Volatility and Spillover: Facts of Market Life

by William G. Foote and Brian V. Wholey

Abstract Volatility and the interaction of markets continue to beguile traders, managers, investors, policy makers and regulators, 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 both uncertainty and variability of market events, especially so-called outliers.

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.[1] [2]

The experience of market price risk management is that spillover and other market return characteristics 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.
- 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] 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]

[1]Thanks to numerous colleagues and financial engineering students at Manhattan College and Syracuse University, to Gil Bassett, the organizing committee and the participants of the R in Finance Conference, Chicago IL, June 3, 2022 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]Jan Tinbergen (Tinbergen, 1959) "As a matter of course every attempt to generalize or stylize in a venture. The artistry in the work of the social economist lies in this stylizing. Some attempts have been made which could not be handled, some proved to be unrealistic. We have to steer clear of these rocks. Stylizing is necessary however and the alternative is sterility."(p. 41) Nicholas Kaldor (Kaldor, 1961) later coined the term "stylized facts." Rama Cont (Cont, 2001) is generally recognized as producing the first cross asset survey of the so-called stylized facts of financial markets. Alexander McNeil et al. (McNeil et al., 2015) devote Chapter 3 to a discussion about the existence of market factors for univariate and multivariate financial time series.

[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. N.N. Taleb (Taleb, 2018) details the use of a very different approach to including thick tailed analysis into the management of risk.

[5]Alexander McNeil et al. (McNeil et al., 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.

- 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 in one market invariablely 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. We should observe observations that exhibit both high levels of unsystematic variability and uncertainty. Some interpreters might attenuate the influenc of these observations as rare outliers.

An initial work flow

We will use the following steps to analyze our focus on 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]

A working example

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

The aluminium recycling company just acquired a renewable energy manufacturing and services company to expand its earnings opportunities. The renewables division invests in wind, clean energy technologies (very similar to the aluminium recycling clean technologies), and solar, a very new field for the company. The CFO 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 and uncertainty?
2. Does one market affect another?
3. What stylized facts of these markets might apply 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.

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

Getting into the 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")
price_tbl <- tq_get(symbols, get  = "stock.prices",
                       from = "2012-01-31",
                       to   = "2022-02-25") %>%
  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))
#
corr_vols <- merge(r_corr, r_vols)
corr_vols_tbl <- corr_vols %>% as_tibble() %>%
  mutate(date = index(corr_vols))
#
save(corr_tbl, file="corr_tbl")
save(vols_tbl, file="vols_tbl")
write_csv(corr_vols_tbl, "corr_vols_tbl.csv")
```

The wrinkle in this wrangling occurs with the use of `storage_mode()` to reinvested the returns with a numeric attribute. Two save options are deployed: a pure R data `save()` for the correlations and volatilities separately, and `write_csv()`. These will allow us to place frozen data into a vault for further review and reference during the decision making cycle. The gather() function still operates without much ado, while the pivot_wide() could as easily been deployed as well, but with some minor intermediate requirements for continuity of data placement in columns.

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 in which we generate a tabular summary of the shape of within-month correlations.

```
load( file="corr_tbl" )
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 0.838 0.0973 -1.67  3.55  0.431 0.810 0.852 0.903 0.985 0.0934
#> 2 TAN_ICLN 0.718 0.180  -0.825 0.327 0.126 0.596 0.754 0.856 0.982 0.260
#> 3 TAN_PBW  0.789 0.134  -0.782 0.111 0.373 0.701 0.807 0.898 0.981 0.197
```

We reuse the code and substitute vols for corr to summarize the volatility data.

```
load( file="vols_tbl" )
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   0.0142 0.00752  3.47 19.6  0.00519 0.00989 0.0127 0.0165 0.0671 0.00664
#> 2 PBW    0.0178 0.00929  2.71 12.1  0.00803 0.0116  0.0157 0.0204 0.0763 0.00879
#> 3 TAN    0.0221 0.00963  1.77  6.90 0.00769 0.0150  0.0215 0.0273 0.0758 0.0123
```

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

```
corr_tbl %>% ggplot(aes(x = date, y = corr, color = assets)) +
  geom_line() +
  theme( legend.position="none") +
  facet_wrap(~assets)

