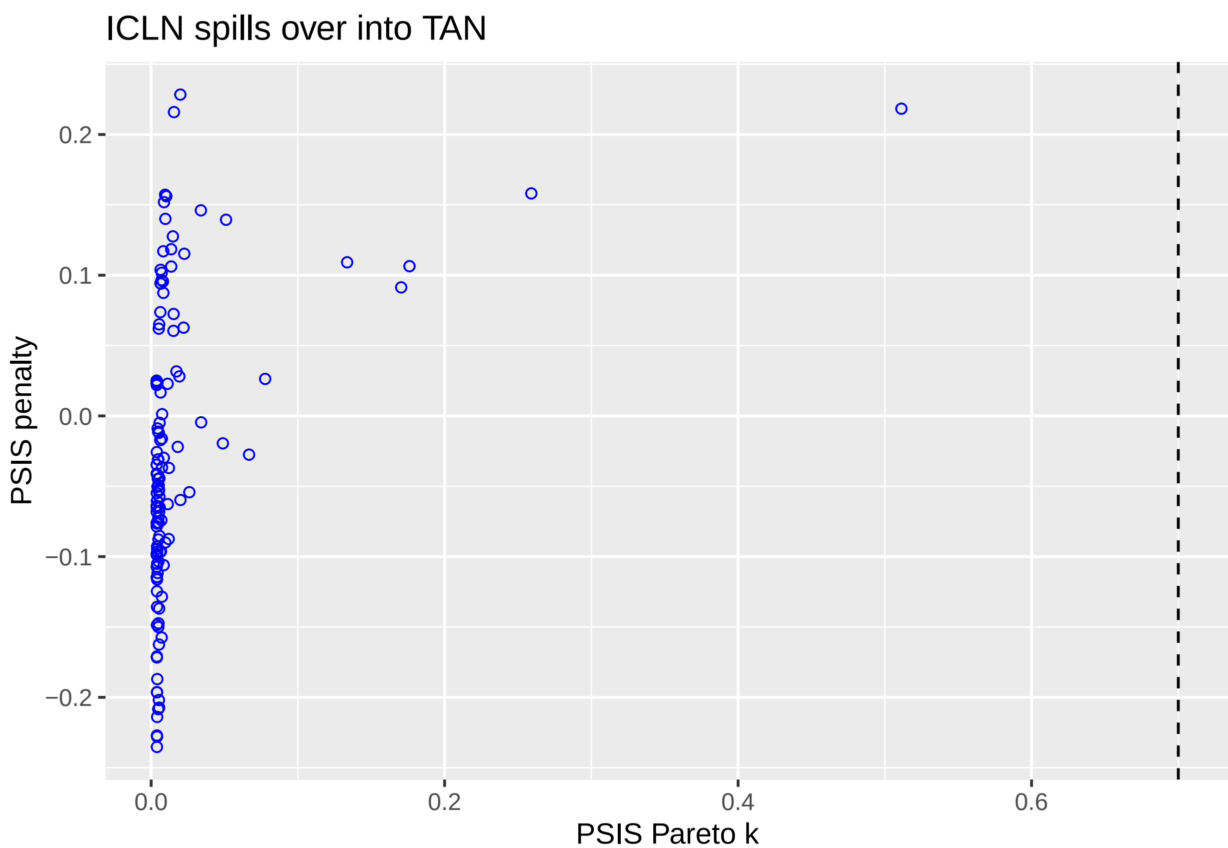
Wind and solar seem to have similar spilloever characteristics relative to clean technologies. All of the models have similarities to the quantile regressions.

