Fabric Risk Industry Project

Shrey Samdani

September 23, 2021

Project Proposal

We all know there are discrete events that affect the market. Macro events like recessions, market-driven breaks such as occur from over-leverage, and revaluations in fundamentals (e.g. "irrational exuberance"). Each type has a different signature in terms of how it affects prices. For example, market-driven events tend to drop and recover faster, while macro events tend to move slowly over a long time period.

Historically, these events appear with a particular frequency, and can each be modeled as a Poisson arrival. The result is what is like what is called a compound Poisson process. The difference is that when an event arrives, it is not a one-period shock - its effect persists.

In this project, we will attempt to solve and research the following question:

What happens to the distribution of equity returns when we introduce non-Gaussian large shock events, versus the usual assumption of a Gaussian distribution?

We wish to observe changes both for the price paths and also for the ensemble distribution (the distribution of returns projected out to some period in the future). Do we see, for example, fatter tails; does risk grow at less than the square root of time?

We can research this question in three phases:

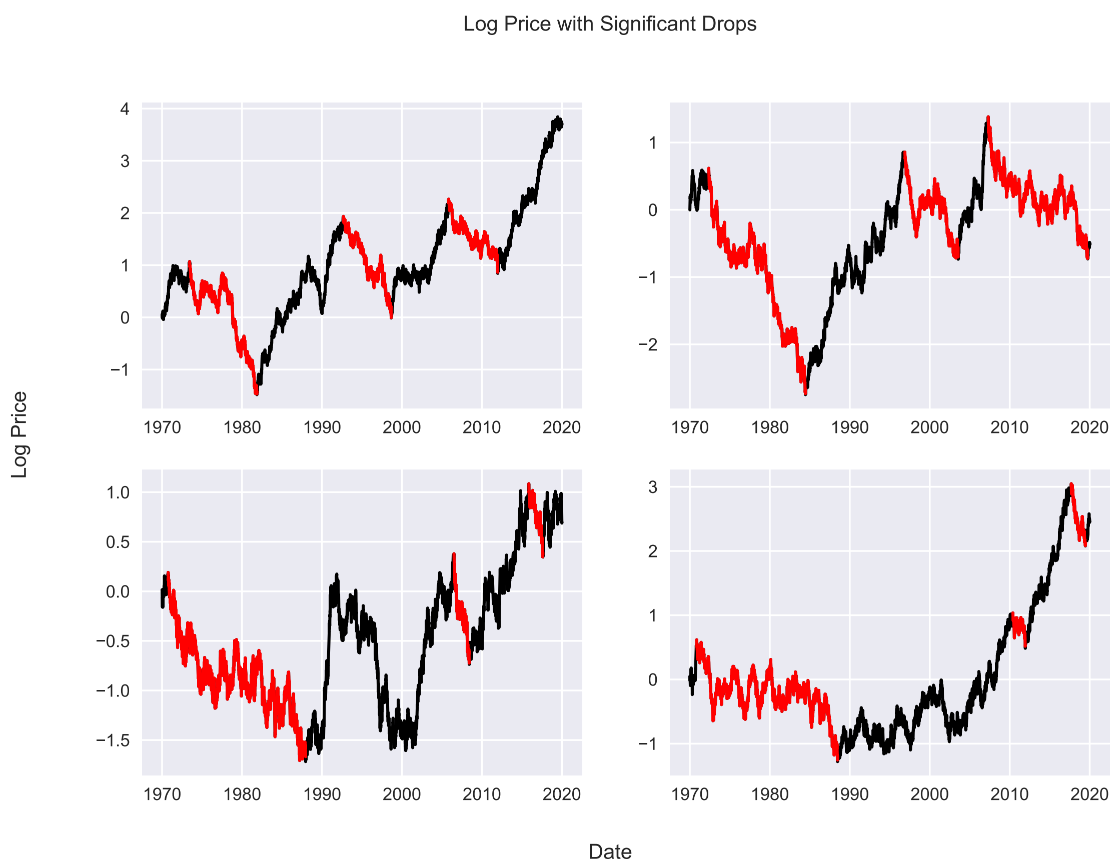
1. Run modified Monte-Carlo simulations of a price paths by simulating Poisson events with a time to a minimum price and a time to recovery and see how that affects the distributions.
    (a) For simplicity, we can start by making the event constant (same shock and time to recovery for every event)
    (b) Next, we can add additional complexity/realism in our simulations by varying the shock parameters (effectively simulating the shocks with a Compound Poisson Process). Additionally, we can modify the volatility for the price paths when they are in the down period of events.
2. Add more realism by setting the events based on historical analysis of the return characteristics of market events.
3. Modify #1 and #2 to see if the amount of run-up in prices before the event changes the nature of its signature. In particular, we modify the extremity or frequency of events based on the current path. For example, if there is a large run-up before the event, do the events tend to be more extreme?
    (a) We will additionally test for path dependence by looking at historical events.

As we get more involved with the project, we may see that we do not see fruitful results in our original question. However, the process of researching will likely give rise to additional questions. In such scenarios, we will attempt to follow down these new paths/questions and inspect any insights they may give.

Project Motivation

As mentioned in the proposal, we have seen shocks in the past that may not follow the normal distribution of log prices. As an example, let us consider the number of 35% price drops from 1970 to 2020 in the S&P 500. There are 5 such instances of these shocks. As a result, if the distribution is correctly specified, we should see an average of 5 price shocks per price path simulated with this normal distribution.

In specific, since we are looking at the log prices, we would want to count the number of instances that a shock of greater than $-\text{Log}[1-0.35] = 0.4308$ occurs in each log price path. Shown below are 4 such simulations of log prices from 1970 to 2020. The volatility has been set to match the S&P 500 (the red coloring shows a significant price drop).



Here, we show the frequency table of the number of price drops for each path of 10000 simulated log price paths.

Number of Drops	Percentage
1	0.8807
2	0.0648
3	0.0255
4	0.0144
5	0.0070
6	0.0033
7	0.0024
8	0.0011
9	0.0006
10	0.0001
11	0.0001

We can see that there is a mean number of 1.2323 shocks over 10000 simulations in the 50 year period spanned above. Clearly, this number is different from the 5 observed shocks in the actual S&P 500, indicating that the normal assumption of log prices is incorrect. In this report, we will attempt to better describe the distribution.

1 Background

Before we get into the project specifics, we begin with a background to pretext the analysis and results in this report.

1.1 Definitions

- **Shock:** A shock represents an event that causes a price path to drop by a certain amount. As a numerical example, a shock of $30\% = 0.3$ means that the stock price will drop by 30%.
- **Shock Length:** The shock length is the length of time for which it takes for the drop in price of a shock to begin, reach its lowest point, and recover.
- **Shock Frequency:** The shock frequency represents the average waiting time between one shock and the next.

1.2 Monte-Carlo Simulation

In order to simulate price paths, we employ a method known as Monte Carlo simulation. In this method, we first assume the distribution of stock price path (more specifically, the distribution of the *increment* of a stock price path over a time interval). We then sample from this distribution to compute each price path increment, and combine the increments to create a price path over a specific time horizon.

In our case, we start by assuming that the stock price at time T can be modeled by the following equation:

$$S_T = S_0 e^{\left(\mu - \frac{\sigma^2}{2}\right)\cdot T + \sigma\cdot\sqrt{T}\cdot W_T}, \tag{1}$$

where $W_T \sim N(0, T)$ is a standard Wiener process with μ as the log return and σ as the log return volatility.

Similarly, we can see that an increment on the logarithmic stock price will follow the following equation:

$$\mathrm{Log}[S_{t,t+\Delta}] = \mathrm{Log}[S_t] + \left(\mu - \frac{\sigma^2}{2}\right)\cdot\Delta + \sigma\cdot\sqrt{\Delta}\cdot W_\Delta \tag{2}$$

We can use this fact to create sample logarithmic price paths for a stock given the proper parameters. If done properly, the terminal logarithmic asset prices should form a normal distribution.

Note: For the rest of this paper, we will assume that $S_0 = 1$. Any other value of S_0 will just result in a vertical shift in the log price path.

1.3 Poisson Process

As mentioned in the proposal, we will be modeling the time between two shocks as a Poisson process. Simply put, the Poisson process is a stochastic process that models the times at which events/arrivals enter a system. Additionally, the Poisson process has the property that the waiting time between such events/arrivals can be modeled by an exponential distribution with density

$$f(x;\lambda) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases} \tag{3}$$

As such, we model the waiting time between stocks using the above exponential distribution.

1.3.1 Compound Poisson Process

Above, we are essentially combining a poisson process with a Wiener process, where each arrival of the poisson process constitutes a "shock". However, the assumption here is that each shock is equivalent, which may not necessarily be the case. In our case, as we will see, we model the shock as a random variable in itself. This changes the process from a poisson process to a compound poisson process, since we have an indendent random variable that constitutes the arrival value of the poisson process.

Mathematically, we can represent this as the following:

$$\text{Log}[S'_{t,t+\Delta}] = \text{Log}[S'_t] + \text{Log}[S_{t,t+\Delta}] + \sum_{i=1}^{N(t)} Y_i(t+\Delta), \tag{4}$$

Where $S'(t)$ represents the shocked price at time t, $N(t)$ refers to the poisson process from section 1.3 and $Y_i(t+\Delta)$ is a random variable that represents the value of shock i at time period $t+\Delta$.

We will test different variants of Y in this report.

As an interesting side note, we find that this is very similar to the model represented in a paper[1] by S. James Press, where a Wiener process (like ours from above) was combined with a compound poisson gaussian process ($Y \sim N(\mu, \sigma)$). In the paper, it was mentioned that such a model can be used for "the study of accident causation; in the study of repeated occurrences in industrial situations; and in the study of the behavior of security price fluctuations," showing that such a model can be applied outside of just the distribution of logged prices.

In a separate paper [2] by the same author, we find that a very similar model has been used in order to model the distribution of logged security prices. As this model is simpler than the one we employ in this report, a closed form solution for the first 4 moments were attained.

[1] S. James Press. "A Modified Compound Poisson Process with Normal Compounding." Journal of the American Statistical Association, vol. 63, no. 322, [American Statistical Association, Taylor & Francis, Ltd.], 1968, pp. 607–13, `https://doi.org/10.2307/2284031`.

[2] S. James Press. "A Compound Events Model for Security Prices." The Journal of Business, vol. 40, no. 3, University of Chicago Press, 1967, pp. 317–35, `http://www.jstor.org/stable/2351754`.

Surprisingly enough, as we will see, the distributional properties of these 4 moments closely resemble those from the resutls in this report. In particular, both models are skewed (nonzero skew), leptokurtic (positive kurtosis), more peaked at its mean than a normal distribution, have high probability mass in the tails, and can be multimodal.

1.4 Annualization

Looking at equation (1), we can see that the log return scales with T and the log return volatility scales with $\sqrt{T}$. To keep numbers more standardized, we can annualize mean by scaling it by a factor of T. We do this by dividing the terminal price mean by T (measured in years) and by dividing volatility by $\sqrt{T}$ (measured in years). As an example (used later in the report), consider that we have daily parametrs μ_{daily} and σ_{daily}. As there are 252 trading days in a year, we model each day as 1/252 years and come up with the following equations:

$$\mu_{daily} \cdot 252 = \mu_{annual} \tag{5}$$

$$\sigma_{daily} \cdot \sqrt{252} = \sigma_{annual} \tag{6}$$

2 Constant Shocks

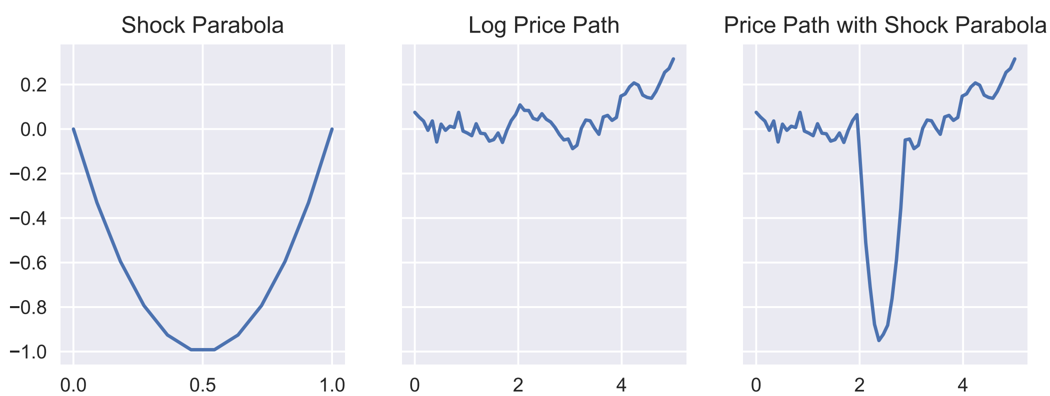
We begin by noting how the behavior of the distribution changes as we add in constant shocks (every shock to the price path will be of the same length and amplitude). For our purposes, we will note changes in the distribution by looking at the first four moments, with emphasis on skew and kurtosis.

In this section, we will use these sample parameters:

- $S_0 = 1$
- $\Delta_T = \frac{1}{12} = 1$ month
- $\sigma = 0.15$
- $\mu = 0.07$
- number of simulations $= n = 10^5$

2.1 Modeling the shock

The shock itself will be modeled by a parabola. We choose a parabola as it is the most intuitive to understand; a parabola is a common continuous and differentiable function. The amplitude of the parabola will be the shock, and it's width will be the shock length. We then take this parabola and overlay it on a log price path. As a visual interpretation, we can look at the following graphs (shock of 1 and time horizon 1).



The parabola on the left has been applied to the price path in the middle to create the price path in right.

For reference, this is the equation for the parabola:

$$y = -4\frac{\text{shock size}}{(\text{shock length})^2} \cdot x^2 + 4\frac{\text{shock size}}{\text{shock length}} \cdot x; \quad y \leq 0 \tag{7}$$

One consideration we also have to take into account is whether or not to allow a "grace period" after one shock occurs - if should or should not be able to allow overlapping shocks.

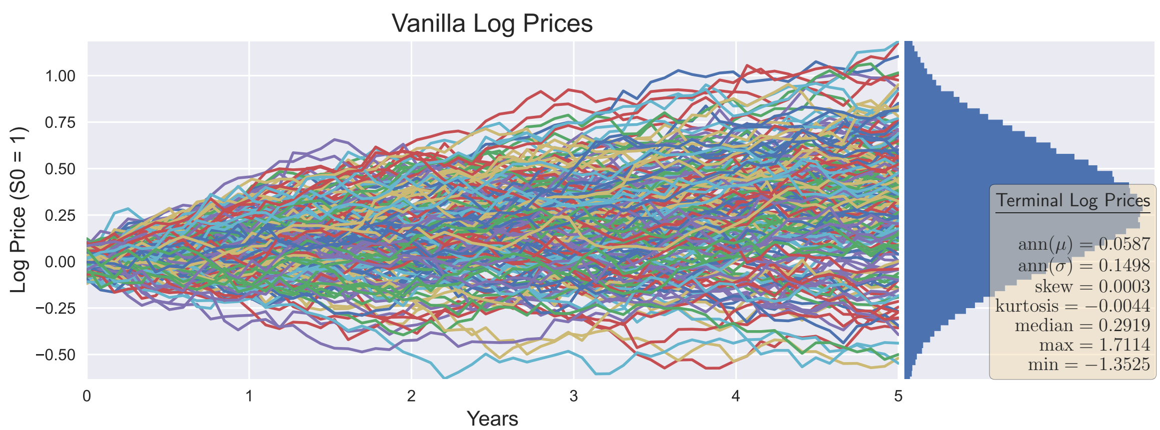
In practice, we have seen that one shock can begin before the previous one has ended, and in our simulations we model the same thing. This also allows for a more "pure" process, leading to more continuous graphs without breaks.