#
```

These plots not only support the summary statistics, but they also illustrate the phenomenon of volatility clustering effectively with clumping of observations in distribution tails, changes in trends, peaks and valley across time (albeit with eye-ball econometrics).

We use the vols column of vols_tbl.

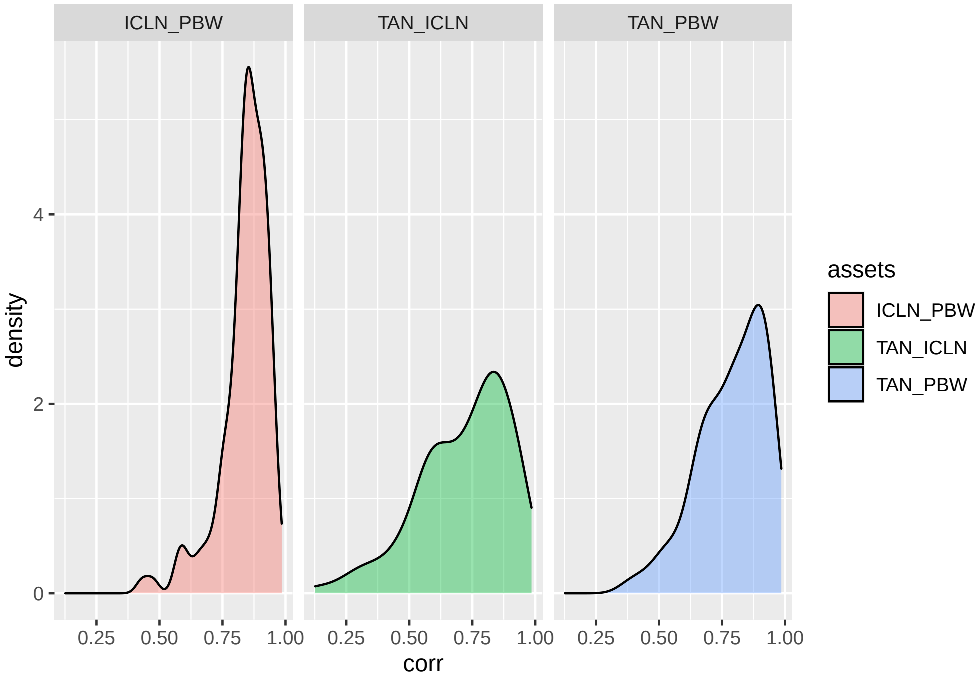

Figure 1: Correlation roams the frequency domain

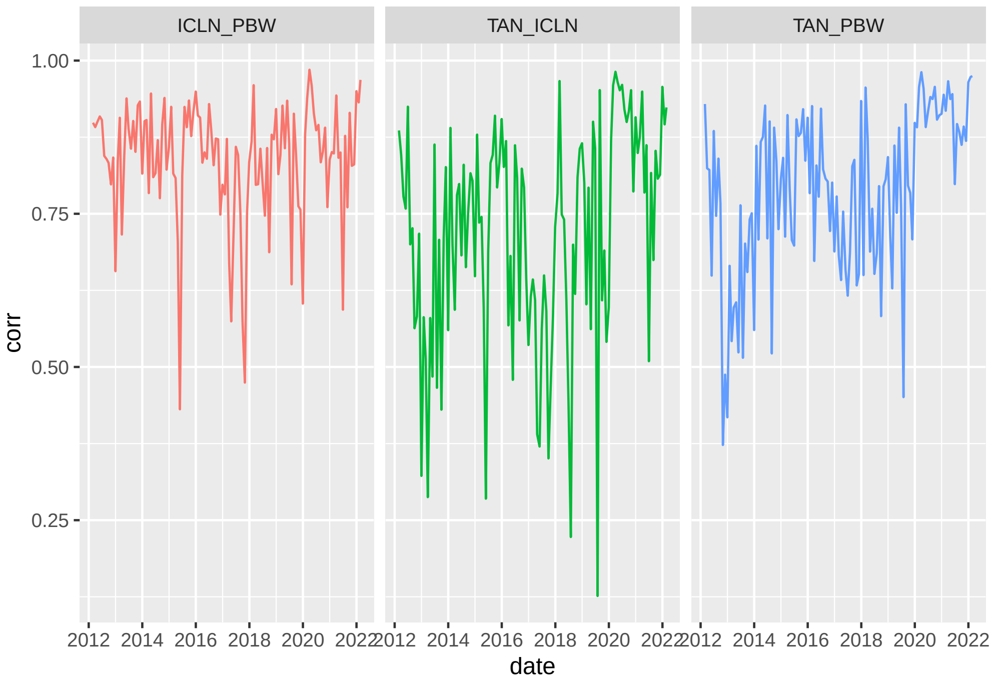

Figure 2: Correlation moves across the frequency domain

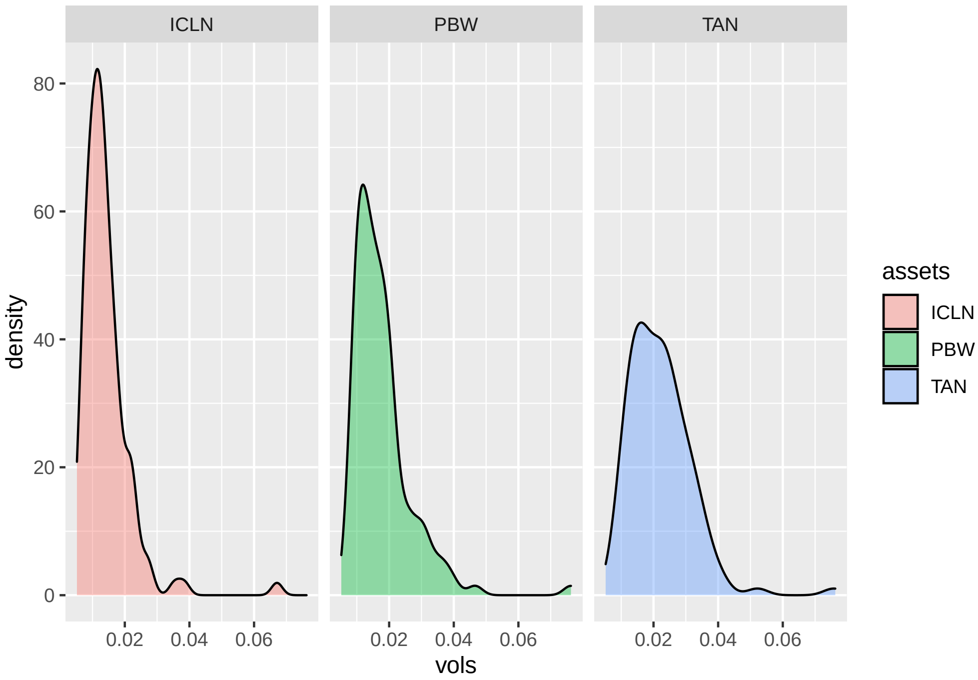

Figure 3: Volatility roams the frequency domain

```
#
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 seem to 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. The footprint of uncertainty can be seen in the extreme tails of both sets of distributions. All of these point to the existing stylized facts of univariate financial market returns.

Do volatility and correlation persist?

with the TAN-ICLN interactions and using the ggplot2 function `ggtsdisplay()` from the `forecast` package, we get all of these visualizations at one stop along the way.

What might we say about correlation persistence?

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

We visualize the TAN volatilities next.

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

What might we judge about volatility persistence?

- Strong and persistent lags over 10 months shows slow decay. Is this some evidence of market memory of risk?