Pareto-Smoothed Importance Sampling and cross validation with the Leave One Out approach yields a trade-off between spillover variability and bias on the y-axis and uncertainty on the x-axis. Thick-tailed, skewed returns distributions become intelligible with this analysis. The extreme uncertainty of outliers (known-unknowns from $k = 0$ to 0.7; unknown-unknowns for $k > 0.7$ in tests reported by Gelman and Vehtari (2013) and Vehtari and Gabry (2015)) contribute to the ability of the ICLN market to spill its uncertainty into the high variations of the TAN market all through the naive mechanism of correlation.[9]

```
## R code 7.34 McElreath2020
#library( plotly )
options( digits=2, scipen=999999)
d <- d_1
set.seed(4284)
m <- m_1
PSIS_m <- PSIS( m, pointwise=TRUE )
PSIS_m <- cbind( PSIS_m, tan_icln=d$tan_icln, icln=d$icln )
set.seed(4284)
#WAIC_m2.2 <- WAIC(m2.2,pointwise=TRUE)
p1 <- PSIS_m %>%
  ggplot( aes( x=penalty, y=k ) ) +
  geom_point( shape=21, color = "blue" ) +
```

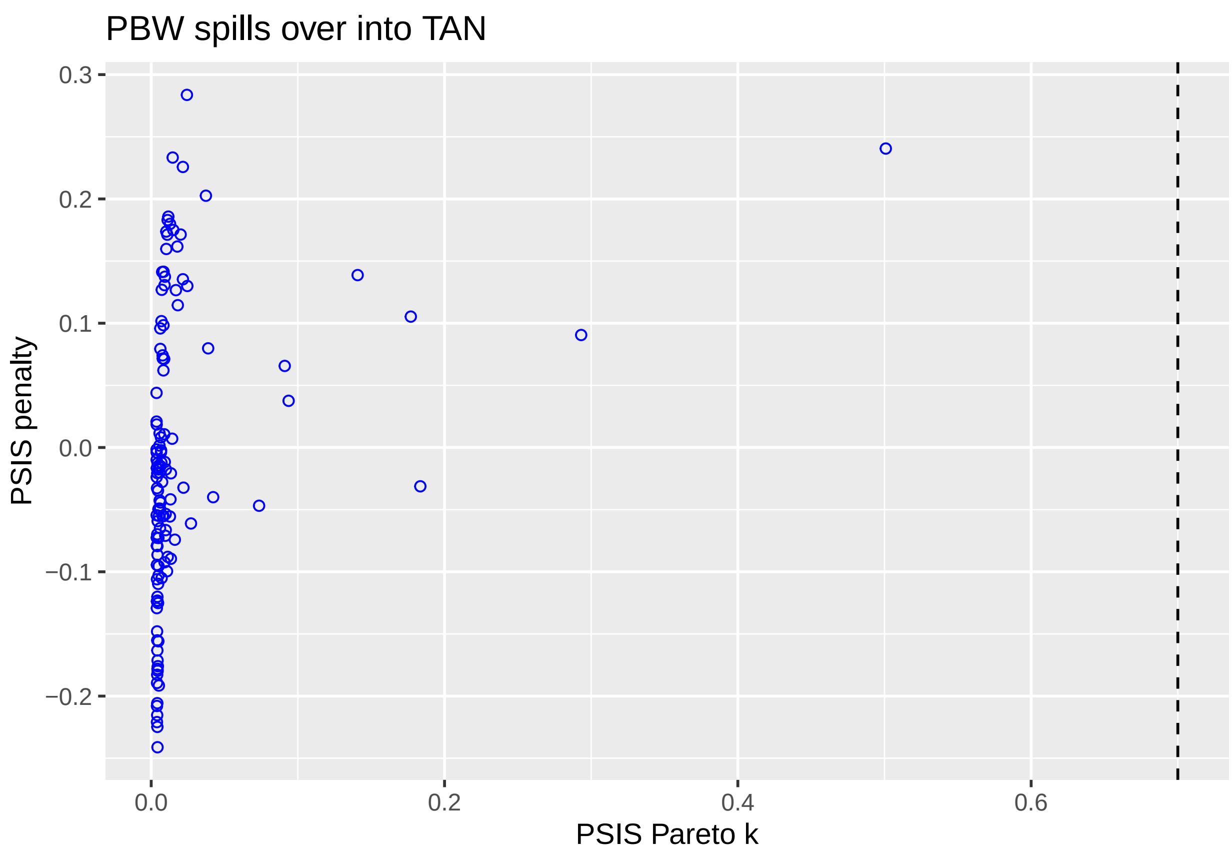
[9]Power law distributions notoriously do not possess first, second, third, or even fourth moments analytically across the GPD parameter space, especially for k (see Embrechts (2000) for examples). Watanabe (2009) develops a theory of statistical learning through which singularities in the space of estimation parameters imply that standard inference using Central Limit Theorems, Gaussian distributions, are inadmissable. The existence of divergence, at least as made intelligible through singularity theory, means we should rely on the more robust median, mean absolute deviation, and inter-quartile ranges (probability intervals) to summarize the outcomes of power law distributions. All of this also aligns well with Taleb's many warnings, examples, and inferences by Taleb (2018).

```
  xlab("PSIS Pareto k") + ylab("PSIS penalty") +
  geom_vline( xintercept = 0.7, linetype = "dashed") +
  ggtitle( "ICLN spills over into TAN" )
p1 #ggplotly( p1 )
```