To put this into context, let's consider a shock of 0.3 (30% drop in price). Since we are looking at log prices, in log space this refers to a shock size of $\log(1-0.3) = \log(0.7)$ (since a drop of 0.3 means that we multiply the price by 0.7). Please see the following table for a list of the price drops in log coordinates. Additionally, since we are treating shocks as symmetric parabolas, the time to the bottom is exactly half of the shock length.

Price Shock	Price Drop (%)	Log Price Shock
0.3	30.0%	0.356675
0.5	50.0%	0.693147
0.7	70.0%	1.203973

2.2 No Shock

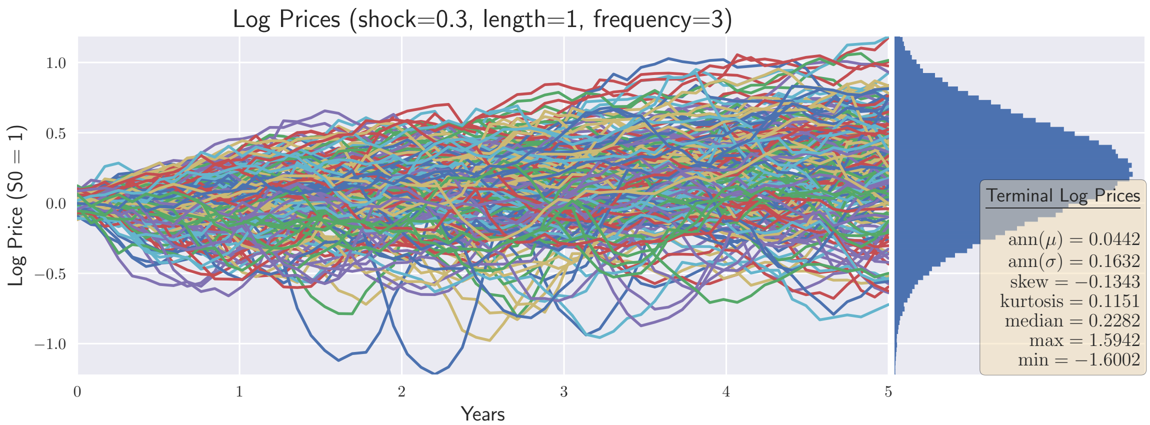
As a control, we first look at the distribution of vanilla (no shock) log prices:

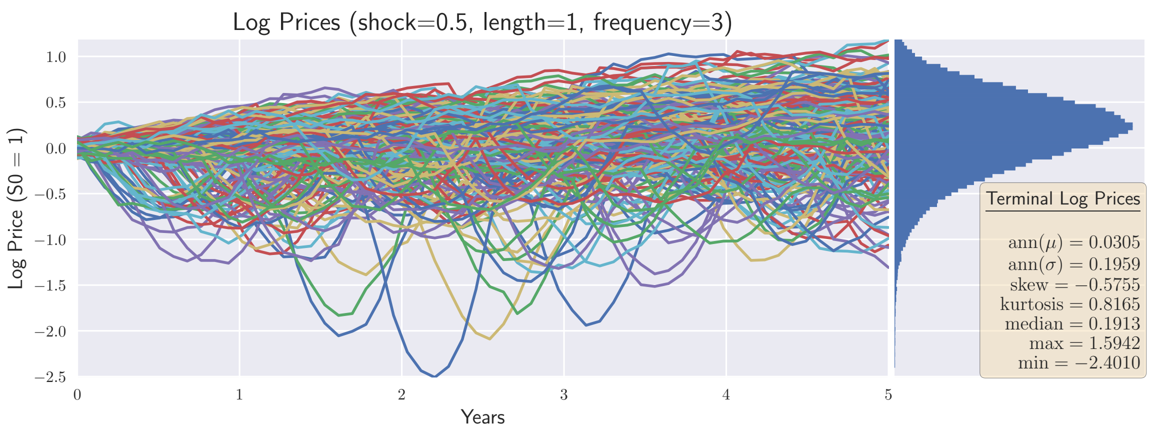


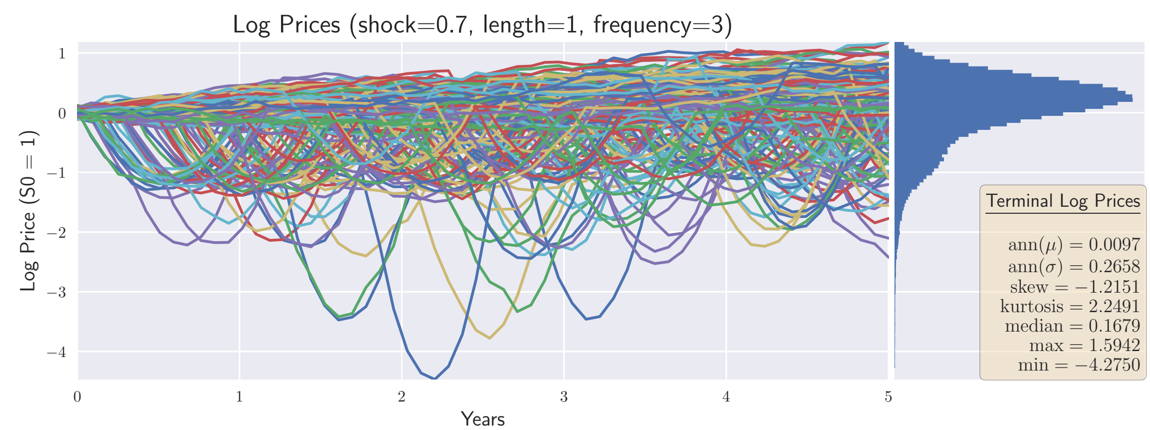
The key things to note here are that skew and kurtosis are both near 0, and approach 0 as $n \to \infty$.

2.3 Adding in constant shocks

Looking at a time horizon of 5 years, a shock frequency of 3 years, a shock length of 1 year, and shocks of size 0.3, 0.5, 0.7, we observe the following changes to the terminal price distributions:







Visually, we can see that adding shocks definitely does affect the distribution of terminal prices. As we increase the size of the shock, the skew decreases and the kurtosis increases, shifting the distribution from normal to a new distribution. Additionally, we see that when increasing the shock to a high number like 0.7, the distribution starts to become slightly bimodal.

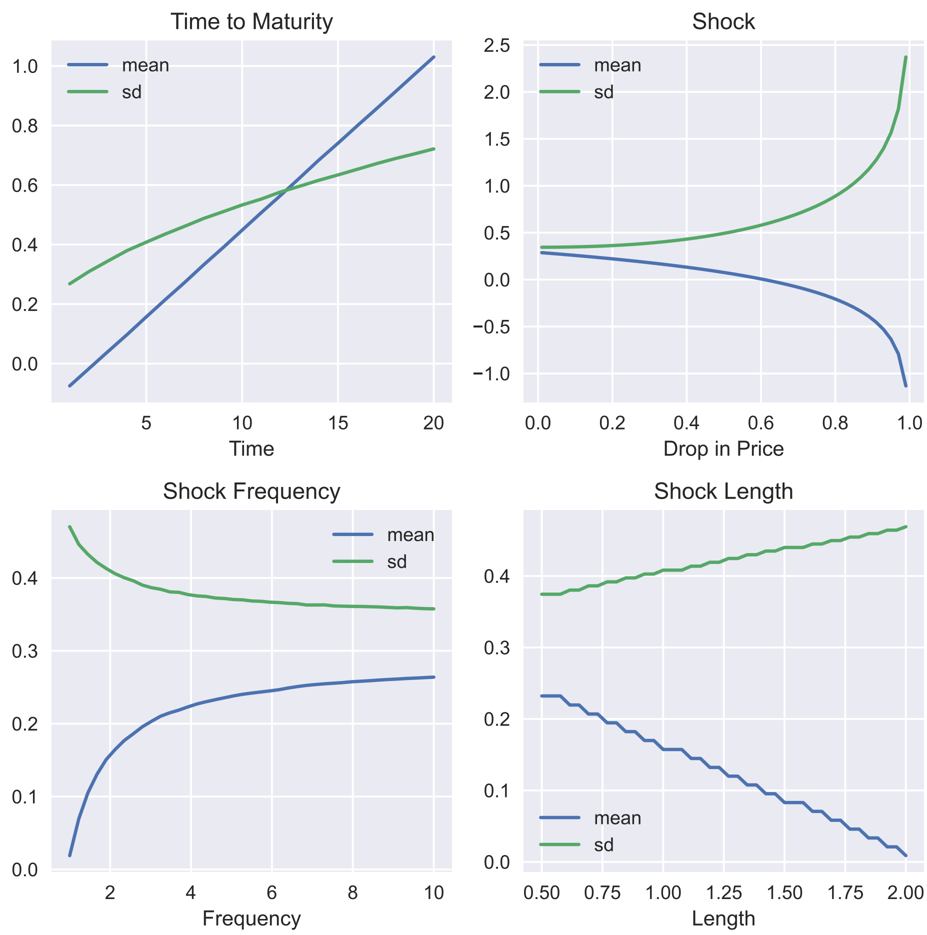
For more graphs and variations, please see the Appendix section 7.1.

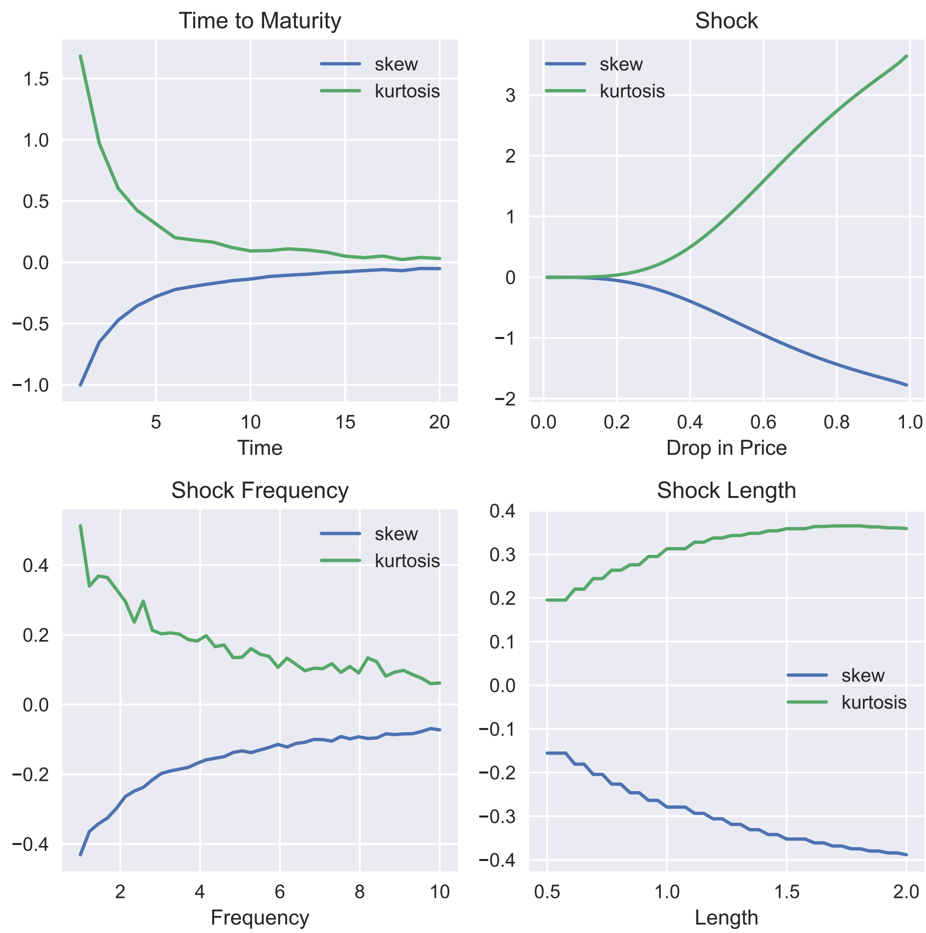
3 Shock Parameters vs First 4 moments

Given that we can compute the first 4 moments for any set of shock parameters, we can create a near continuous equation relating the shock parameters to each of the moments. To observe the effect and shape of each parameter (time horizon, shock, shock frequency, shock length), we vary one parameter at a time, keeping the other parameters constant. For added realism, we have fit the volatility to match the S&P 500. For illustration purposes, we have used the following metrics for the constant shock parameters:

- Price Shock: 35% (log price shock of 0.431)
- Shock Frequency: 2 years
- Shock Length: 1 year
- Time horizon: 5 years

The graphs below show how the first 4 moments of the terminal stock prices vary with changes in each of these parameters. The first four graphs describe the mean and standard deviation (volatility), and the latter four describe the skew and kurtosis. For example, in looking at the very first graph, we can see that both the mean and standard deviation of the terminal prices increase with a longer time horizon.





Clearly, there are some patterns here! Although we have shown the first two moments, our focus will be primarily on the skew and kurtosis.

3.1 Shock

The first two moments of the shock graph change in general as expected: as the shock increases, the mean decreases while the volatility increases. There is somewhat of a logarithmic pattern to each of the two moments.

When looking at the skew and kurtosis, we see that it stays near 0 up until approximately the volatility used in the simulations (0.15). Past this, we see that increasing the shock size actually changes the skew and kurtosis in a linear fashion.

3.2 Time to Maturity

By looking at the graph for the first two moments, we find that the dynamics between time, mean, and volatility are actually mostly unaffected by the shocks! Clearly, the mean scales with time and the volatility scales with the square root of time. In general, the standard deviation is higher and the mean is lower, but they scale in the same way as without the shocks.

The skew and kurtosis are wildly different at low time horizons, but as the time horizon increases we can see that these go to 0. This makes sense: at a large time horizon, the normality of the distribution is higher.

3.3 Shock Frequency

Here, we see nothing particularly surprising. The higher the frequency, the lower the mean and the higher the volatility. This is to be expected, as more frequent shocks will have larger affects in the distribution.

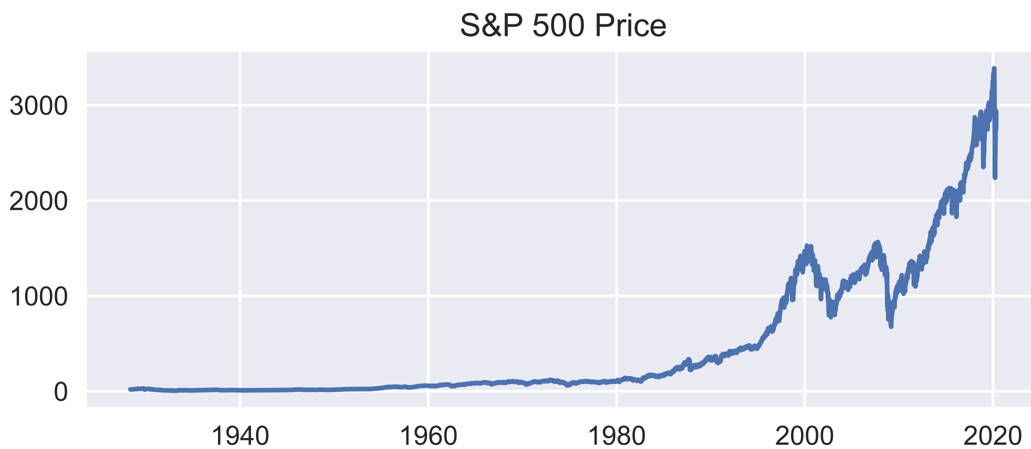
The skew and kurtosis match that of the time to maturity graph. This is expected, since shock frequency and time to maturity are both time relative terms - a high frequency and short time to maturity is similar to a low frequency and long time to maturity.

3.4 Shock Length

Here, the even moments and the odd moments behave in the same way. We see that mean and skew decrease linearly with shock length, and volatility and kurtosis increase with shock length.