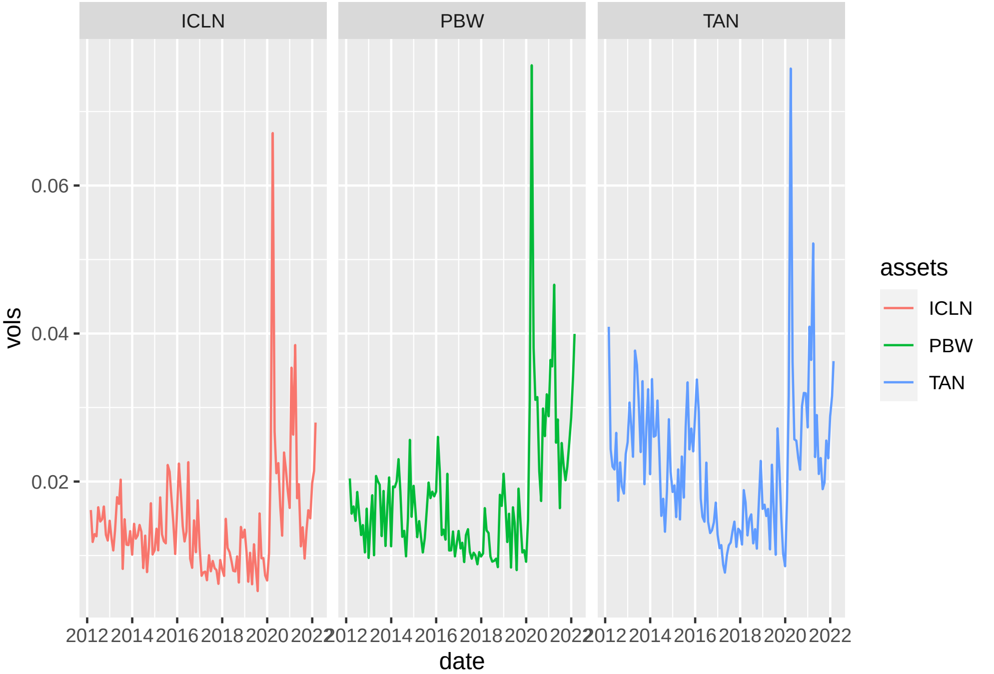

Figure 4: Volatility moves across time.

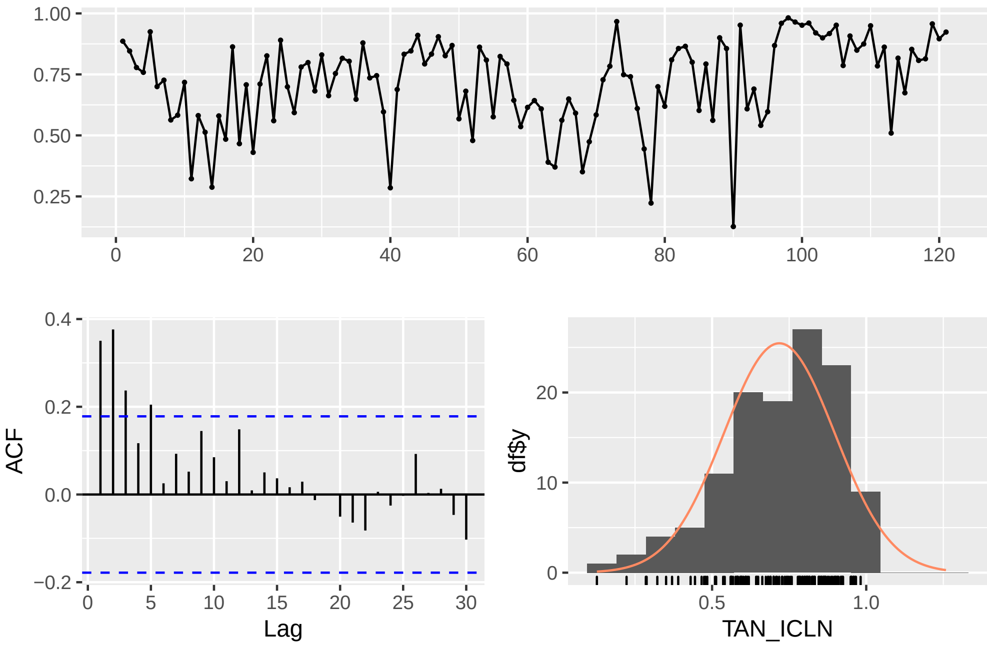

Figure 5: Is the TAN-ICLN relatioship volatile?

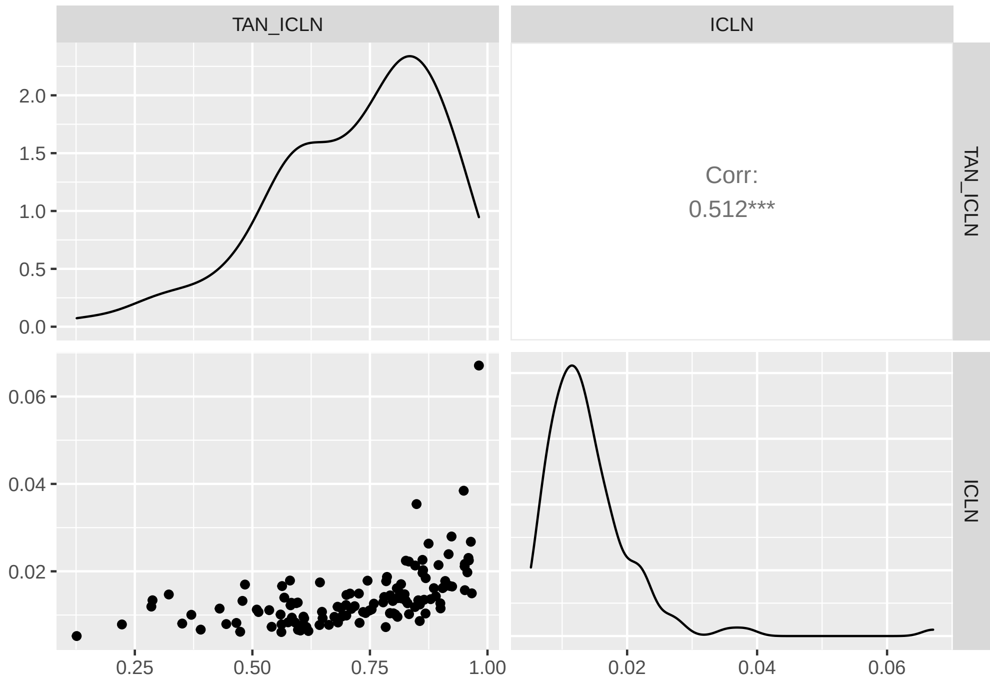

Figure 6: Does ICLN relate to TAN?

- 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 verb *judge* is too strong, but we do seem some recurring patterns in line with previously identified stylizations of market data. The markets would seem to remember whatever it is that correlation and volatility represent to market participants.[7]

Do markets spill into one another?

Market spillover occurs when the volatility of one market, through entanglement,[8] affects the volatility of another market. We have three markets here: TAN, ICLN, and PBW all interacting with one another. We are not asking why, just the question of whether, we observe spillover conditional on our maintained choice of measurements and observational models.

If ICLN is volatile, will TAN be affected? If so, unanticipated competitive movements in one sector (ICLN) will occasion unanticipated movements in another (TAN), here coupled through correlational structures, perhaps to be later modeled with copulae.