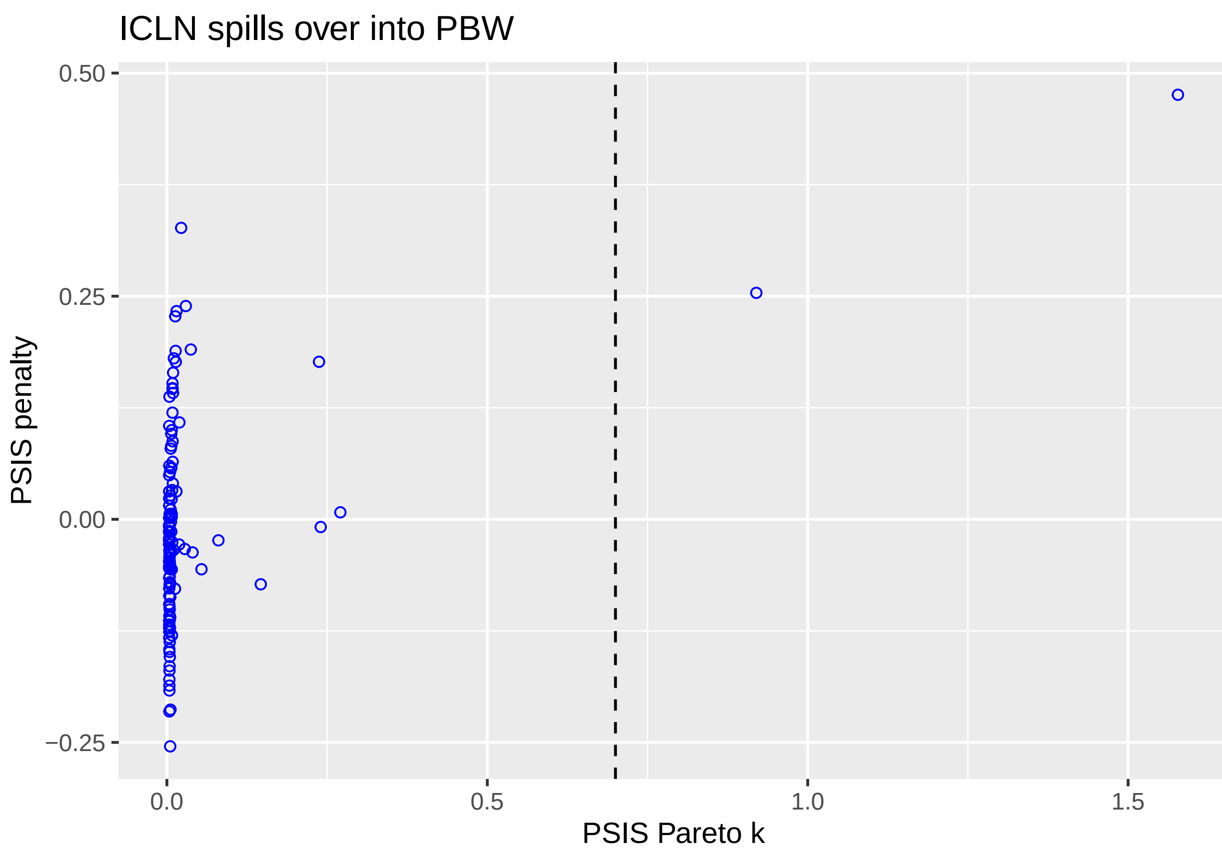
Here are the TAN-PBW outlier results.

```
## R code 7.34 McElreath2020
library( plotly )
options( digits=2, scipen=999999)
d <- d_2
set.seed(4284)
m <- m_2
PSIS_m <- PSIS( m, pointwise=TRUE )
PSIS_m <- cbind( PSIS_m, tan_pbw=d$tan_pbw, icln=d$pbw )
set.seed(4284)
#WAIC_m2.2 <- WAIC(m2.2,pointwise=TRUE)
p1 <- PSIS_m %>%
  ggplot( aes( x=penalty, y=k ) ) +
  geom_point( shape=21, color = "blue" ) +
  xlab("PSIS Pareto k") + ylab("PSIS penalty") +
  geom_vline( xintercept = 0.7, linetype = "dashed") +
  ggtitle( "PBW spills over into TAN" )
p1 #ggplotly( p1 )
```