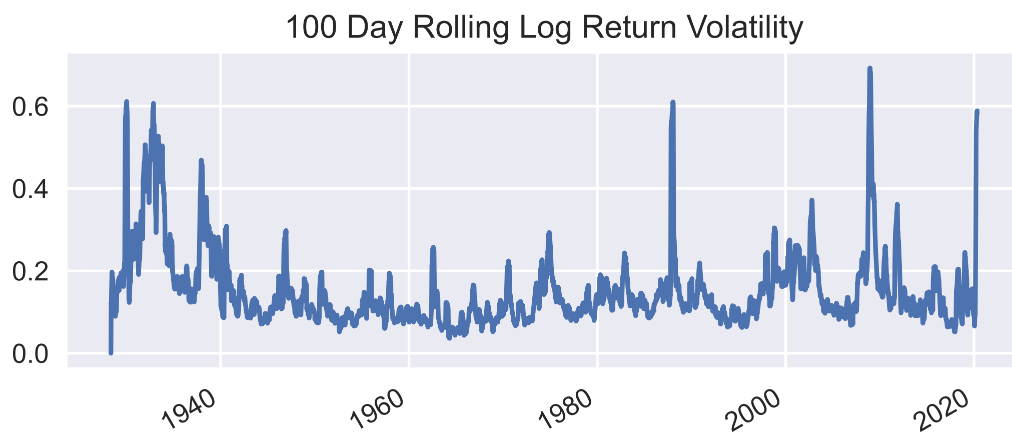
4 Adding Realism (S&P 500)

In order to add more realism into our simulations, we model our parameters using the S&P 500 as a benchmark.



4.1 S&P 500 Volatility

One of the ways we want to add realism is by matching the log return volatility that is described by the S&P 500.



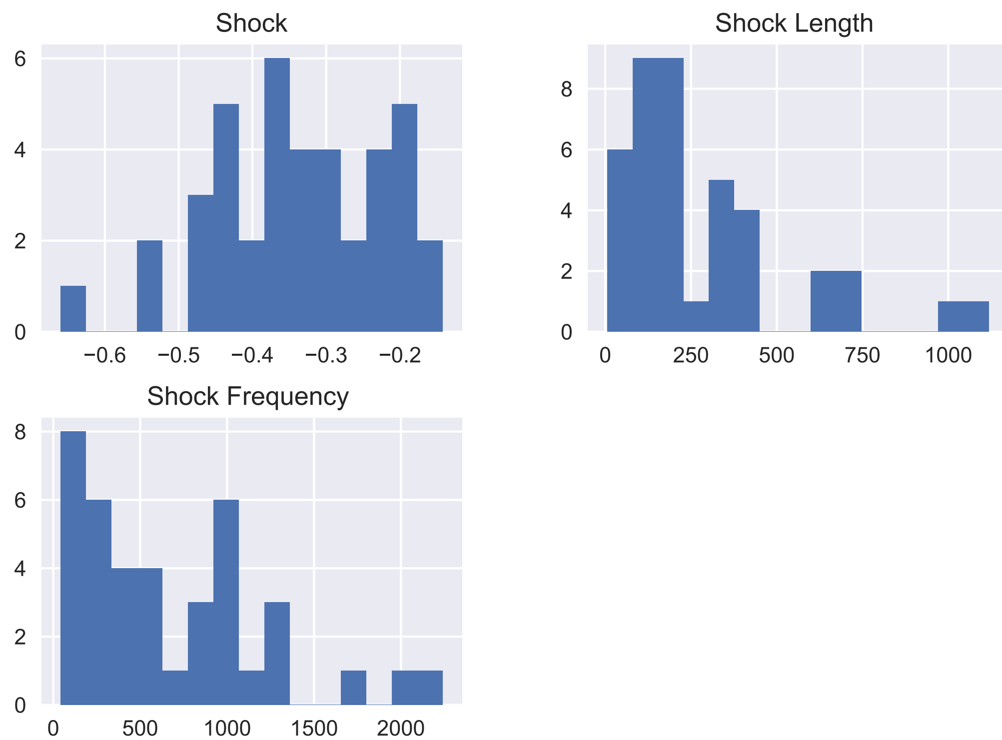
	100 Day Rolling Log Return Volatility
count	24176.000000
mean	0.154320
std	0.094226
min	0.000000
25%	0.096438
50%	0.124824
75%	0.177332
max	0.692608

We can see from above that the annualized volatility hovers around 15% (including shocks). We can compute the volatility without shocks by computing the mean volatility using the bottom 60.0% of values. We find this number to be 10.17%

We can then incorporate these numbers by modifying our simulation to have the respective volatilities when outside or inside of a shock period.

4.2 Shock Length and Shock Magnitude

Furthermore, we can modify the shock length and shock magnitude to match historical data in the S&P 500. Using data about past shocks computed by Richard Bookstaber (40 in total), we observe the following:



	Shock	Shock Length	Shock Frequency
count	40.000000	40	39
mean	-0.338073	284 days 22:12:00	689 days 09:13:50.769230768
std	0.118426	264 days 09:30:48.867891700	528 days 21:50:56.849287536
min	-0.660127	5 days 00:00:00	42 days 00:00:00
25%	-0.422945	107 days 18:00:00	269 days 00:00:00
50%	-0.333899	201 days 00:00:00	553 days 00:00:00
75%	-0.230367	382 days 12:00:00	980 days 12:00:00
max	-0.142036	1119 days 00:00:00	2240 days 00:00:00

Note: Rather than using the annualized return for the shock amount, we annualize the z score. Mathematically, this means that we scale the shock with respect to the square root of the time from the shock start to shock minimum (as opposed to scaling it linearly).

Looking at the histograms, we can see that they can be somewhat modeled by known distributions. We can model the shock and shock length as lognormal distributions, fit to the data from above. Similarly, we can see that the shock frequency does in fact look similar to an exponential distribution, solidifying our assumption about poisson processes from above.

Although these do not look like perfect matches of the distributions, they should be enough for our simulations to estimate a relatively good accuracy. Ideally, we would model this with more than 40 data points.

If we incorporate these shock parameters in, we will effectively replace our poisson process with a compound poisson process, as our shock parabola is not a random variable rather than constant.

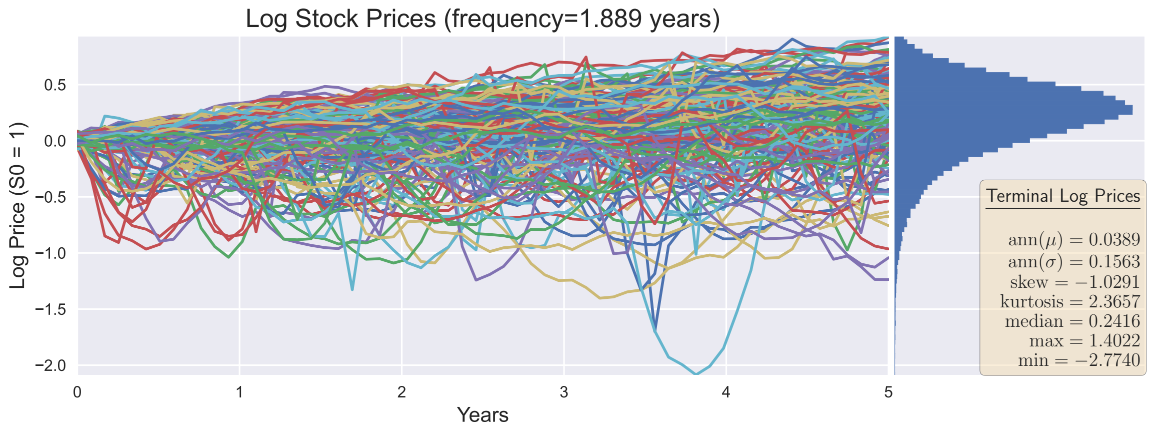
4.3 Incorporating the Realism Factors

Below, we show the simulations using the above metrics:

- **matching in-shock and out-of-shock volatility.** This is done by increasing the volatility in the brownian motion process during the case when we are in a shock.
- **matching historical shock amplitudes.** As mentioned above, for each shock we now draw from a lognormal distribution fit from the shock data above.
- **matching historical shock lengths.** Similar to the shocks, we do this by drawing from a lognormal distribution fit to the historical shock lengths.
- **mathcing the mean frequency of events**. This is done by modifying our exponential distribution we use to pick the waiting times between each shock. We set the mean such that it matches the mean of the waiting times between shocks (approximately 1.9 years).

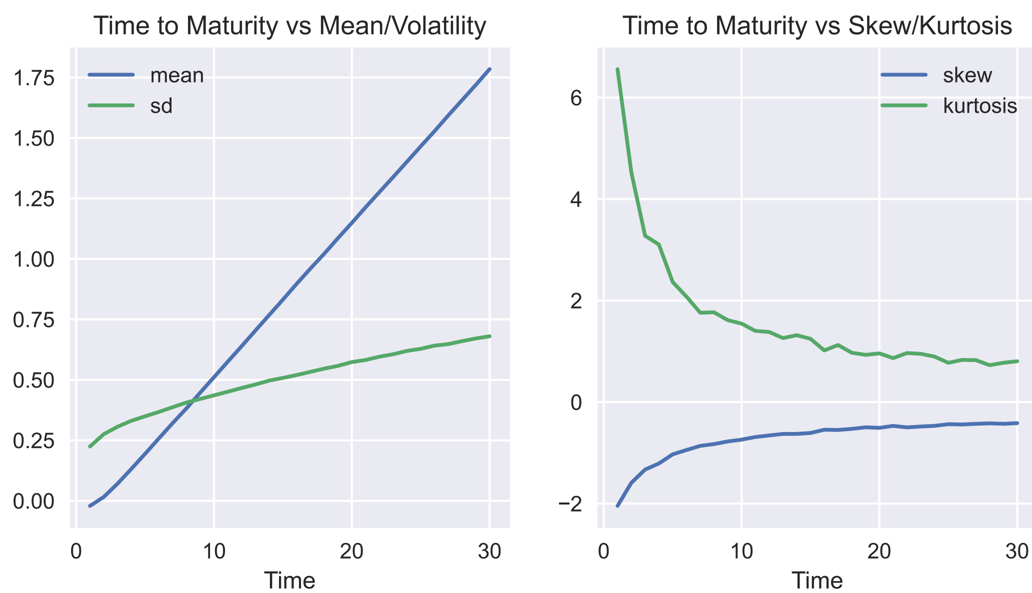
Additionally, we use a log return rate of 7% (the "actuarial rate").

Note: We still use $S_0 = 1$, since we are looking at log prices. If we were to use a different starting price, we would just be shifting the graph up or down.

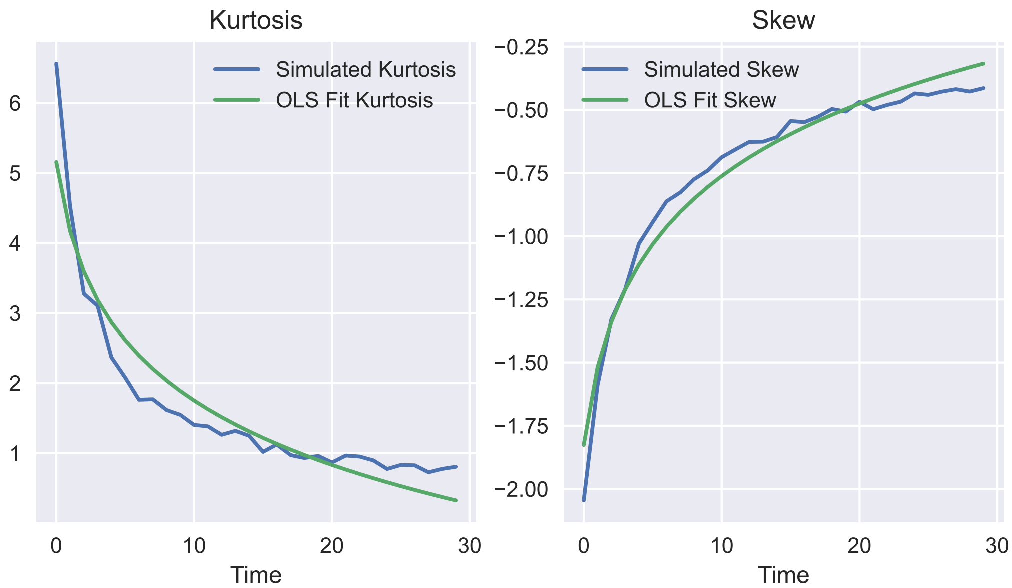


Clearly, we can see that the distribution of terminal prices is in fact affected by the shocks. The skew and kurtosis here are very significant different from 0, and the distribution does not match that of a normal distribution. In fact, this distribution looks more lognormal.

Plotting the moments against time, we get:



We see similar behavior as we did in the simulations from section 3. We can see that the skew and kurtosis seem to vary with the logarithm of time. Fitting this to our data, we get the following:



Dep. Variable:	kurtosis	**R-squared:**	0.897
Model:	OLS	**Adj. R-squared:**	0.894
Method:	Least Squares	**F-statistic:**	245.0
Date:	Sun, 12 Sep 2021	**Prob (F-statistic):**	2.24e-15
Time:	14:42:33	**Log-Likelihood:**	-15.186
No. Observations:	30	**AIC:**	34.37
Df Residuals:	28	**BIC:**	37.17
Df Model:	1		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	5.1552	0.238	21.644	0.000	4.667	5.643
Log_T	-1.4200	0.091	-15.651	0.000	-1.606	-1.234

Omnibus:	13.972	**Durbin-Watson:**	0.418
Prob(Omnibus):	0.001	**Jarque-Bera (JB):**	16.270
Skew:	1.186	**Prob(JB):**	0.000293
Kurtosis:	5.719	**Cond. No.**	9.33

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Dep. Variable:	skew	**R-squared:**	0.964
Model:	OLS	**Adj. R-squared:**	0.963
Method:	Least Squares	**F-statistic:**	759.8
Date:	Sun, 12 Sep 2021	**Prob (F-statistic):**	7.80e-22
Time:	14:42:33	**Log-Likelihood:**	36.698
No. Observations:	30	**AIC:**	-69.40
Df Residuals:	28	**BIC:**	-66.59
Df Model:	1		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	-1.8259	0.042	-43.217	0.000	-1.912	-1.739
Log_T	0.4436	0.016	27.565	0.000	0.411	0.477

Omnibus:	6.348	**Durbin-Watson:**	0.296
Prob(Omnibus):	0.042	**Jarque-Bera (JB):**	4.604
Skew:	-0.865	**Prob(JB):**	0.100
Kurtosis:	3.832	**Cond. No.**	9.33

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

We see very high values of R^2 and a pretty good fit on the graph! We additionally now have an equation to model the skew and volatility as a function of the logarithm of the time horizon.

$$\begin{aligned}\text{Skew} &= -1.8259 + 0.4436 \cdot \text{Log[Time]} \\ \text{s.e. Intercept} &= 0.042 \\ \text{s.e. Log[Time]} &= 0.016\end{aligned}$$

$$\begin{aligned}\text{Kurtosis} &= 5.1552 - 1.42 \cdot \text{Log[Time]} \\ \text{s.e. Intercept} &= 0.238 \\ \text{s.e. Log[Time]} &= 0.091\end{aligned}$$