Let's examine this idea with a simple scatter matrix from the `GGally` package.

What do we observe in this round?

1. Are the correlations and volatilities 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.

[7]Statistical market memory refers to the influence of past prices on current prices. In a returns calculation an assumption is made that the lag structure is simply one time period ago. Prices are non-stationary and differences are stationary, that is, practically speaking, with no trend. This assumption seems to be further upheld by zero expected values of returns. However, higher moments of returns with a long duration of persisting influences of past observations on current realizations, might suggest a fractional difference return structure. (Hurst, 1951) is usually credited for the first statistical application of fractional differencing to time series and (Granger and Joyeux, 1980) for applications to financial and economic time series.

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

3. Are there potentially multiple regions of outliers? Yes, again, 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.

Thinking in Quantiles

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.[9]

- Just like lm()for Ordinary Least 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.

```
#>
#> Call: rq(formula = log(y) ~ log(x), tau = taus)
#>
#> tau: [1] 0.1
#>
#> Coefficients:
#>             coefficients lower bd upper bd
#> (Intercept) 1.0690       1.0581   2.0200
#> log(x)      0.4011       0.3368   0.6394
#>
#> Call: rq(formula = log(y) ~ log(x), tau = taus)
#>
#> tau: [1] 0.25
#>
#> Coefficients:
#>             coefficients lower bd upper bd
#> (Intercept) 1.2320       0.7583   1.7384
#> log(x)      0.3792       0.2820   0.5383
#>
#> Call: rq(formula = log(y) ~ log(x), tau = taus)
#>
#> tau: [1] 0.5
#>
#> Coefficients:
#>             coefficients lower bd upper bd
#> (Intercept) 1.1585       0.8946   1.3260
#> log(x)      0.3340       0.2746   0.3780
#>
#> Call: rq(formula = log(y) ~ log(x), tau = taus)
#>
#> tau: [1] 0.75
#>
#> Coefficients:
#>             coefficients lower bd upper bd
#> (Intercept) 1.0588       0.6348   1.5228
#> log(x)      0.2893       0.1732   0.4059
#>
#> Call: rq(formula = log(y) ~ log(x), tau = taus)
#>
#> tau: [1] 0.9
#>
```

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

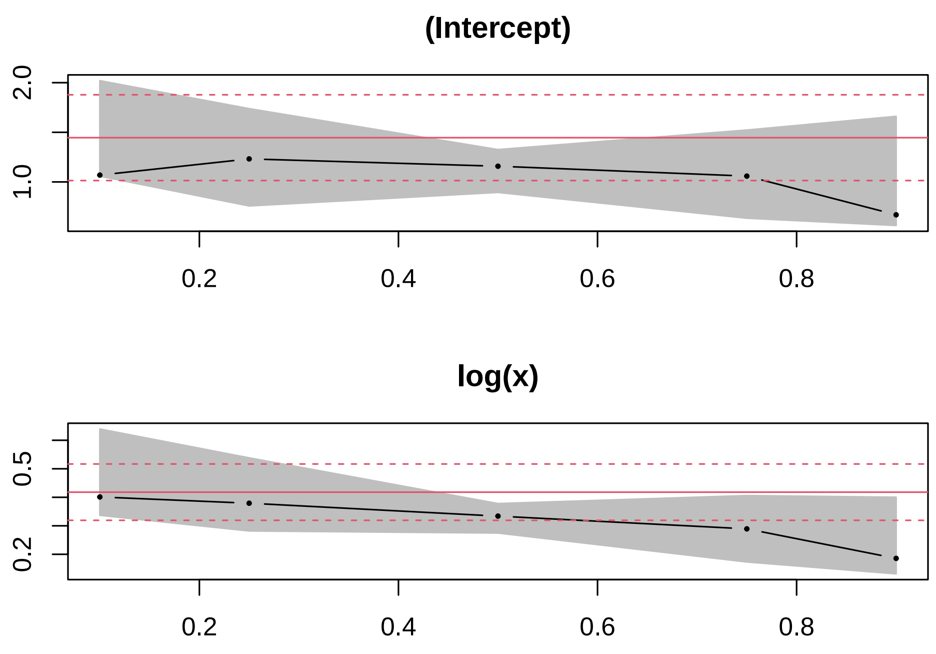

Figure 7: Volatility impact across quantiles.

```
#> Coefficients:
#>             coefficients lower bd upper bd
#> (Intercept) 0.6690       0.5615   1.6606
#> log(x)      0.1857       0.1316   0.4002
```

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 probability 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 probably guessthat preparing for market risk is a very high risk management priority. We may leave those determinations to decision makers for the time being given the narrow scope of this initial and provisional analysis.

One more plot will help further inform the market spillover questions.

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.

Bayesian thoughts

Alternatively, we can examine the impact of the riskiness of one market on the other using a purely probabilistic model. Up to this point we have implemented a robust model of market interaction, albeit in a frequentist mood. 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

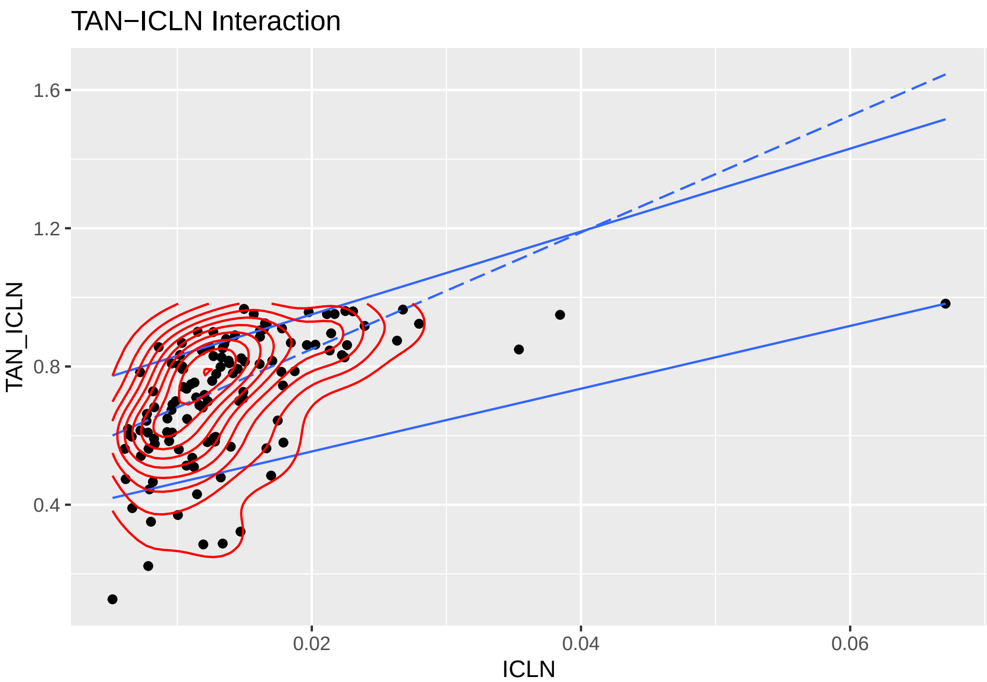

Figure 8: Topography of spillover.