and

```
## R code 7.34 McElreath2020
#library( plotly )
options( digits=2, scipen=999999)
d <- d_3
set.seed(4284)
m <- m_3
PSIS_m <- PSIS( m, pointwise=TRUE )
PSIS_m <- cbind( PSIS_m, icln_pbw=d$icln_pbw, pbw=d$pbw )
set.seed(4284)
#WAIC_m2.2 <- WAIC(m2.2,pointwise=TRUE)
p1 <- PSIS_m %>%
  ggplot( aes( x=penalty, y=k ) ) +
  geom_point( shape=21, color = "blue" ) +
  xlab("PSIS Pareto k") + ylab("PSIS penalty") +
  geom_vline( xintercept = 0.7, linetype = "dashed") +
  ggtitle( "ICLN spills over into PBW" )
p1 #ggplotly( p1 )
```



Our next stop is to look at the joint probability of spillover effects across the three markets.

6 Industry risk infrastructure

We now invoke the full generative model, including the jointly considered markets. We index each market and consider all of the impact parameters as jointly determined. They thus share information across markets through the total probability of observing all three markets in the presence of all of the market interaction parameters.

```
#library(rethinking)
corr_vol_spill <- read_csv( "market-spillovers.csv" )
d_4 <- corr_vol_spill %>%
  tibble(
    corr = log( abs(corrs) ),
    vol = log( vols ),
    mid = mids
)
#log vols and (abs) corrs will allow us to interpret coefficients as elasticities
```

```
m_4 <- quap(
  alist(
    corr ~ dnorm( mu, sigma ),
    mu <- a[mid] + b[mid] *vol,
    a[mid] ~ dnorm( 0, 1 ),
    b[mid] ~ dnorm( 0, 1 ),
    sigma[mid] ~ dexp( 1 )
  ),
  data = d_4
```

```
)
precis( m_4, depth = 2 )
```

```
##           mean    sd  5.5% 94.5%
## a[1]      1.36 0.183 1.066  1.65
## a[2]      0.81 0.165 0.544  1.07
## a[3]      0.48 0.168 0.216  0.75
## b[1]      0.40 0.042 0.330  0.46
## b[2]      0.26 0.040 0.193  0.32
## b[3]      0.16 0.040 0.097  0.23
## sigma[1] 0.21 0.013 0.186  0.23
## sigma[2] 0.17 0.011 0.152  0.19
## sigma[3] 0.19 0.013 0.174  0.21
```

```
#options( digits = 2 )
#cov2cor( vcov( m_4 ) )
```