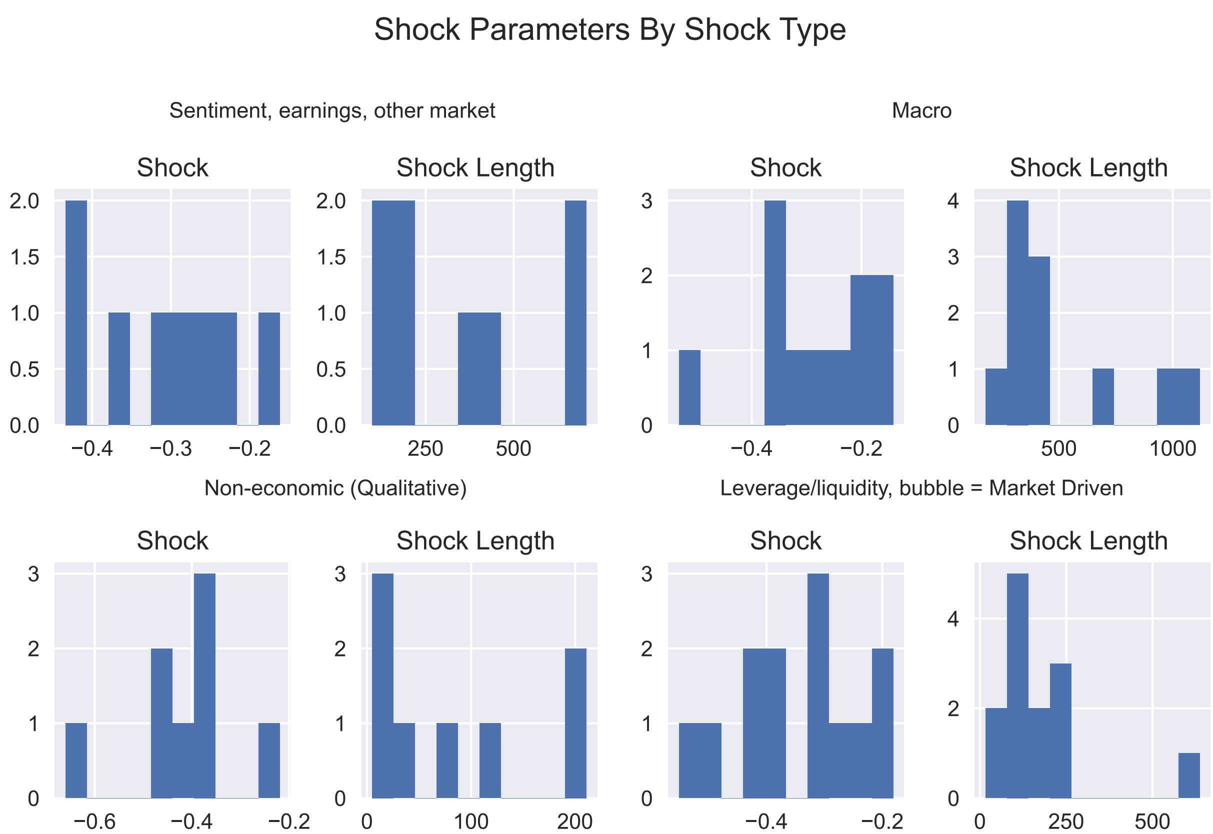
We can see that the skew starts low at around -1.8 and gradually increases with the logarithm of time (eventually reaching zero). Similarly, the kurtosis starts at around 5 and gradually decreases with the logarithm of time (eventually reaching 0). Given the relatively low standard errors and the high correlation, we assume that these parameters are close to the true values (assuming that the distribution is correct). The interesting fact here is that we can see that there is a strong relationship between the logarithm of the time horizon and higher moments.

4.4 Modeling the shock type

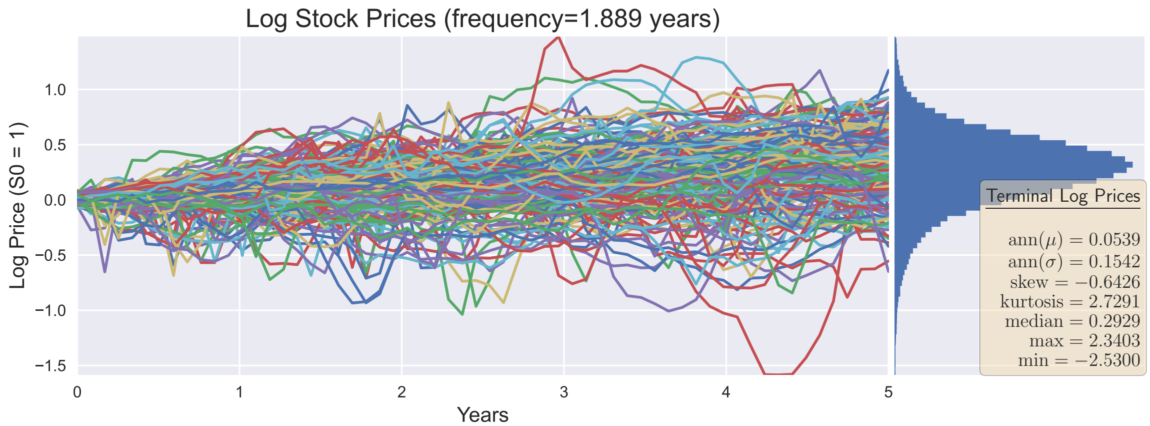
To add increased realism, we will further compound the shock by conditioning it on shock type. The shock type (Sentiment, Macro, Non-economic, or Market driven) will be chosen by a uniform random variable. The shock parameters will then be chosen based on the shock type.

Below, even without many data points, we see that the for all shock types except the first one, the shocks follow somewhat of a lognormal distribution. For the shock type regarding "Catalyst: Sentiment, earnings, other market", we see that this one is more of a uniform distribution. We can first sample the shock type based on the probabilities below and then sample the shock frequency and shock amplitude based on the type of shock.

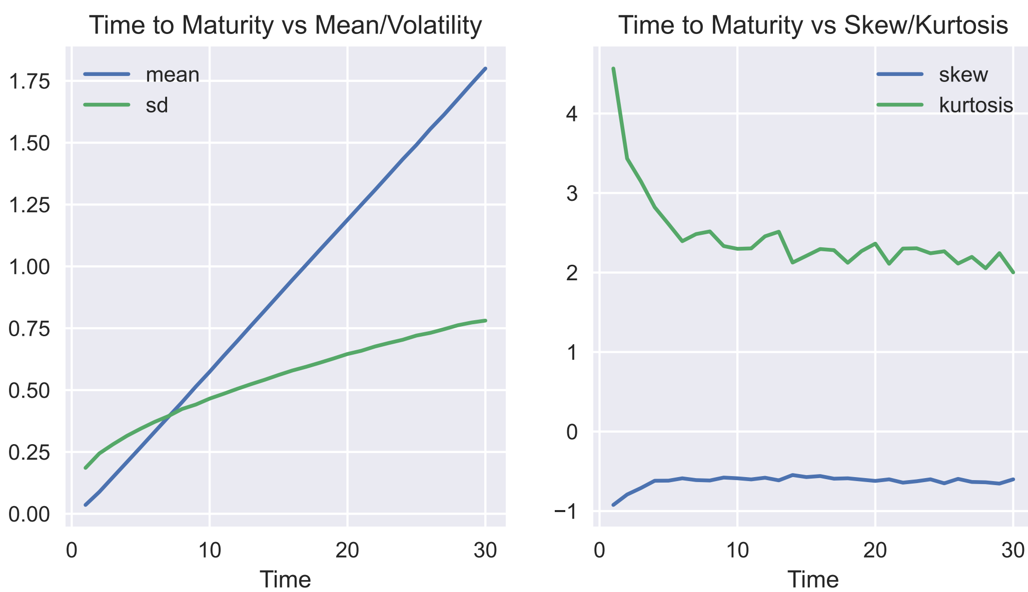
Shock Type	Frequency
Sentiment, earnings, other market	0.200
Macro	0.275
Non-economic (Qualitative)	0.200
Leverage/liquidity, bubble = Market Driven	0.325



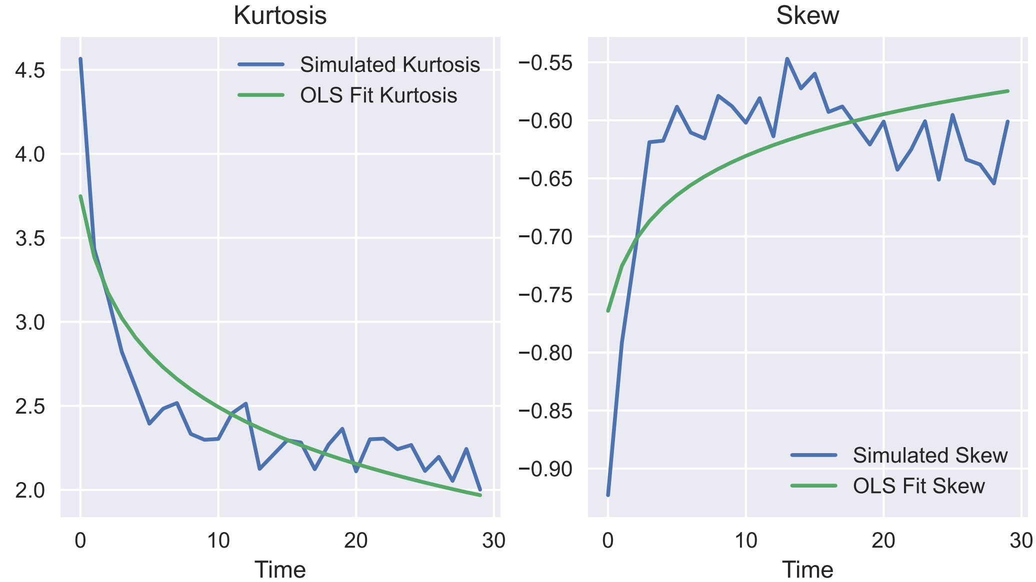
Simulating the price path, we see that it does have nonzero skew and kurtosis.



Similar to what we did in section 4.3, we plot the moments of these simulations as a function of time and fit the skew and kurtosis using the logarithm of time as our independent variable.



Additionally, the regression results are shown here.



Dep. Variable:	kurtosis	**R-squared:**	0.782
Model:	OLS	**Adj. R-squared:**	0.775
Method:	Least Squares	**F-statistic:**	100.7
Date:	Sun, 12 Sep 2021	**Prob (F-statistic):**	8.91e-11
Time:	15:34:47	**Log-Likelihood:**	1.4308
No. Observations:	30	**AIC:**	1.138
Df Residuals:	28	**BIC:**	3.941
Df Model:	1		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	3.7480	0.137	27.380	0.000	3.468	4.028
Log_T	-0.5233	0.052	-10.035	0.000	-0.630	-0.416

Omnibus:	14.345	**Durbin-Watson:**	0.760
Prob(Omnibus):	0.001	**Jarque-Bera (JB):**	18.308
Skew:	1.142	**Prob(JB):**	0.000106
Kurtosis:	6.070	**Cond. No.**	9.33

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Dep. Variable:	skew	**R-squared:**	0.427
Model:	OLS	**Adj. R-squared:**	0.406
Method:	Least Squares	**F-statistic:**	20.83
Date:	Sun, 12 Sep 2021	**Prob (F-statistic):**	9.13e-05
Time:	15:34:47	**Log-Likelihood:**	45.011
No. Observations:	30	**AIC:**	-86.02
Df Residuals:	28	**BIC:**	-83.22
Df Model:	1		

	coef	std err	t	P> |t|	[0.025	0.975]
Intercept	-0.7641	0.032	-23.861	0.000	-0.830	-0.699
Log_T	0.0557	0.012	4.564	0.000	0.031	0.081

Omnibus:	4.417	**Durbin-Watson:**	0.489
Prob(Omnibus):	0.110	**Jarque-Bera (JB):**	3.024
Skew:	-0.749	**Prob(JB):**	0.220
Kurtosis:	3.420	**Cond. No.**	9.33

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Overall, these results are not as nice as the ones we saw in the previous section. The reason for this, most likely, is that we don't quite have all the data needed to properly simulate what the shocks look like. As such, we should take the results from these simulations with a grain of salt.

Looking at these graphs, the same trends hold for the first two moments - they scale with time and the square root of time respectively. For skew and kurtosis, we see that at all time horizons, the skew and kurtosis is actually nonzero. Skew stays relatively constant around -0.5 with a jump from at time = 1. The kurtosis models that of a logarithmic graph much more, with a relatively high R^2 value.

These results are not as accurate, but have the potential to be more accurate with the increase of data.

5 Extentions / Additional Realism

5.1 Run Up before the shock

Another topic of interest that would add more realism to the simulations is adding the effect of a run up in determining the shock parameters. We hypothesize that having a large run up will lead to a larger drop.

We categorize a run-up by statistics leading up to the shock itself. This includes the return and volatility at certain time periods up to 2 years before the shock. Unfortunately, we did not find that the run up significantly affects the shock, as shown by the low correlations between the variables. For more details, please see the graphs and regressions in the appendix section 7.2.

6 Results

6.1 Overall

We see that by incorporating shocks into the price simulations, the distribution shifts based on the time horizon. In general, we see that the normal assumption of log prices/returns is rejected when incorporating shocks. The resulting distribution does not seem to be well defined, as it has an inconsistent shape based on the shock parameters; we see that for high shock amplitudes, the distribution even goes from unimodal to bimodal.

An additional fact we saw is that the run-up before a shock has no affect on the resulting distribution of log prices.

6.2 Short time horizons

Judging by the time graph in the previous section, we can see that looking at short term time horizons is definitely affected by the presence of these non-normal shocks. In the short term, skew decreases and kurtosis increases. We can see that these will affect significantly investments looking out 10-15 years. As a whole, by assuming a normal distribution, we are understating our risk in such investments.

6.3 Long time horizons

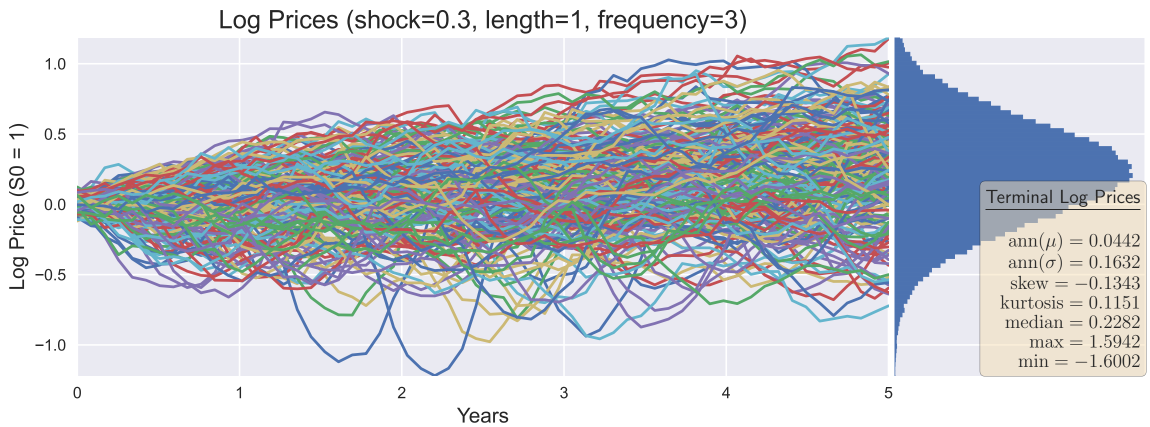
When looking at time horizons greater than 10-15 years, we see that the normal assumption is not rejected. By visually observing the histograms at such time horizons and by looking at the higher moments, we see that they match those that would be found in a normal distribution. As such, investors engaged in long term investing need not take the shocks into account, as they will get "smooted out" over time.