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.

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

The markets are indexed with $[i]$, 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 for a mixed-random effects hierarchical model of the probable market risk infrastructure.

```
#>          mean      sd   5.5%  94.5%
#> a      0.6852 0.02041 0.6526 0.7178
#> b      2.3142 0.93485 0.8201 3.8083
#> sigma  0.1711 0.01140 0.1529 0.1893
```

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.

```
#>          mean       sd   5.5%  94.5%
#> a      0.7283 0.017975 0.6996 0.7571
#> b      3.4013 0.805484 2.1140 4.6886
#> sigma  0.1190 0.008101 0.1061 0.1320
```

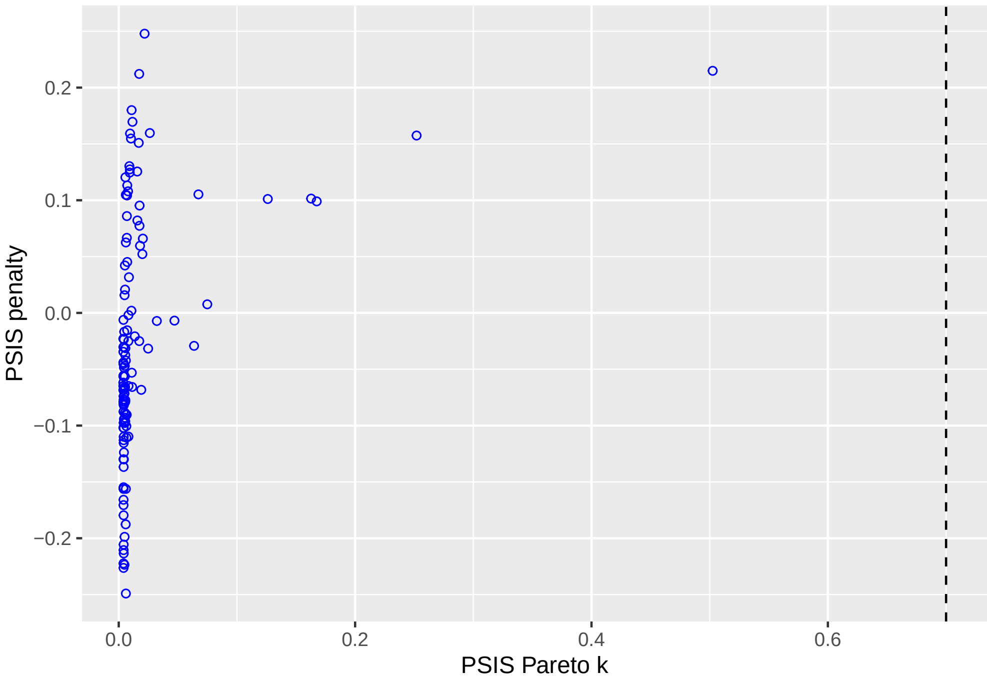

Figure 9: ICLN spills over into TAN.

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

```
#>         mean       sd    5.5%   94.5%
#> a      0.7889 0.014410 0.76588 0.81194
#> b      2.7783 0.672538 1.70349 3.85319
#> sigma  0.0880 0.005809 0.07871 0.09728
```

Wind and solar seem to have similar spillover 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 $k > 0.7$) 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.[10]

Here are the TAN-PBW outlier results.

and

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

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.

The data structure stacks the three markets on top of one another with `corrs` and `vols` observations and an index of 1, 2 and 3 in `mids` for each market. We can `pivot_longer()` the `corr_vols_tbl` and

[10] Power law distributions notoriously do not possess first, second, third, or even fourth moments analytically across the GPD parameter space, especially for k. Sumio Watanabe (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 N.N. Taleb (Taleb, 2018).

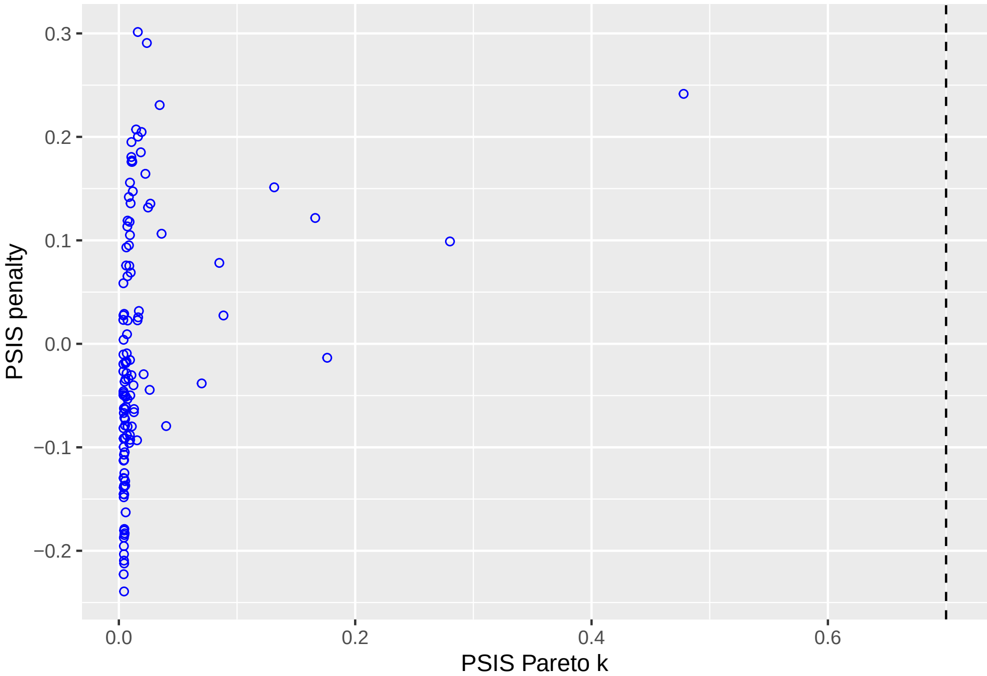

Figure 10: PBW spills over into TAN.