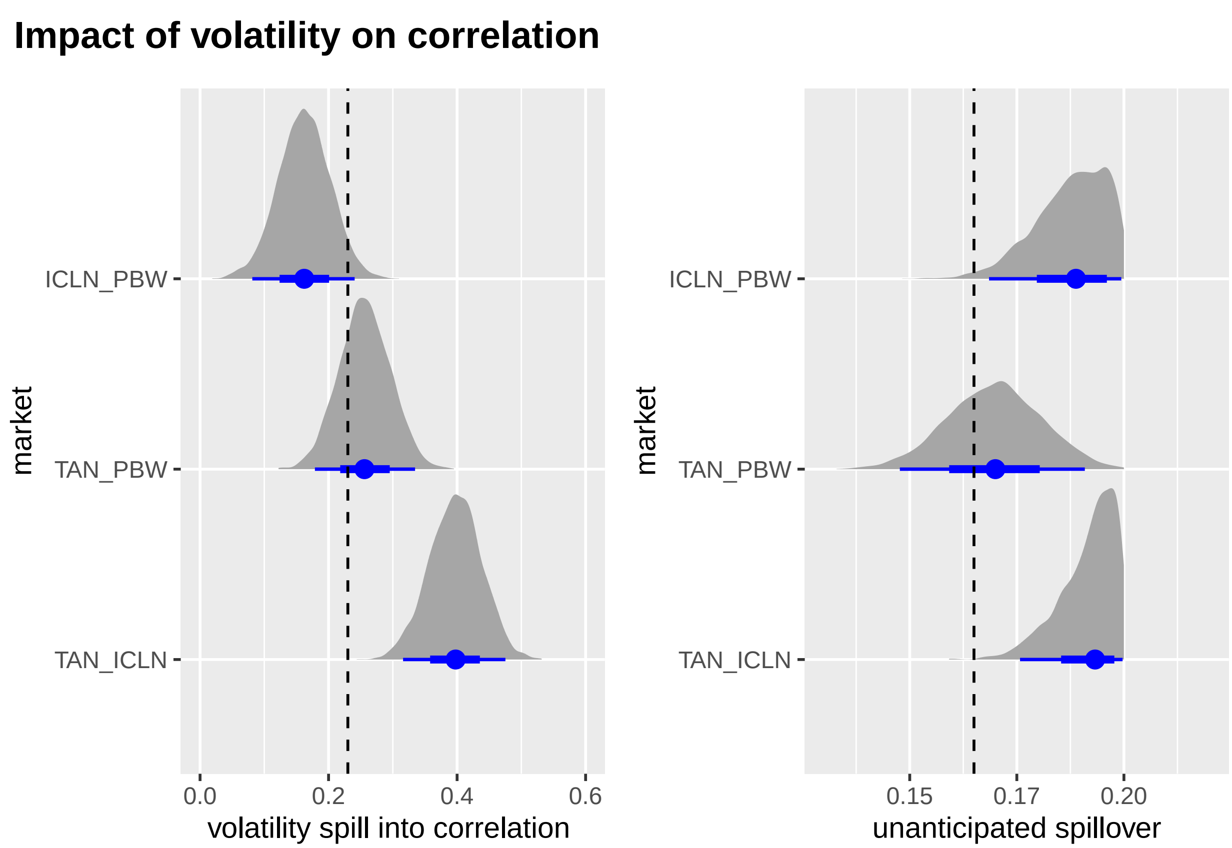
The Bayesian approach integrates the marginal parameters for each of the markets by using the total probability of observing the data across the markets and the range of potential parameter values. Thus the parameters share information across markets for systematic and ideosyncratic components.