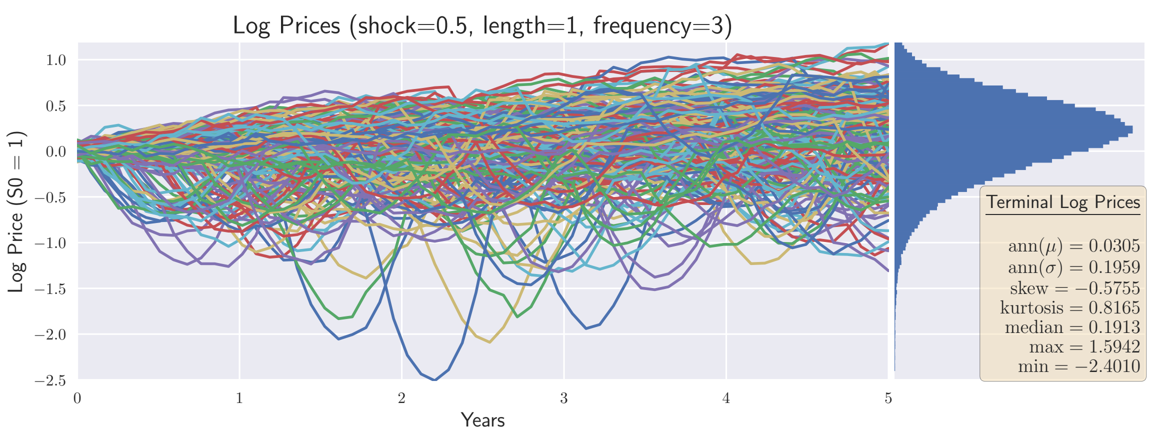
7 Appendix

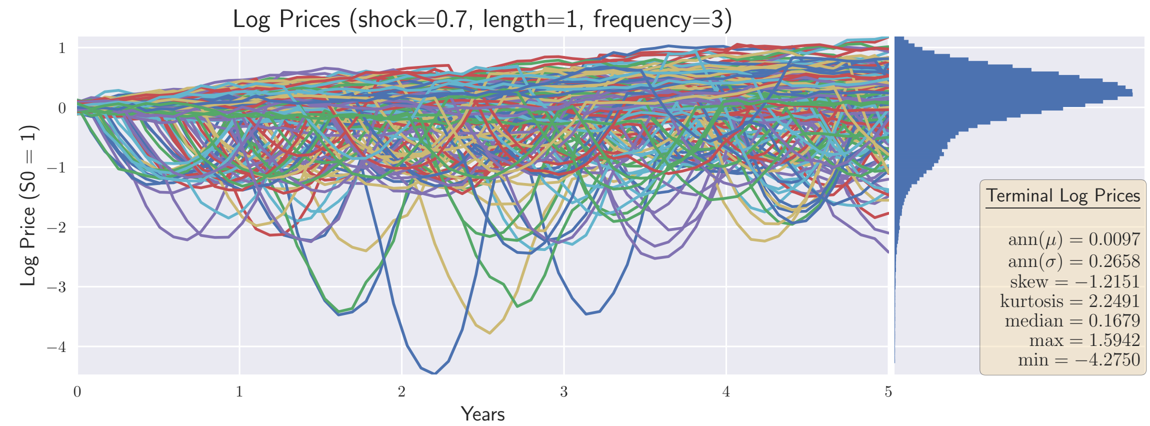
7.1 Constant Shock Additional Graphs

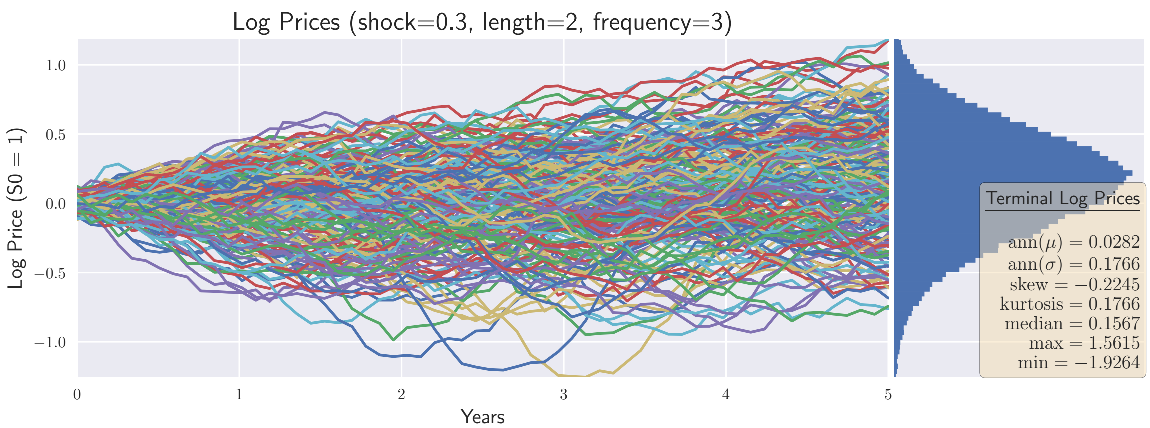
These graphs show the distributions and moments of terminal stock prices given different scenarios. In specific, we try all combinations of the following parameters:

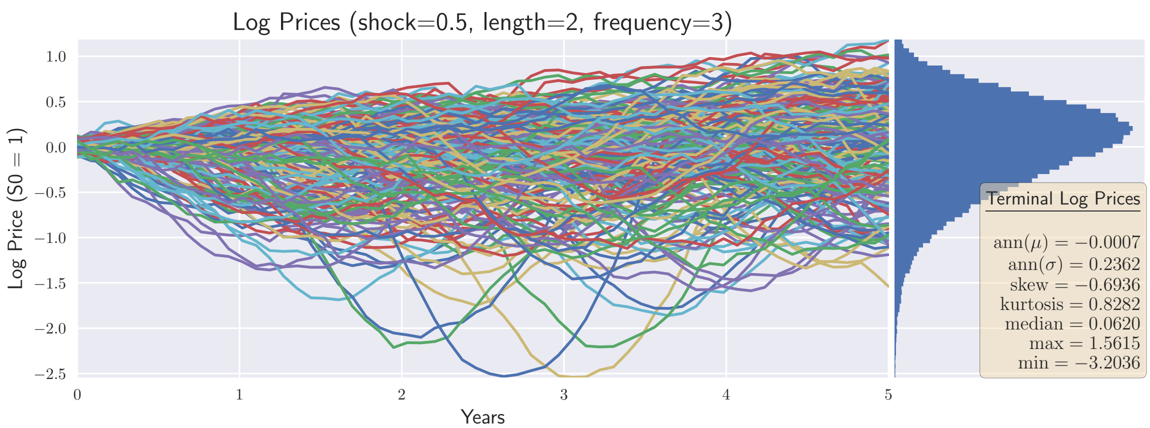
- **Time Horizon**: 5 years, 30 years
- **Shock Frequency**: 3 years, 7 years
- **Shock Lengths**: 1 year, 2 years
- **Price Shocks**: 30%, 50%, 70%

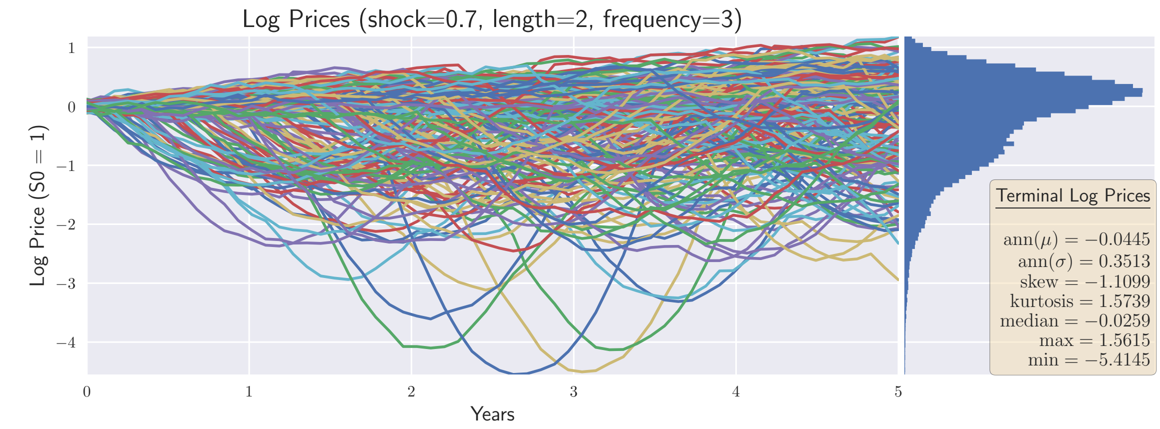


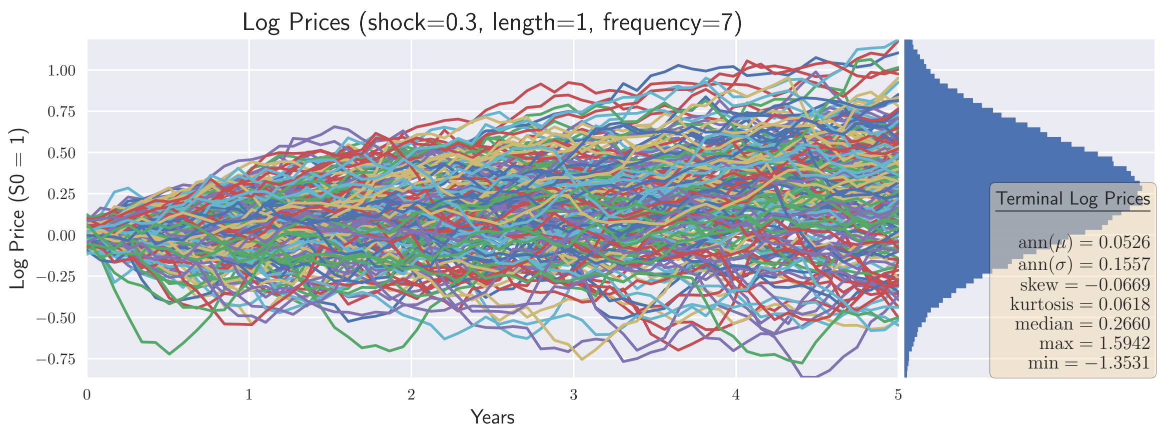


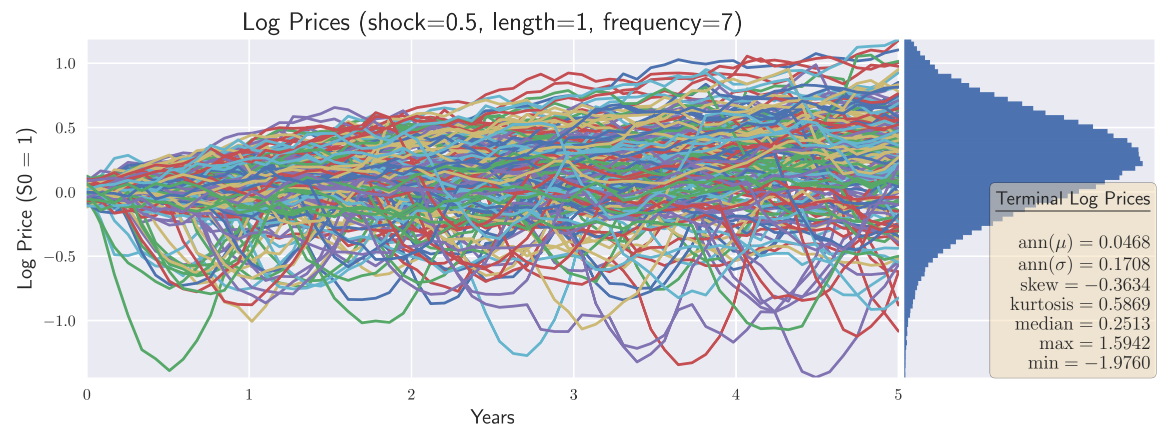

Log Prices (shock=0.7, length=1, frequency=3)
Log Price (S0 = 1)
Years
1
0
−1
−2
−3
−4
0
1
2
3
4
5
Terminal Log Prices
ann(μ) = 0.0097
ann(σ) = 0.2658
skew = −1.2151
kurtosis = 2.2491
median = 0.1679
max = 1.5942
min = −4.2750

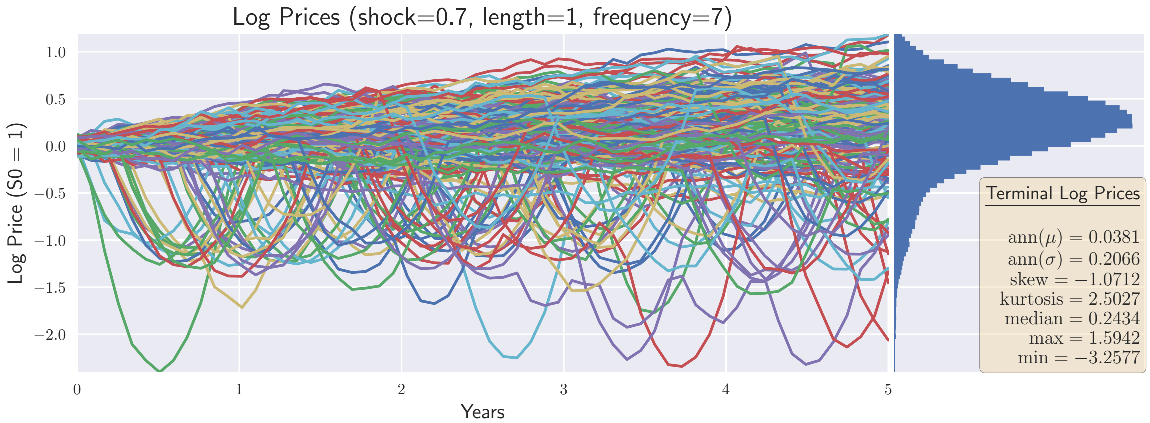

Log Prices (shock=0.3, length=2, frequency=3)
Log Price (S0 = 1)
Years
1.0
0.5
0.0
−0.5
−1.0
0
1
2
3
4
5
Terminal Log Prices
ann(μ) = 0.0282
ann(σ) = 0.1766
skew = −0.2245
kurtosis = 0.1766
median = 0.1567
max = 1.5615
min = −1.9264

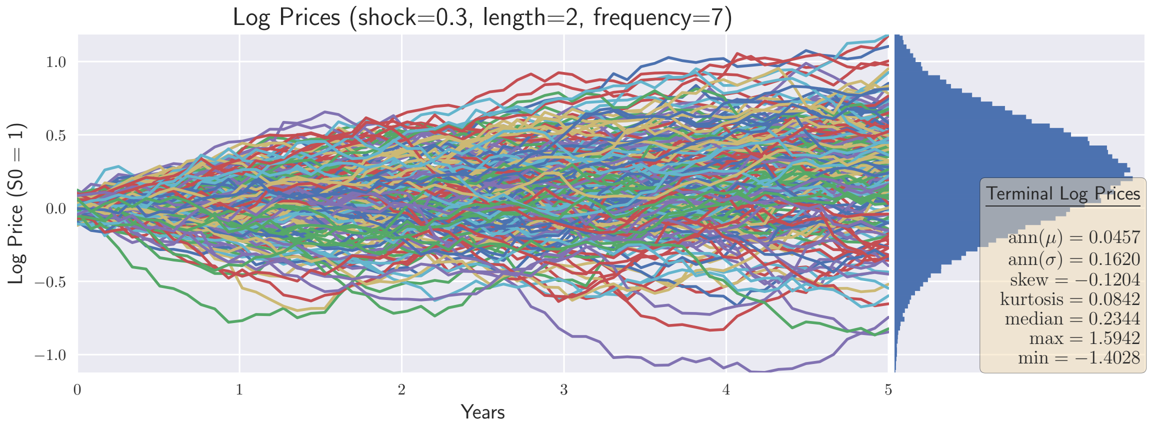

Log Prices (shock=0.5, length=2, frequency=3)
Log Price (S0 = 1)
Years
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
−2.5
0
1
2
3
4
5
Terminal Log Prices
ann(μ) = −0.0007
ann(σ) = 0.2362
skew = −0.6936
kurtosis = 0.8282
median = 0.0620
max = 1.5615
min = −3.2036