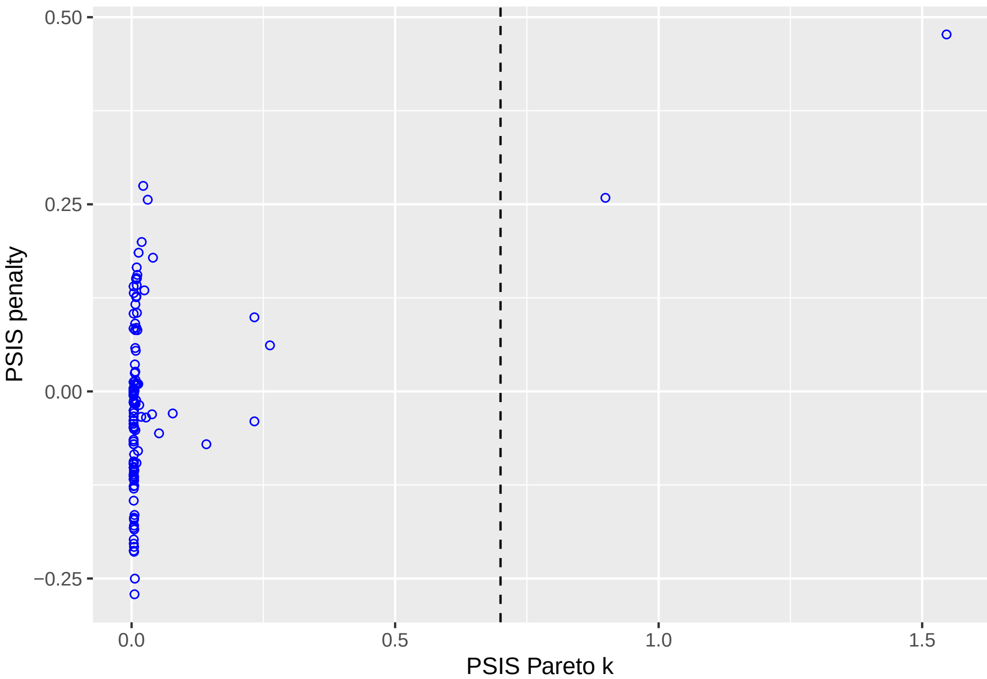

Figure 11: ICLN spills over into PBW.

set out the index against each of the three markets.

Our next model will estimate regression parameters within and across markets.

We use a quadratic approximation technique to estimate the parameters, which, when drawn will produce the marginal distributioins summarized in the `precis()`. It is possible for the optimizer to develop non-optimal ranges of results with this technique, mainly non-positive definite `sigma`. We save a plausible result of a run to replicate results.

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 will place two plots side-by-side, much like a `ggplot2` facet, all deposited from Claus Wilke's very helpful `cowplot` package vignette.[11]

```
#
# 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.

To ground us in a more practical answer. We might interpret a 10% move either in ICLN or PBW volatility as inducing at least a 30% move 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, given this data, and this model.

Again we can review the role of each of the observations in the bias-uncertainty trade-off. 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.

Most of the uncertainty Pareto k) and variability (penalty) 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.[12]

[11] Here's the online site at `https://wilkelab.org/cowplot/articles/plot_grid.html`.

[12] PSIS exploits the distribution of potentially outlying and influential observations using the Generalized Pareto

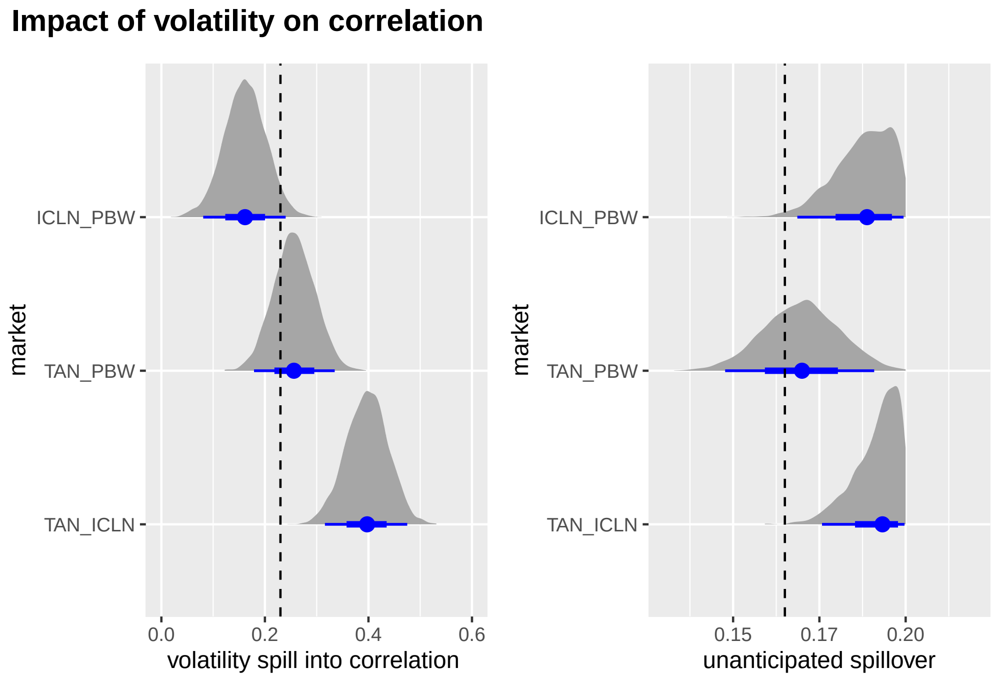

Figure 12: Splitting spillover into systematic and unsystematic aspects.

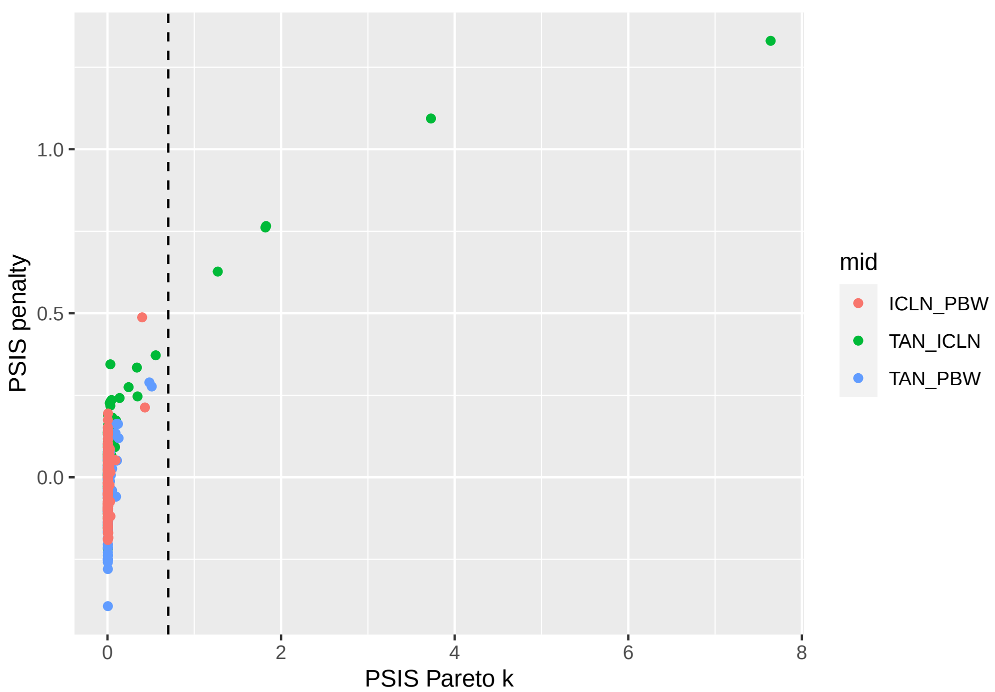

Figure 13: Market Spillover: industry risk and uncertainty

A penultimate thought

We can explore even stranger new statistical spaces by allowing for the correlation of the intercept and slopes of, and across, each market dyad, a random mixed effects model. Here is code directly from Richard McElreath (McElreath, 2020) using his (Stan) `ulam()` mapping function into the Stan Probabilistic Programming environment (Carpenter et al., 2017). We sample correlations between the regression parameters to help us discern the differences and similarities between correlation trends (just the intercept) and the variable impact of volatility on correlations (just the slope). The LKJ (Lewandowski et al., 2009) vine and onion sampler will generate beta distributions of off diagonal correlated samples of the two parameters joined here with a multi-normal copula.