This helper function based on the `cowplot` package will place two plots side-by-side, much like a `ggplot2` facet.

```
#
# draw marginal samples for parameters
#
# ggplot helper
library(cowplot)
plot_2_grid <- function( plot1, plot2, title_1= "Default" ){
    # build side by side plots
    plot_row <- plot_grid(plot1, plot2)
    # now add the title
    title <- ggdraw() +
    draw_label(
      title_1,
      fontface = 'bold',
      x = 0,
      hjust = 0
    ) +
    theme(
      # add margin on the left of the drawing canvas,
      # so title is aligned with left edge of first plot
      plot.margin = margin(0, 0, 0, 7)
    )
    plot_grid(
    title, plot_row,
    ncol = 1,
    # rel_heights values control vertical title margins
    rel_heights = c(0.1, 1)
    )
}
```

The comparison of slope impact parameters across the three markets in the left panel indicates the high plausibility of information sharing across the markets on a systematic basis. In the right panel are the standard deviations, a measure of the shared information of an unsystematic nature. The joint distribution of these marginal parameters indicate a high degree of sharing of return volatility among the markets.

```
library(tidybayes.rethinking)
m_draws <- m_4 %>%
  spread_draws( a[mid], b[mid], sigma[mid] )
mid <- as.factor(m_draws$mid)
levels(mid) <- list( "TAN_ICLN"=1, "TAN_PBW"=2, "ICLN_PBW"=3 )
m_draws$mid <- mid
# plot grid of two parameters
p1 <- m_draws %>%
  ggplot(aes(x = b, y = as.factor(mid) ) ) +
  stat_halfeye( color = "blue") +
  xlab("volatility spill into correlation ") + ylab("market") +
  xlim( 0.0, 0.6) +
  geom_vline( xintercept = 0.23 , linetype = "dashed") + geom_vline( xintercept = 4.00, linetype = "dasl
# build sigma_V comparison
p2 <- m_draws %>%
  filter( sigma < 0.2) %>%
  ggplot(aes(x = sigma, y = as.factor(mid) ) ) +
  stat_halfeye( color = "blue") +
  xlab("unanticipated spillover") + ylab("market") +
  xlim( 0.13, 0.22) +
  geom_vline( xintercept = 0.12 , linetype = "dashed") + geom_vline( xintercept = 0.165, linetype = "da:
plot_2_grid( p1, p2, title_1 = "Impact of volatility on correlation")
```