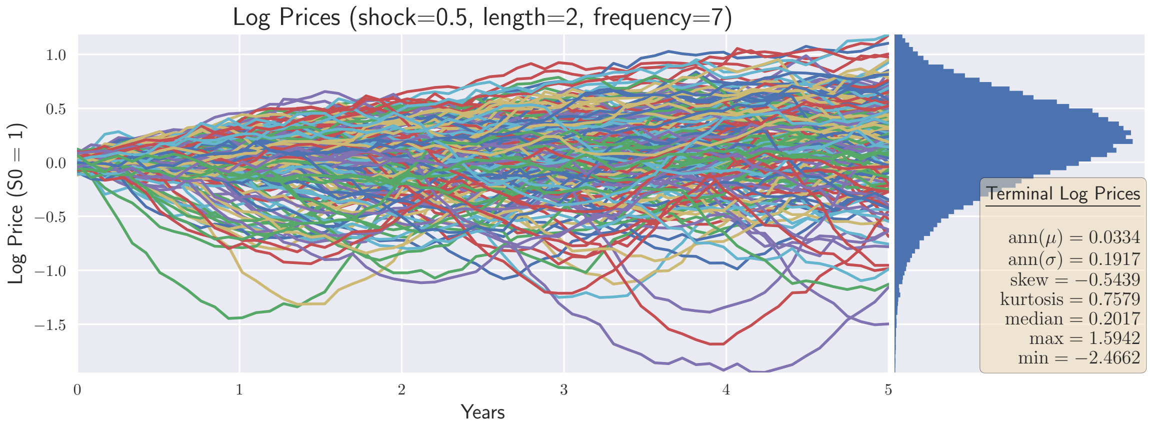

Log Prices (shock=0.7, length=2, frequency=3)
Log Price (S0 = 1)
Years
1
0
−1
−2
−3
−4
0
1
2
3
4
5
Terminal Log Prices
ann(μ) = −0.0445
ann(σ) = 0.3513
skew = −1.1099
kurtosis = 1.5739
median = −0.0259
max = 1.5615
min = −5.4145

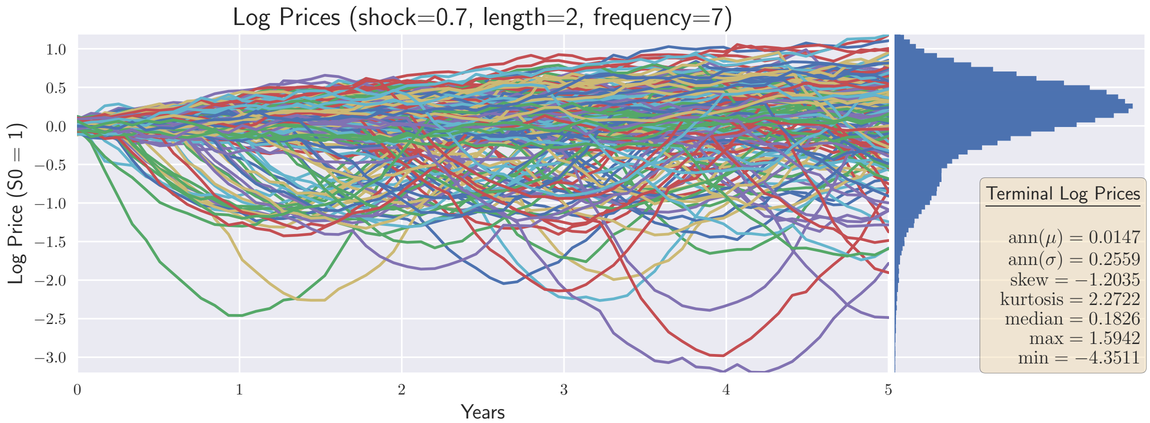

Log Prices (shock=0.3, length=1, frequency=7)
Log Price (S0 = 1)
Years
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
0
1
2
3
4
5
Terminal Log Prices
ann(μ) = 0.0526
ann(σ) = 0.1557
skew = −0.0669
kurtosis = 0.0618
median = 0.2660
max = 1.5942
min = −1.3531

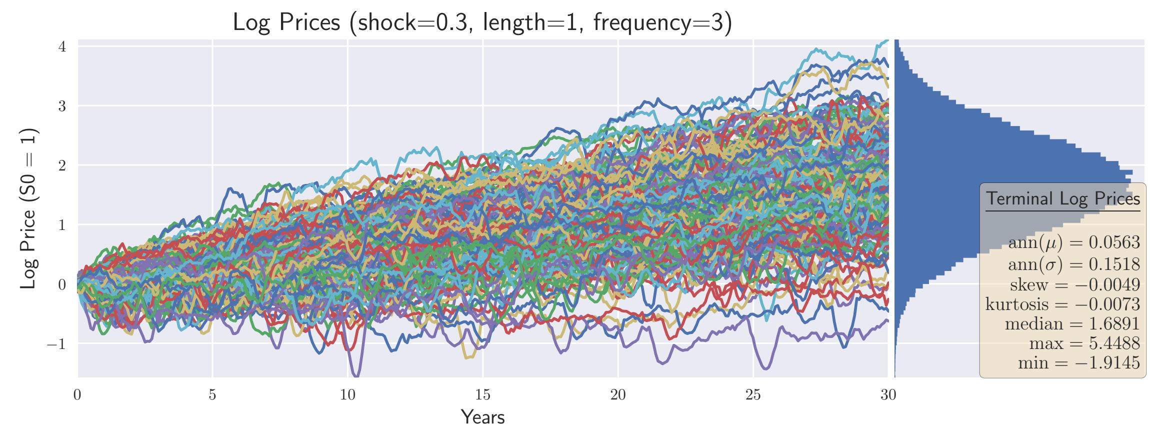

Log Prices (shock=0.5, length=1, frequency=7)
Log Price (S0 = 1)
Years
1.0
0.5
0.0
−0.5
−1.0
0
1
2
3
4
5
Terminal Log Prices
ann(μ) = 0.0468
ann(σ) = 0.1708
skew = −0.3634
kurtosis = 0.5869
median = 0.2513
max = 1.5942
min = −1.9760

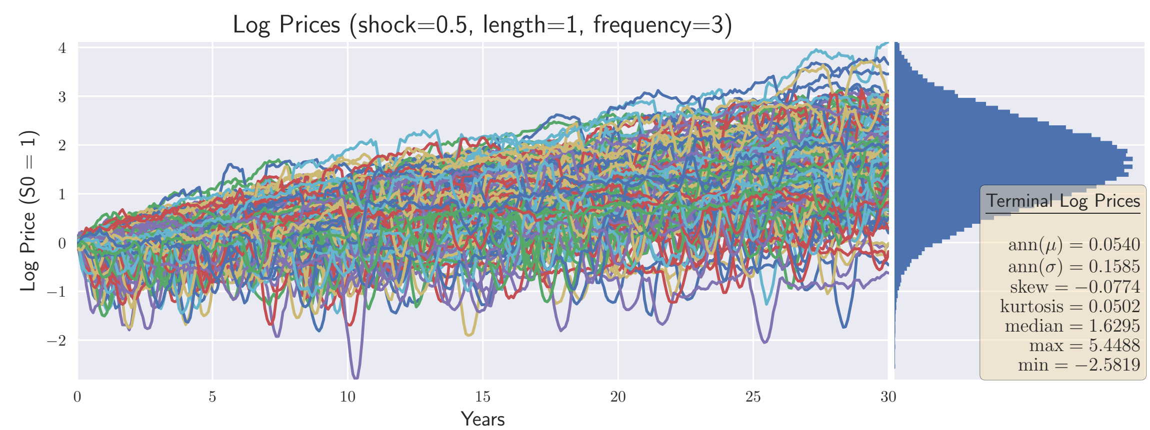

Log Prices (shock=0.7, length=1, frequency=7)
Log Price (S0 = 1)
Years
Terminal Log Prices
ann(μ) = 0.0381
ann(σ) = 0.2066
skew = −1.0712
kurtosis = 2.5027
median = 0.2434
max = 1.5942
min = −3.2577

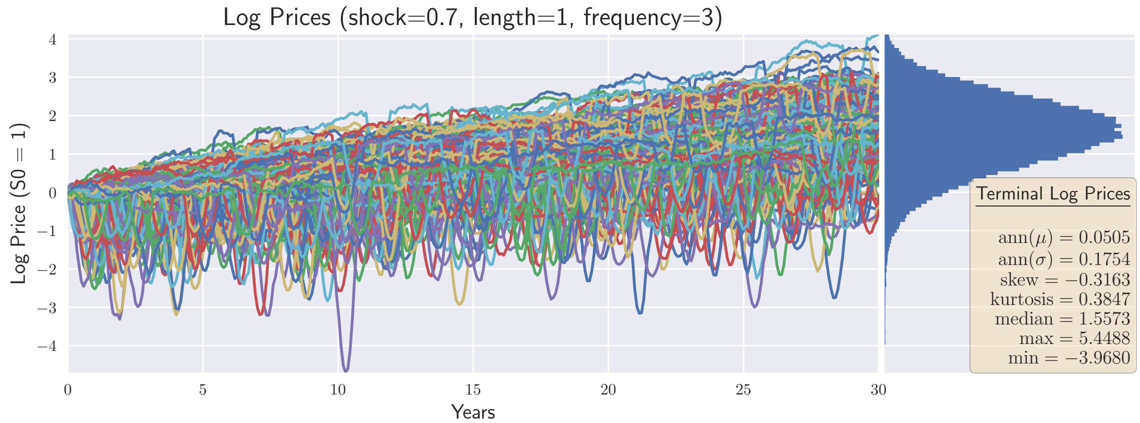

Log Prices (shock=0.3, length=2, frequency=7)
Log Price (S0 = 1)
Years
Terminal Log Prices
ann(μ) = 0.0457
ann(σ) = 0.1620
skew = −0.1204
kurtosis = 0.0842
median = 0.2344
max = 1.5942
min = −1.4028

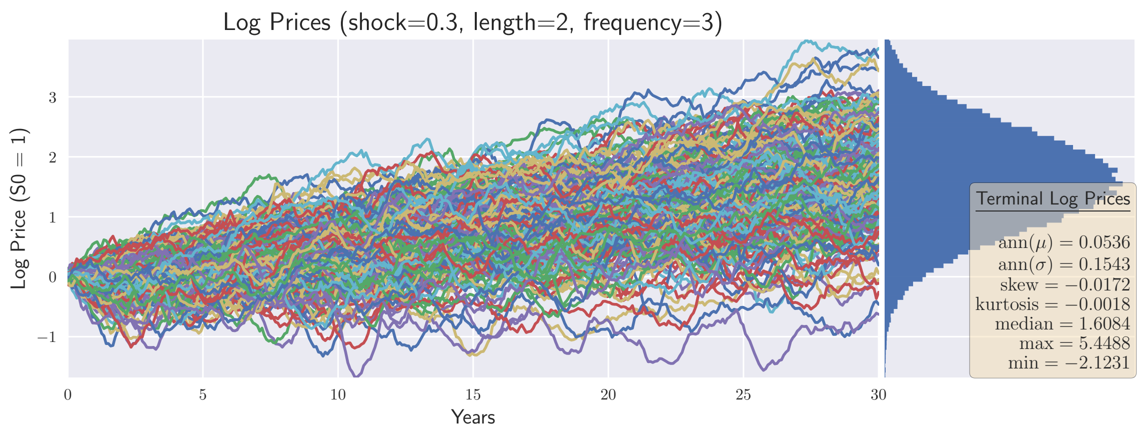

Log Prices (shock=0.5, length=2, frequency=7)
Log Price (S0 = 1)
Years
Terminal Log Prices
ann(μ) = 0.0334
ann(σ) = 0.1917
skew = −0.5439
kurtosis = 0.7579
median = 0.2017
max = 1.5942
min = −2.4662

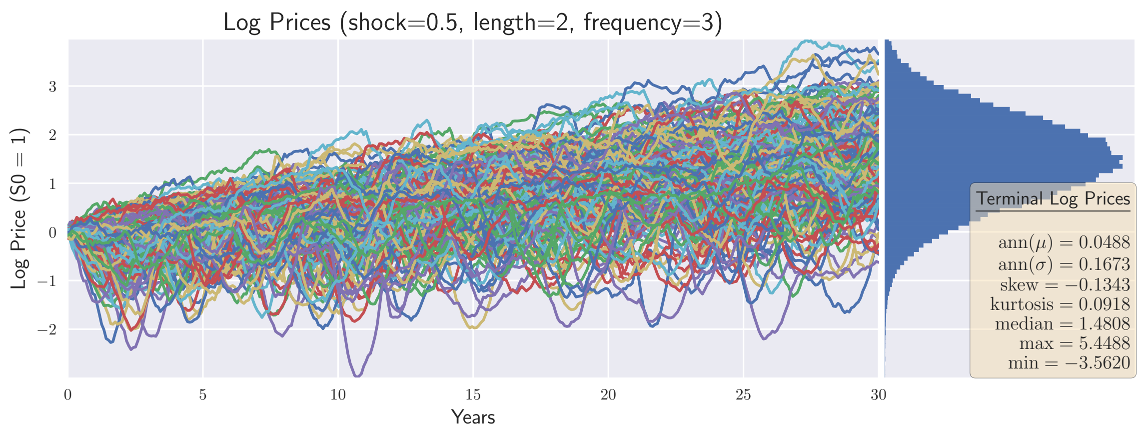

Log Prices (shock=0.7, length=2, frequency=7)
Log Price (S0 = 1)
Years
Terminal Log Prices
ann(μ) = 0.0147
ann(σ) = 0.2559
skew = −1.2035
kurtosis = 2.2722
median = 0.1826
max = 1.5942
min = −4.3511


Log Prices (shock=0.3, length=1, frequency=3)
Log Price (S0 = 1)
Years
Terminal Log Prices
ann(μ) = 0.0563
ann(σ) = 0.1518
skew = −0.0049
kurtosis = −0.0073
median = 1.6891
max = 5.4488
min = −1.9145

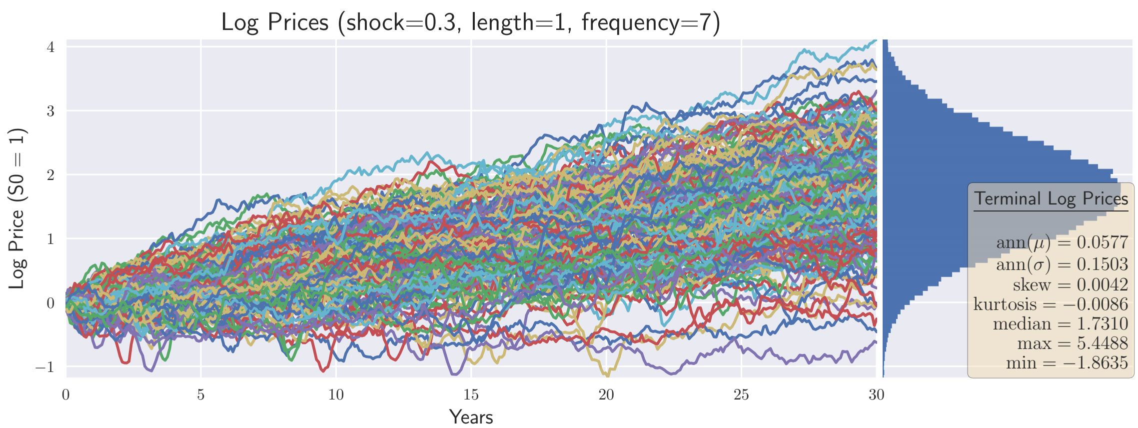

Log Prices (shock=0.5, length=1, frequency=3)
Log Price (S0 = 1)
Years
Terminal Log Prices
ann(μ) = 0.0540
ann(σ) = 0.1585
skew = −0.0774
kurtosis = 0.0502
median = 1.6295
max = 5.4488
min = −2.5819

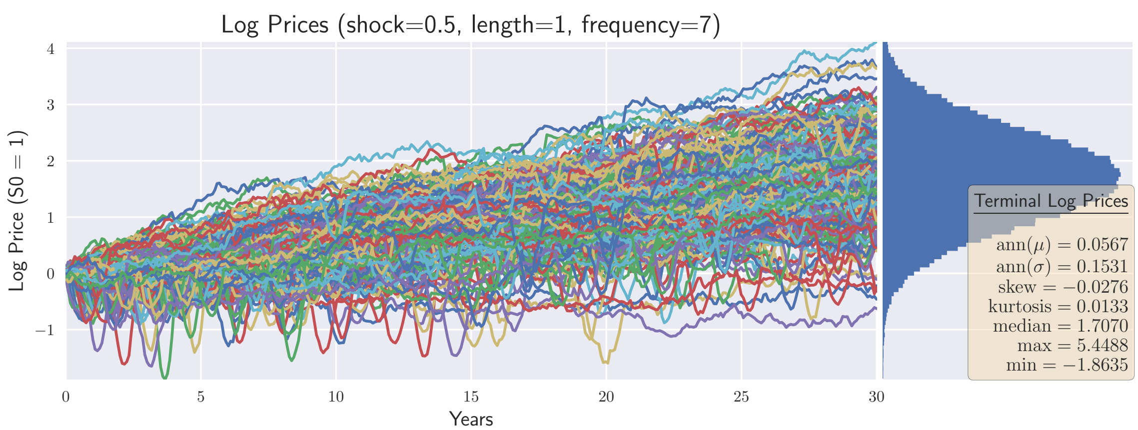

Log Prices (shock=0.7, length=1, frequency=3)
Log Price (S0 = 1)
Years
Terminal Log Prices
ann(μ) = 0.0505
ann(σ) = 0.1754
skew = −0.3163
kurtosis = 0.3847
median = 1.5573
max = 5.4488
min = −3.9680