```
corr_vol_spill <- read_csv( "market-spillovers.csv" )
d_4 <- corr_vol_spill %>%
  tibble(
    corr = log( abs(corrs) ),
    vol = log( vols ),
    mid = mids
)
m_5 <- ulam(
  alist(
    corr ~ dnorm( mu_corr, sigma_corr ),
    mu_corr <- v[mid,1] + v[mid,2]*vol,
    sigma_corr ~ dexp( 1 ),
    vector[2]:v[mid] ~ multi_normal( c(abar,bbar), Rho, sigma ),
    abar ~ normal( 0, 1 ),
    bbar ~ normal( 0, 1 ),
    sigma ~ dexp( 1 ),
    Rho ~ lkjcorr(2)
  ),
  data = d_4, log_lik = TRUE
)
save( m_5, file = "m_5")
```

We save the `ulam` run to help with timely document processing. The routine takes about 10 seconds for just 1000 Monte Carlo runs in one Markov Chain. The `vector[2]:v[mid]` deposits a transformed variable into Stan code model block depicted here.

```
data{
    int dates[366];
    int mids[366];
    vector[366] vols;
    vector[366] corrs;
    vector[366] corr;
    vector[366] vol;
    int mid[366];
}
parameters{
    real<lower=0> sigma_corr;
    vector[2] v[3];
    real abar;
    real bbar;
    vector<lower=0>[2] sigma;
    corr_matrix[2] Rho;
}
model{
    vector[366] mu_corr;
```

distribution 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 informational 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 (Watanabe, 2009).

Table 1: WAIC: which model leaks less information?

model	WAIC	SE	dWAIC	dSE	pWAIC	weight
LKJ-m_5	-149	72	0	NA	14	1
no-LKJ-m_4	-138	76	11	11	25	0

```
    Rho ~ lkj_corr( 2 );
    sigma ~ exponential( 1 );
    bbar ~ normal( 0 , 1 );
    abar ~ normal( 0 , 1 );
    {
    vector[2] MU;
    MU = [ abar , bbar ]';
    v ~ multi_normal( MU , quad_form_diag(Rho , sigma) );
    }
    sigma_corr ~ exponential( 1 );
    for ( i in 1:366 ) {
        mu_corr[i] = v[mid[i], 1] + v[mid[i], 2] * vol[i];
    }
    corr ~ normal( mu_corr , sigma_corr );
}
generated quantities{
    vector[366] log_lik;
    vector[366] mu_corr;
    for ( i in 1:366 ) {
        mu_corr[i] = v[mid[i], 1] + v[mid[i], 2] * vol[i];
    }
    for ( i in 1:366 ) log_lik[i] = normal_lpdf( corr[i] | mu_corr[i] , sigma_corr );
```

Loading the model data, we can now view the way markets mix.