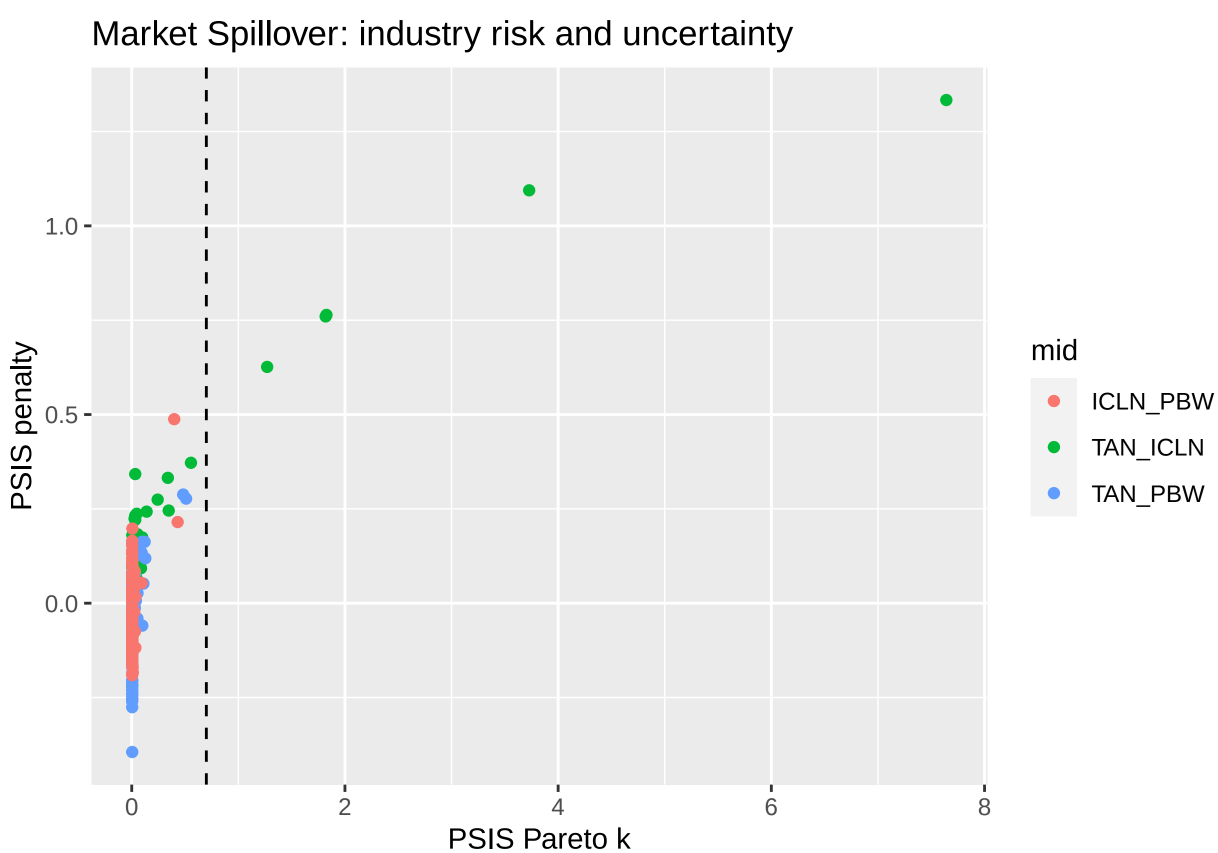

```
#
ggsave( "compare.png" )
```

To ground us in a more practical answer. We can say that a 10% move either in ICLN or PBW volatility will induce at least a 30% in the volatility of TAN, with moves as low as about 20% and as high as about 45%. A

10% move in ICLN will induce a bit more than an 18% move in PBW, with moves as low as 6% and as high as over 30%. These intervals are credible with 89% probability.

Again we can review the role of each of the observations on the bias-uncertainty tradeoff. High penalty-high k observations will obscure efforts to predict correlations. The market to watch in this regard is the ICLN-PBW pair, which coincides with the high volatilities both of PBW and ICLN evident in the σ distributions we viewed above.

```
## R code 7.34 McElreath2020
#library( plotly )
options( digits=2, scipen=999999)
d <- d_4
set.seed(4284)
m <- m_4
PSIS_m <- PSIS( m, pointwise=TRUE )
PSIS_m <- cbind( PSIS_m, mid = d$markets )
set.seed(4284)
#WAIC_m2.2 <- WAIC(m2.2,pointwise=TRUE)
p1 <- PSIS_m %>%
  ggplot( aes( x=penalty, y=k, fill = mid, color = mid ) ) +
  geom_point( shape=21 ) +
  xlab("PSIS Pareto k") + ylab("PSIS penalty") +
  geom_vline( xintercept = 0.7, linetype = "dashed") +
  ggtitle( "Market Spillover: industry risk and uncertainty" )
p1 #ggplotly( p1 )
```



Most of the uncertainty is found in wind (PBW) volatility on solar (TAN: green). The least uncertain market seems to be the relationship between PBW and ICLN.

Finally, at least for this exercise, we can compare the Wide Area Information Criterion, also known as the Watanabe-Akaike Information Criterion, or WAIC for short, of covariance and non-covariance dependent models of market structure.[10]

7 What have we accomplished?

We can provide provisional answers to the CFO's initial questions using the work flows developed here.