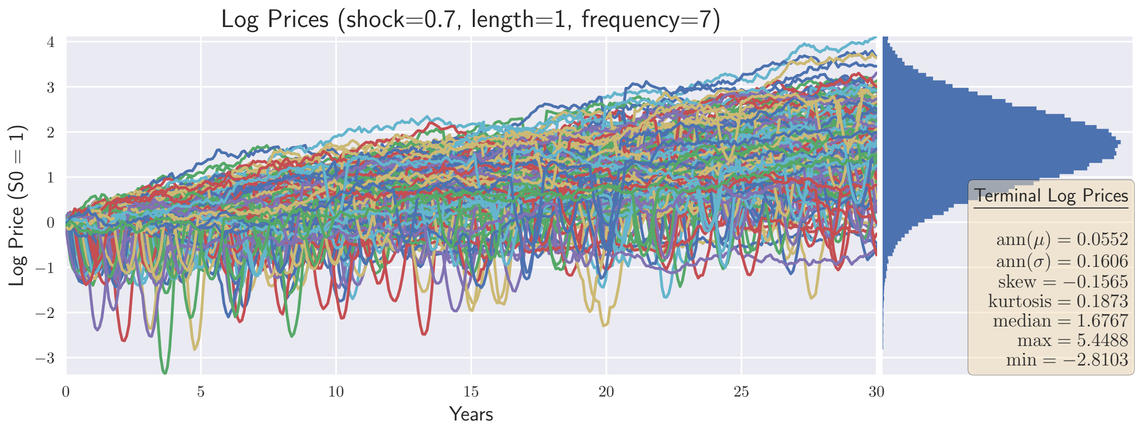

Log Prices (shock=0.3, length=2, frequency=3)
Log Price (S0 = 1)
Years
Terminal Log Prices
ann(μ) = 0.0536
ann(σ) = 0.1543
skew = −0.0172
kurtosis = −0.0018
median = 1.6084
max = 5.4488
min = −2.1231

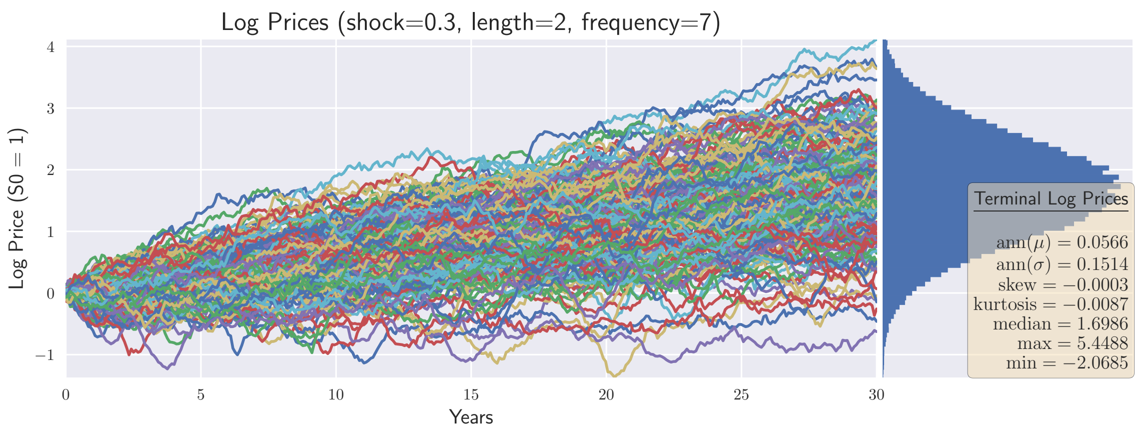

Log Prices (shock=0.5, length=2, frequency=3)
Log Price (S0 = 1)
Years
Terminal Log Prices
ann(μ) = 0.0488
ann(σ) = 0.1673
skew = −0.1343
kurtosis = 0.0918
median = 1.4808
max = 5.4488
min = −3.5620

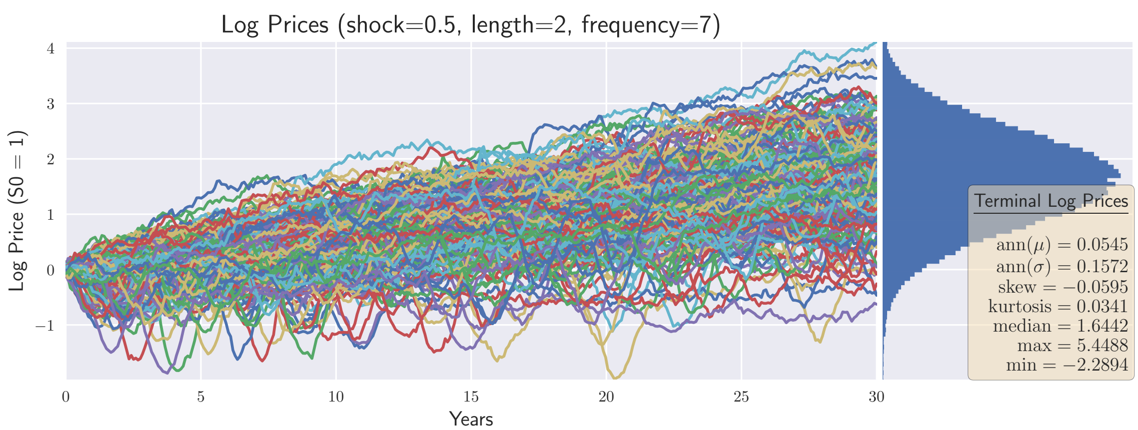

Log Prices (shock=0.7, length=2, frequency=3)
Log Price (S0 = 1)
Years
Terminal Log Prices
ann(μ) = 0.0414
ann(σ) = 0.1988
skew = −0.4305
kurtosis = 0.4539
median = 1.3201
max = 5.4488
min = −6.3122

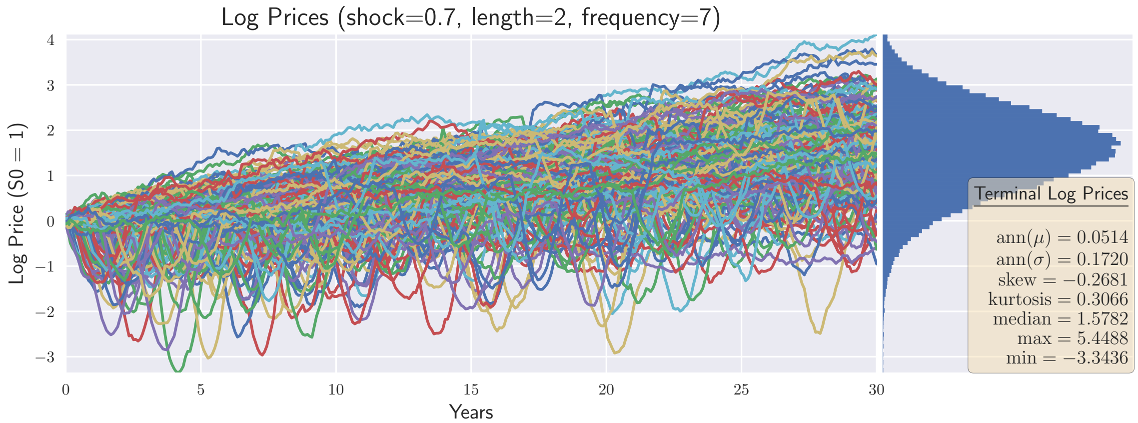

Log Prices (shock=0.3, length=1, frequency=7)
Log Price (S0 = 1)
Years
Terminal Log Prices
ann(μ) = 0.0577
ann(σ) = 0.1503
skew = 0.0042
kurtosis = −0.0086
median = 1.7310
max = 5.4488
min = −1.8635


Log Prices (shock=0.5, length=1, frequency=7)
Log Price (S0 = 1)
Years
Terminal Log Prices
ann(μ) = 0.0567
ann(σ) = 0.1531
skew = −0.0276
kurtosis = 0.0133
median = 1.7070
max = 5.4488
min = −1.8635


Log Prices (shock=0.7, length=1, frequency=7)
Log Price (S0 = 1)
Years
Terminal Log Prices
ann(μ) = 0.0552
ann(σ) = 0.1606
skew = −0.1565
kurtosis = 0.1873
median = 1.6767
max = 5.4488
min = −2.8103


Log Prices (shock=0.3, length=2, frequency=7)
Log Price (S0 = 1)
Years
Terminal Log Prices
ann(μ) = 0.0566
ann(σ) = 0.1514
skew = −0.0003
kurtosis = −0.0087
median = 1.6986
max = 5.4488
min = −2.0685


Log Prices (shock=0.5, length=2, frequency=7)
Log Price (S0 = 1)
Years
Terminal Log Prices
ann(μ) = 0.0545
ann(σ) = 0.1572
skew = −0.0595
kurtosis = 0.0341
median = 1.6442
max = 5.4488
min = −2.2894



Log Prices (shock=0.7, length=2, frequency=7)
Log Price (S0 = 1)
Years
Terminal Log Prices
ann(μ) = 0.0514
ann(σ) = 0.1720
skew = −0.2681
kurtosis = 0.3066
median = 1.5782
max = 5.4488
min = −3.3436

7.2 Run Up Plots and Regressions

Here, we try to model the shock parameters (shock amplitude, shock length, time between shocks) as a function of the moments of time periods leading up to the shock itself. We take the first three moments from 30, 90, 180, 360, 540, and 720 days before the shock and regress against each of the above parameters. For more visibility, we also plot the features below along with their correlations.

Note: The number at the end of each independent variable refers to the number of days before the shock that the statistic was computed at.

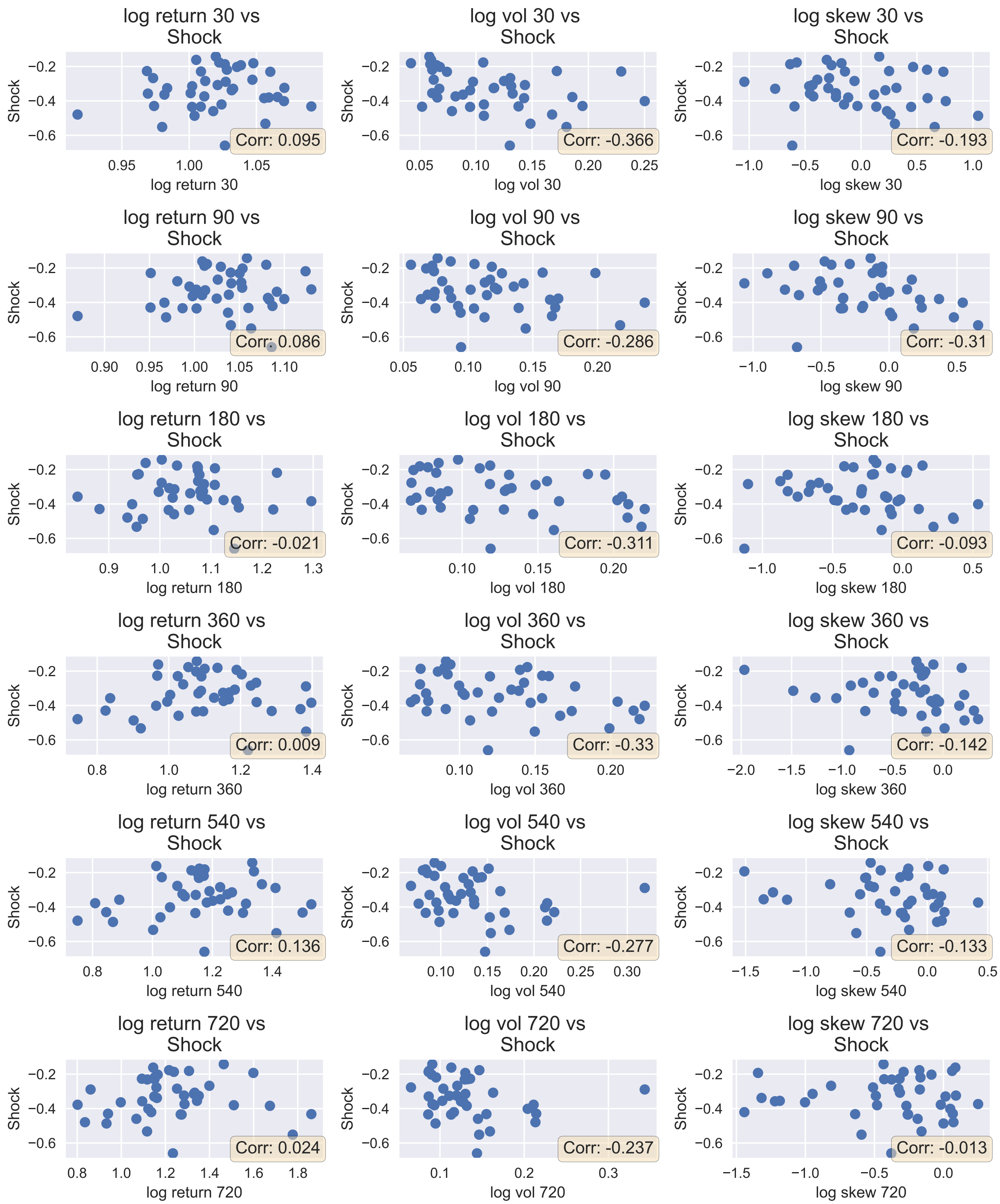
7.2.1 Run Up vs Shock

Dep. Variable:	Shock	**R-squared:**	0.478
Model:	OLS	**Adj. R-squared:**	0.030
Method:	Least Squares	**F-statistic:**	13.65
Date:	Fri, 10 Sep 2021	**Prob (F-statistic):**	8.19e-08
Time:	10:45:11	**Log-Likelihood:**	42.070
No. Observations:	40	**AIC:**	-46.14
Df Residuals:	21	**BIC:**	-14.05
Df Model:	18		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	-1.2095	0.440	-2.747	0.012	-2.125	-0.294
log_return_30	0.4122	0.690	0.597	0.557	-1.022	1.847
log_vol_30	0.2266	1.057	0.214	0.832	-1.972	2.425
log_skew_30	-0.0673	0.045	-1.505	0.147	-0.160	0.026
log_return_90	0.7239	0.717	1.010	0.324	-0.766	2.214
log_vol_90	-0.1105	1.987	-0.056	0.956	-4.243	4.022
log_skew_90	-0.1362	0.073	-1.874	0.075	-0.287	0.015
log_return_180	-0.1969	0.272	-0.723	0.478	-0.763	0.369
log_vol_180	0.9373	1.346	0.696	0.494	-1.862	3.736
log_skew_180	0.1365	0.089	1.537	0.139	-0.048	0.321
log_return_360	-0.3440	0.330	-1.044	0.308	-1.029	0.341
log_vol_360	0.7955	1.572	0.506	0.618	-2.474	4.065
log_skew_360	0.0890	0.126	0.703	0.489	-0.174	0.352
log_return_540	0.6251	0.280	2.234	0.036	0.043	1.207
log_vol_540	-8.3815	3.627	-2.311	0.031	-15.925	-0.838
log_skew_540	-0.2390	0.134	-1.788	0.088	-0.517	0.039
log_return_720	-0.2835	0.162	-1.748	0.095	-0.621	0.054
log_vol_720	6.1793	2.872	2.152	0.043	0.207	12.151
log_skew_720	0.1382	0.059	2.337	0.029	0.015	0.261