```
#>            mean     sd   5.5% 94.5% n_eff Rhat4
#> sigma_corr 0.19 0.0075  0.182  0.21   383     1
#> v[1,1]     1.28 0.1989  0.964  1.61   188     1
#> v[1,2]     0.38 0.0455  0.309  0.46   185     1
#> v[2,1]     0.82 0.1560  0.595  1.07   228     1
#> v[2,2]     0.26 0.0376  0.205  0.32   227     1
#> v[3,1]     0.59 0.1706  0.330  0.87   153     1
#> v[3,2]     0.19 0.0408  0.123  0.25   149     1
#> abar       0.78 0.3236  0.178  1.25   140     1
#> bbar       0.26 0.1111  0.096  0.43   362     1
#> sigma[1]   0.58 0.4156  0.150  1.38   234     1
#> sigma[2]   0.20 0.1597  0.050  0.46   202     1
#> Rho[1,1]   1.00 0.0000  1.000  1.00   NaN   NaN
#> Rho[1,2]   0.29 0.4237 -0.482  0.85   353     1
#> Rho[2,1]   0.29 0.4237 -0.482  0.85   353     1
#> Rho[2,2]   1.00 0.0000  1.000  1.00   NaN   NaN
```

What is evident here is the way correlated parameters will force their sampled marginal distributions to concentrate. For markets this means we can better discern ranges of impact (slope), trend (intercept), and even variability (sigma_corr) across and within market spillovers. We can also look at the two models in terms of their risk-uncertainty characteristics.

If we compare the two WAIC's we see some of the lower information loss in `m_5` predictive power versus `m_4`. WE can use `rethinking::compare( m_4, m_5 )` to perform the calculations

Model `m_5` is slightly better at prediction than `m_4, if that is what we want. The` `m_5` model generates, again, some very highly uncertain regions at high risk, and again with solar (TAN) and clean technologies (ICLN). But some of the extreme uncertainty we saw in `m_4` seems incorporated into `m_5` with fewer outlying observations here. When we compare the two WAIC's we see some of the lower information loss in `m_5` predictive power versus 'm_4' is probably due to those less uncertain outliers.

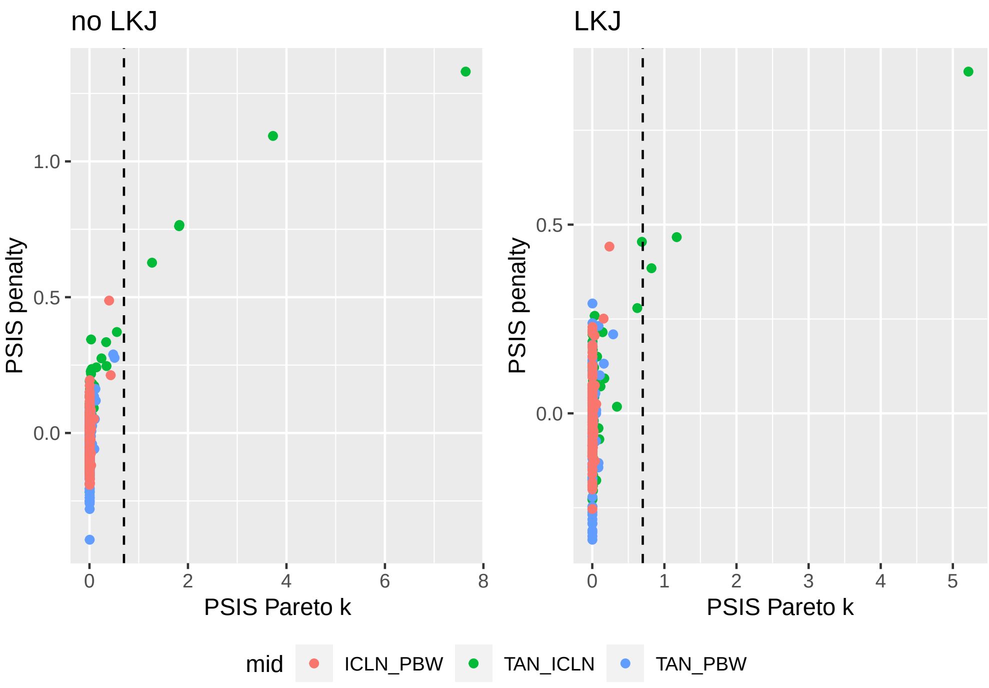

Figure 14: Trading off risk, uncertainty, and models.

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 experimented with the idea 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, into models 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 optimization constraints.

Bibliography

B. Carpenter, A. Gelman, M. D. Hoffman, D. Lee, B. Goodrich, B. M., M. Brubaker, J. Guo, P. Li, and A. Riddell. Stan: A probabilistic programming language. *Journal of Statistical Software*, 76(1), 2017. [p2, 16]

R. Cont. Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance*, 1, 2001. [p1]

C. Granger and R. Joyeux. An introduction to long-range time series models and fractional differencing. *Journal of Time Series Analysis*, 1(1):15–29, 1980. URL //dx.doi.org/10.1111/j.1467-9892.1980.tb00297. [p8]

H. Hurst. Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers*, 5(5):770–799, 1951. [p8]

N. Kaldor. *Capital Accumulation and Economic Growth*, pages 177–222. St. Martin's Press, New York, 1961. URL https://doi.org/10.1007/978-349-08452-4_10. [p1]

R. Koenker. *Quantile Regression (Econometric Society Monographs)*. Cambridge University Press, 2005. [p9]

D. Lewandowski, D. Kurowicka, and H. Joe. Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis*, 100(9):1989–2001, 2009. ISSN 0047-259X. doi: https://doi.org/10.1016/j.jmva.2009.04.008. URL https://www.sciencedirect.com/science/article/pii/S0047259X09000876. [p16]

R. McElreath. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. CRC Press, second edition edition, Mar. 2020. ISBN 978-0-429-63914-2. URL https://xcelab.net/rm/statistical-rethinking/. [p2, 16]

A. McNeil, R. Frey, and P. Embrechts. *Quantitative Risk Management: Concepts, Techniques and Tools - Revised Edition*. Princeton Series in Finance. Princeton University Press, 2015. ISBN 9781400866281. URL https://books.google.com/books?id=SfJnBgAAQBAJ. [p1]

D. Orrell. *Quantum Economics and Finance: An Applied Mathematics Introduction*. Panda Ohana Publishing, 2020. ISBN 9781916081611. URL https://books.google.com/books?id=OZaMzQEACAAJ. [p8]

Stan Development Team. Rstan: The r interface to stan (r package version 2.17.3) [computer software]. 2020a. URL http://mc-stan.org/. [p2]

Stan Development Team. The stan core library (version 2.18.0). 2020b. URL http://mc-stan.org/. [p2]

N. N. Taleb. The statistical consequences of fat tails: Papers and commentary. *The technical inconcerto*, 1, 2018. URL https://www.bibsonomy.org/bibtex/21e63d63eaa2cca30d30e8038bcf7b53f/schmidt2. [p1, 12]

J. Tinbergen. *An Economic Policy for 1936*, pages 39–84. 1959. URL http://hdl.handle.net/1765/15944. [p1]

S. Watanabe. *Algebraic Geometry and Statistical Learning Theory*. Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press, 2009. ISBN 9780521864671. URL https://books.google.com/books?id=GnnUHasAtUQC. [p12, 16]

William G. Foote
Manhattan College
Department of Business Analytics
Riverdale NY, USA 10461
wfoote01@manhattan.edu

Brian V. Wholey
Manhattan College
Department of Business Analytics
Riverdale NY, USA 10461
bwholey01@manhattan.edu