- Univariate volatility clustering is compatible with highly kurtotis, and thus highly volatile volatility.
- Volatile volatility in each asset spills into each market separately. These affects are measured using both quantile regression and Bayesian statistical techniques. The two approaches seem to agree on the probable existence of market spillover.
- We tested the hypothesis that one market's riskiness affects another market, probably. They do, both in systematic, β, and idiosyncratic, σ, modes.
- We find that a strongly coupled market risk structure persists across TAN, PBW and ICLN assets in each pair of market combinations.
- TAN is the most sensitive market with highly volatile responses to ICLN and PBS volatility. Both ICLN and PBW are far less sensitive to moves within and across their markets.

The CFO should be wary that strong negative movements in one market will probably persist and spill over into negative movements across the three markets. Earnings based on this model will themselves exhibit the results of volatility clustering and market spillover.

Our next step could well be to build volatility clustering, and other thick-tailed outcomes, to derive implied capital requirements through market risk channels. Such requirements would inform the risk budgeting assignments that might be built into delegations of authority, portfolio limits, even into optimiation constraints.

References

Carpenter, Gelman, B. 2017. "Stan: A Probabilistic Programming Language." *Journal of Statistical Software* 76 (1).

Cont, R. 2001. "Empirical Properties of Asset Returns: Stylized Facts and Statistical Issues." *Quantitative Finance* 1.

Embrechts, Haan, Paul. 2000. "Modellinf Multivariate Extremes." In *Extremes and Integrated Risk Management*, edited by Paul Embrechts, 59–67. RISK Publications.

Gelman, J. Hwang, A., and A. Vehtari. 2013. "Understanding Predictive Information Criteria for Bayesian Models."

Koenker, Roger. 2005. *Quantile Regression (Econometric Society Monographs)*. Cambridge University Press.

McElreath, Richard. 2020. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. Second edition. CRC Press. `https://xcelab.net/rm/statistical-rethinking/`.

McNeil, A. J., R. Frey, and P. Embrechts. 2015. *Quantitative Risk Management: Concepts, Techniques and Tools - Revised Edition*. Princeton Series in Finance. Princeton University Press. `https://books.google.com/books?id=SfJnBgAAQBAJ`.

Orrell, D. 2020. *Quantum Economics and Finance: An Applied Mathematics Introduction*. Panda Ohana Publishing. `https://books.google.com/books?id=OZaMzQEACAAJ`.

[10]PSIS exploits the distribution of potentially outlying and influential observations using the GPD to model and measure the data point-wise with the shape parameter $k = \xi$. Any point with $k > 0.5$ will have infinite variance and thus contribute to a concentration of points – the thick tail. Related to this idea, WAIC is the log-posterior-predictive density (*lppd*, that is, the Bayesian deviance) and a penalty proportional to the variance in posterior predictions:

$$WAIC(y, \Theta) = -2(lppd - \underbrace{\Sigma_i var_\theta \, log \; p(y_i|\theta))}_{penalty}$$

The penalty is related to the number of free parameters in the simulations, as Watanabe (2009) demonstrates.

Taleb, Nassim Nicholas. 2018. "The Statistical Consequences of Fat Tails: Papers and Commentary." Edited by Stem Academic Publishing. *The Technical Inconcerto* 1.

Team, Stan Development. 2020a. "RStan: The r Interface to Stan (r Package Version 2.17.3) [Computer Software]." http://mc-stan.org/.

———. 2020b. "The Stan Core Library (Version 2.18.0)." http://mc-stan.org/.

Vehtari, A. Gelman, A., and J. Gabry. 2015. "Efficient Implementation of Leave-One-Out Cross-Validation and WAIC for Evaluating Fitted Bayesian Models."

Watanabe, S. 2009. *Algebraic Geometry and Statistical Learning Theory*. Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press. https://books.google.com/books?id=GnnUHasAtUQC.