Omnibus:	0.275	**Durbin-Watson:**	1.531
Prob(Omnibus):	0.871	**Jarque-Bera (JB):**	0.202
Skew:	-0.160	**Prob(JB):**	0.904
Kurtosis:	2.861	**Cond. No.**	1.09e+03

Notes:
[1] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using 1 lags and without small sample correction
[2] The condition number is large, 1.09e+03. This might indicate that there are strong multicollinearity or other numerical problems.


log return 30 vs Shock
Corr: 0.095
log return 30
log vol 30 vs Shock
Corr: -0.366
log vol 30
log skew 30 vs Shock
Corr: -0.193
log skew 30
log return 90 vs Shock
Corr: 0.086
log return 90
log vol 90 vs Shock
Corr: -0.286
log vol 90
log skew 90 vs Shock
Corr: -0.31
log skew 90
log return 180 vs Shock
Corr: -0.021
log return 180
log vol 180 vs Shock
Corr: -0.311
log vol 180
log skew 180 vs Shock
Corr: -0.093
log skew 180
log return 360 vs Shock
Corr: 0.009
log return 360
log vol 360 vs Shock
Corr: -0.33
log vol 360
log skew 360 vs Shock
Corr: -0.142
log skew 360
log return 540 vs Shock
Corr: 0.136
log return 540
log vol 540 vs Shock
Corr: -0.277
log vol 540
log skew 540 vs Shock
Corr: -0.133
log skew 540
log return 720 vs Shock
Corr: 0.024
log return 720
log vol 720 vs Shock
Corr: -0.237
log vol 720
log skew 720 vs Shock
Corr: -0.013
log skew 720
Shock

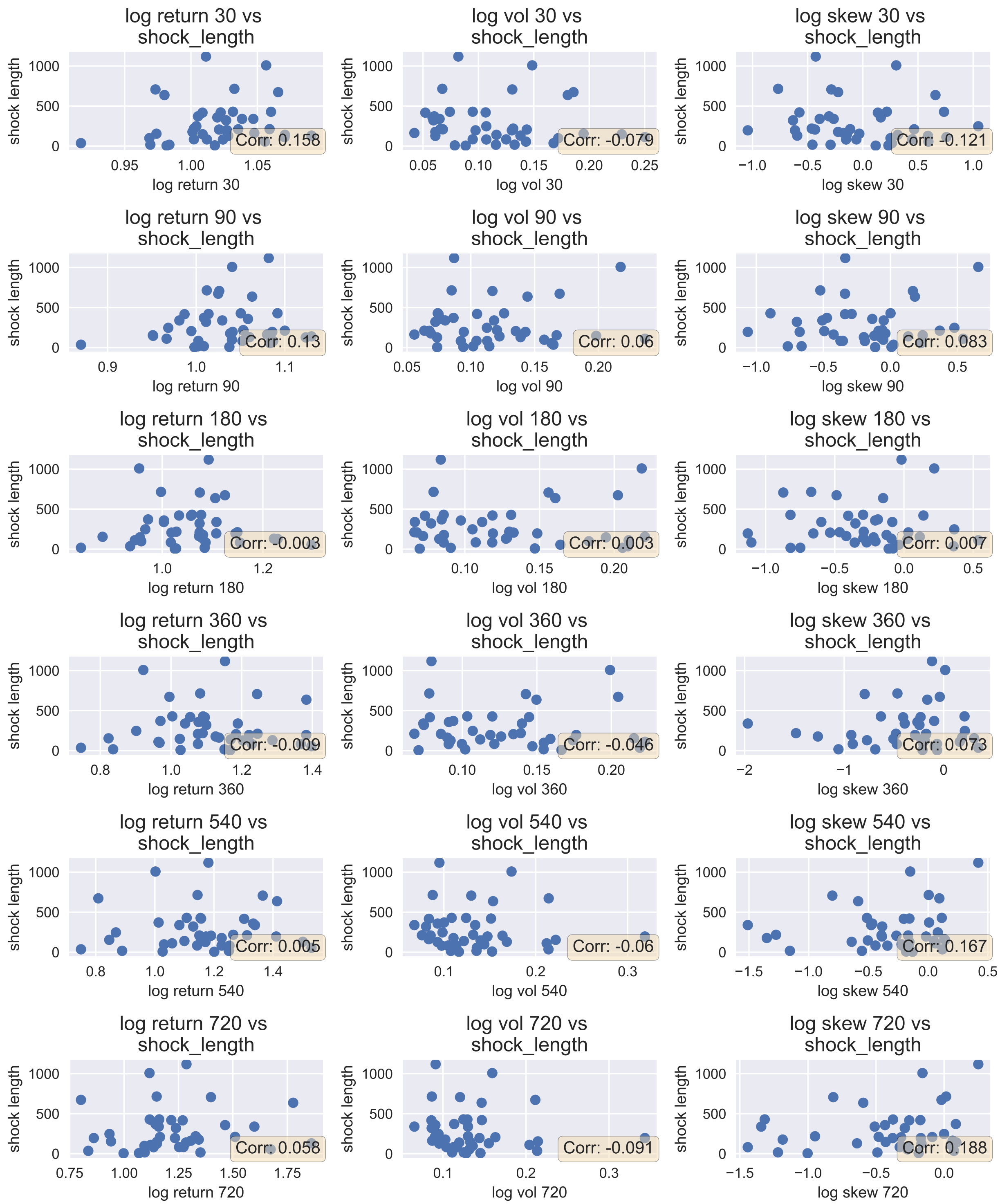
7.2.2 Run Up vs shock length

Dep. Variable:	shock_length	**R-squared:**	0.402
Model:	OLS	**Adj. R-squared:**	-0.111
Method:	Least Squares	**F-statistic:**	4.778
Date:	Fri, 10 Sep 2021	**Prob (F-statistic):**	0.000452
Time:	10:45:14	**Log-Likelihood:**	-269.08
No. Observations:	40	**AIC:**	576.2
Df Residuals:	21	**BIC:**	608.2
Df Model:	18		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	1257.6305	1517.638	0.829	0.417	-1898.471	4413.732
log_return_30	-2391.2059	1921.591	-1.244	0.227	-6387.372	1604.960
log_vol_30	-5513.6034	1251.091	-4.407	0.000	-8115.390	-2911.817
log_skew_30	-258.4412	104.589	-2.471	0.022	-475.945	-40.937
log_return_90	1696.9486	1298.102	1.307	0.205	-1002.602	4396.499
log_vol_90	1.014e+04	2358.022	4.302	0.000	5240.671	1.5e+04
log_skew_90	226.2487	135.354	1.672	0.109	-55.236	507.733
log_return_180	313.7184	1042.050	0.301	0.766	-1853.343	2480.780
log_vol_180	-2720.9735	3064.989	-0.888	0.385	-9094.966	3653.019
log_skew_180	-205.9186	194.805	-1.057	0.302	-611.037	199.200
log_return_360	-352.4511	772.052	-0.457	0.653	-1958.021	1253.119
log_vol_360	-3458.9894	3984.146	-0.868	0.395	-1.17e+04	4826.495
log_skew_360	101.1268	222.837	0.454	0.655	-362.287	564.541
log_return_540	-366.6337	670.226	-0.547	0.590	-1760.444	1027.177
log_vol_540	1.669e+04	5486.136	3.042	0.006	5281.212	2.81e+04
log_skew_540	252.2295	318.213	0.793	0.437	-409.531	913.990
log_return_720	217.2719	397.769	0.546	0.591	-609.935	1044.479
log_vol_720	-1.524e+04	4349.605	-3.504	0.002	-2.43e+04	-6197.136
log_skew_720	-150.6652	109.861	-1.371	0.185	-379.134	77.804

Omnibus:	4.541	**Durbin-Watson:**	1.714
Prob(Omnibus):	0.103	**Jarque-Bera (JB):**	3.558
Skew:	0.720	**Prob(JB):**	0.169
Kurtosis:	3.245	**Cond. No.**	1.09e+03

Notes:
[1] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using 1 lags and without small sample correction
[2] The condition number is large, 1.09e+03. This might indicate that there are strong multicollinearity or other numerical problems.



log return 30 vs shock_length
Corr: 0.158
log return 30
log vol 30 vs shock_length
Corr: -0.079
log vol 30
log skew 30 vs shock_length
Corr: -0.121
log skew 30
log return 90 vs shock_length
Corr: 0.13
log return 90
log vol 90 vs shock_length
Corr: 0.06
log vol 90
log skew 90 vs shock_length
Corr: 0.083
log skew 90
log return 180 vs shock_length
Corr: -0.003
log return 180
log vol 180 vs shock_length
Corr: 0.003
log vol 180
log skew 180 vs shock_length
Corr: 0.007
log skew 180
log return 360 vs shock_length
Corr: -0.009
log return 360
log vol 360 vs shock_length
Corr: -0.046
log vol 360
log skew 360 vs shock_length
Corr: 0.073
log skew 360
log return 540 vs shock_length
Corr: 0.005
log return 540
log vol 540 vs shock_length
Corr: -0.06
log vol 540
log skew 540 vs shock_length
Corr: 0.167
log skew 540
log return 720 vs shock_length
Corr: 0.058
log return 720
log vol 720 vs shock_length
Corr: -0.091
log vol 720
log skew 720 vs shock_length
Corr: 0.188
log skew 720
shock length

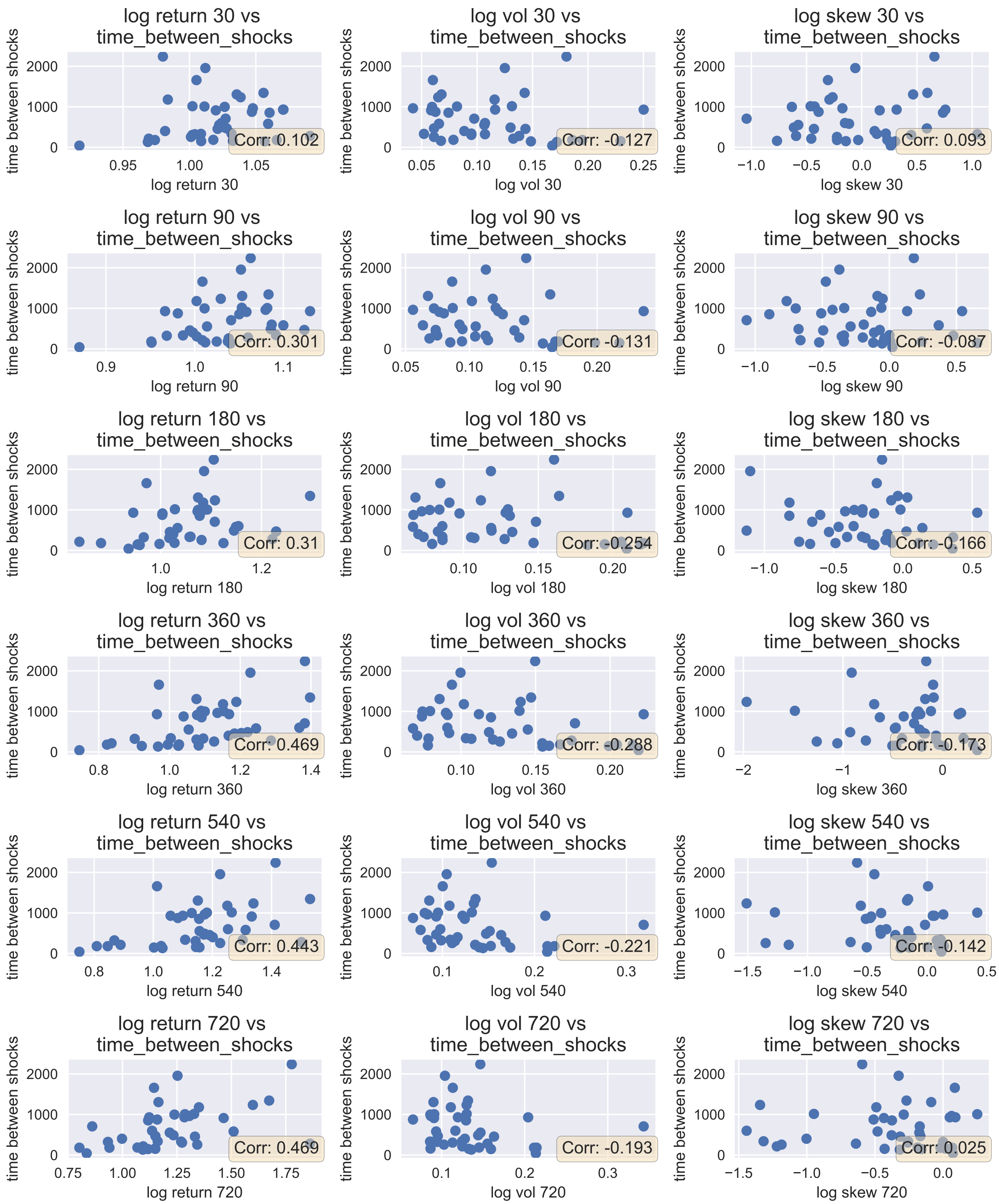
7.2.3 Run Up vs time between shocks

Dep. Variable:	time_between_shocks	**R-squared:**	0.574
Model:	OLS	**Adj. R-squared:**	0.191
Method:	Least Squares	**F-statistic:**	12.45
Date:	Fri, 10 Sep 2021	**Prob (F-statistic):**	3.21e-07
Time:	10:45:16	**Log-Likelihood:**	-282.76
No. Observations:	39	**AIC:**	603.5
Df Residuals:	20	**BIC:**	635.1
Df Model:	18		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	672.8921	2158.207	0.312	0.758	-3829.049	5174.833
log_return_30	-2250.6451	2871.272	-0.784	0.442	-8240.013	3738.723
log_vol_30	1617.2515	2409.071	0.671	0.510	-3407.983	6642.486
log_skew_30	358.2887	159.744	2.243	0.036	25.068	691.509
log_return_90	2032.6567	1982.146	1.025	0.317	-2102.027	6167.340
log_vol_90	3635.9719	4582.403	0.793	0.437	-5922.754	1.32e+04
log_skew_90	-441.4086	409.574	-1.078	0.294	-1295.765	412.948
log_return_180	-2570.7583	1261.248	-2.038	0.055	-5201.676	60.160
log_vol_180	-2550.2883	5686.007	-0.449	0.659	-1.44e+04	9310.515
log_skew_180	269.9614	431.614	0.625	0.539	-630.370	1170.292
log_return_360	2585.6531	1166.255	2.217	0.038	152.889	5018.417
log_vol_360	325.2984	6448.690	0.050	0.960	-1.31e+04	1.38e+04
log_skew_360	-315.0153	337.969	-0.932	0.362	-1020.006	389.976
log_return_540	-1905.1180	910.000	-2.094	0.049	-3803.345	-6.891
log_vol_540	-3.383e+04	1.6e+04	-2.119	0.047	-6.71e+04	-521.923
log_skew_540	-296.8366	436.653	-0.680	0.504	-1207.678	614.005
log_return_720	1769.8558	596.049	2.969	0.008	526.519	3013.193
log_vol_720	3.258e+04	1.46e+04	2.234	0.037	2163.905	6.3e+04
log_skew_720	651.9104	239.118	2.726	0.013	153.120	1150.701

Omnibus:	0.210	**Durbin-Watson:**	1.746
Prob(Omnibus):	0.901	**Jarque-Bera (JB):**	0.054
Skew:	0.087	**Prob(JB):**	0.973
Kurtosis:	2.946	**Cond. No.**	1.07e+03

Notes:
[1] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using 1 lags and without small sample correction
[2] The condition number is large, 1.07e+03. This might indicate that there are strong multicollinearity or other numerical problems.



Note: In order to reduce serial correlation, regression results use the Newey-West covariance matrix.