Mechanical & Industrial Engineering
UNIVERSITY OF TORONTO

Data-driven distributionally robust risk parity portfolio optimization

2020 INFORMS Annual Meeting

November 12, 2020

Giorgio Costa and Roy H. Kwon

Department of Mechanical and Industrial Engineering,

University of Toronto

What is this presentation about?

- We will discuss optimal asset allocation strategies.
- In particular, we discuss a strategy known as risk parity.
- We rely on estimated inputs, making us susceptible to estimation error.
- We introduce distributional robustness to mitigate the impact of uncertainty.
- The user is able to define their preferred statistical measure for ‘robustness’.
- We present a novel algorithm to solve the resulting problem.

Optimal asset allocation

- **Mean–variance optimization (MVO)**: Construct optimal portfolios as a trade-off between risk and expected return.

- However, MVO is susceptible to estimation errors in both parametric measures of return and risk.
    - In particular the measure of return is considered to be quite unreliable.

Optimal asset allocation

- **Mean–variance optimization (MVO)**: Construct optimal portfolios as a trade-off between risk and expected return.
- However, MVO is susceptible to estimation errors in both parametric measures of return and risk.
  - In particular the measure of return is considered to be quite unreliable.
- Alternative: **Risk parity**
  - Equalizes the risk contributions of each asset
  - Does not require a return measure.
  - However, it is still susceptible to estimation errors in the risk measure.

Distributional robustness

Introduce robustness at the parameter estimation step

- We use a discrete probability distribution to model the 'weights' associated with each scenario in our dataset.
  - E.g., consider a simple estimate of the expected value and variance of a <u>discrete</u> random variable:

$$\mu = \sum_{i=1}^{n} p_i \cdot x_i$$

$$\sigma^2 = \sum_{i=1}^{n} p_i \cdot (x_i - \mu)^2$$

Distributional robustness

Introduce robustness at the parameter estimation step

- We use a discrete probability distribution to model the 'weights' associated with each scenario in our dataset.
  - E.g., consider a simple estimate of the expected value and variance of a <u>discrete</u> random variable:

$$\mu = \sum_{i=1}^{n} p_i \cdot x_i$$
$$\sigma^2 = \sum_{i=1}^{n} p_i \cdot (x_i - \mu)^2$$

  - If we have raw data, we assume scenarios are equally likely, $p_i = 1/n$
  - What if we break away from this assumption?

Introduce robustness at the parameter estimation step

- This approach avoids making assumptions or imposing a structure on our data and the underlying distribution.

- Our financial <u>risk measure</u> is the asset covariance matrix.

Distributional robustness

Introduce robustness at the parameter estimation step

- This approach avoids making assumptions or imposing a structure on our data and the underlying distribution.
- Our financial risk measure is the asset covariance matrix.
- The risk parity problem seeks to equalize the asset risk contributions.
  - We minimize the objective by changing our asset weights to attain risk parity.
  - We can maximize the objective by using the discrete probabilities as adversarial variables

Distributional robustness

Introduce robustness at the parameter estimation step

- This approach avoids making assumptions or imposing a structure on our data and the underlying distribution.

- Our financial risk measure is the asset covariance matrix.

- The risk parity problem seeks to equalize the asset risk contributions.
  - We minimize the objective by changing our asset weights to attain risk parity.
  - We can maximize the objective by using the discrete probabilities as adversarial variables

- We have a minimax problem.

Minimax problem

- Specifically, we have a constrained convex–concave minimax problem.
- The asset weights are constrained by the set of feasible portfolios.
- The discrete probabilities are constrained by

Minimax problem

- Specifically, we have a constrained convex–concave minimax problem.
- The asset weights are constrained by the set of feasible portfolios.
- The discrete probabilities are constrained by
  - The axioms of probability
  - Some statistical distance from our nominal distribution
  - Naturally, the nominal is the ‘equally likely’ distribution

Distributionally robust risk parity

Statistical distance measures

- As long as the statistical distance measure is both convex and bounded, the user can choose their preferred measure.
- We exemplify this through the following three measures
  - Jensen–Shannon (JS) divergence
  - Hellinger (H) distance
  - Total variation (TV) distance
- Set the distance proportional to an investor's confidence level.

Projected gradient descent–ascent algorithm

- We can solve our minimax problem through a projected gradient descent–ascent (PGDA) algorithm.
  - Iteratively descend in the asset weight space, ascend in the probability space.
  - A projection (or similar approach) is necessary due to the constraints.
  - We must define several parameters in both directions.
  - Problem: slow convergence and vanishing gradients in both directions.

Projected gradient ascent with sequential convex programming

- We propose a projected gradient ascent algorithm grounded on sequential convex programming (PGA–SCP).
  - Iteratively ascend in the probability space.
  - Solve for the corresponding risk parity portfolio in the asset weight space after every iteration.
  - In other words, we repeatedly solve the convex risk parity problem.
  - The resulting algorithm is much faster and stable

Comparison between PGDA and PGA–SCP

- Data:
  - Synthetic asset return data
  - We have 200 assets ($n = 200$) and 5,000 observations ($T = 5,000$)
  - The investor confidence is set to 30% ($\delta = 0.3$)

Numerical experiments

Comparison between PGDA and PGA–SCP

	$n = 200$, $T = 5,000$, $\delta = 0.3$					
	JS		Hellinger		TV	
	PGDA	PGA–SCP	PGDA	PGA–SCP	PGDA	PGA–SCP
Time (s)	**157**	**79.9**	569	99.6	235	128
Iterations	59	33	132	28	48	37
Var. ($\times 10^4$)	3.97	9.06	3.67	9.98	0.71	12.9

- The runtime of the proposed PGA–SCP is considerably lower.

Comparison between PGDA and PGA–SCP

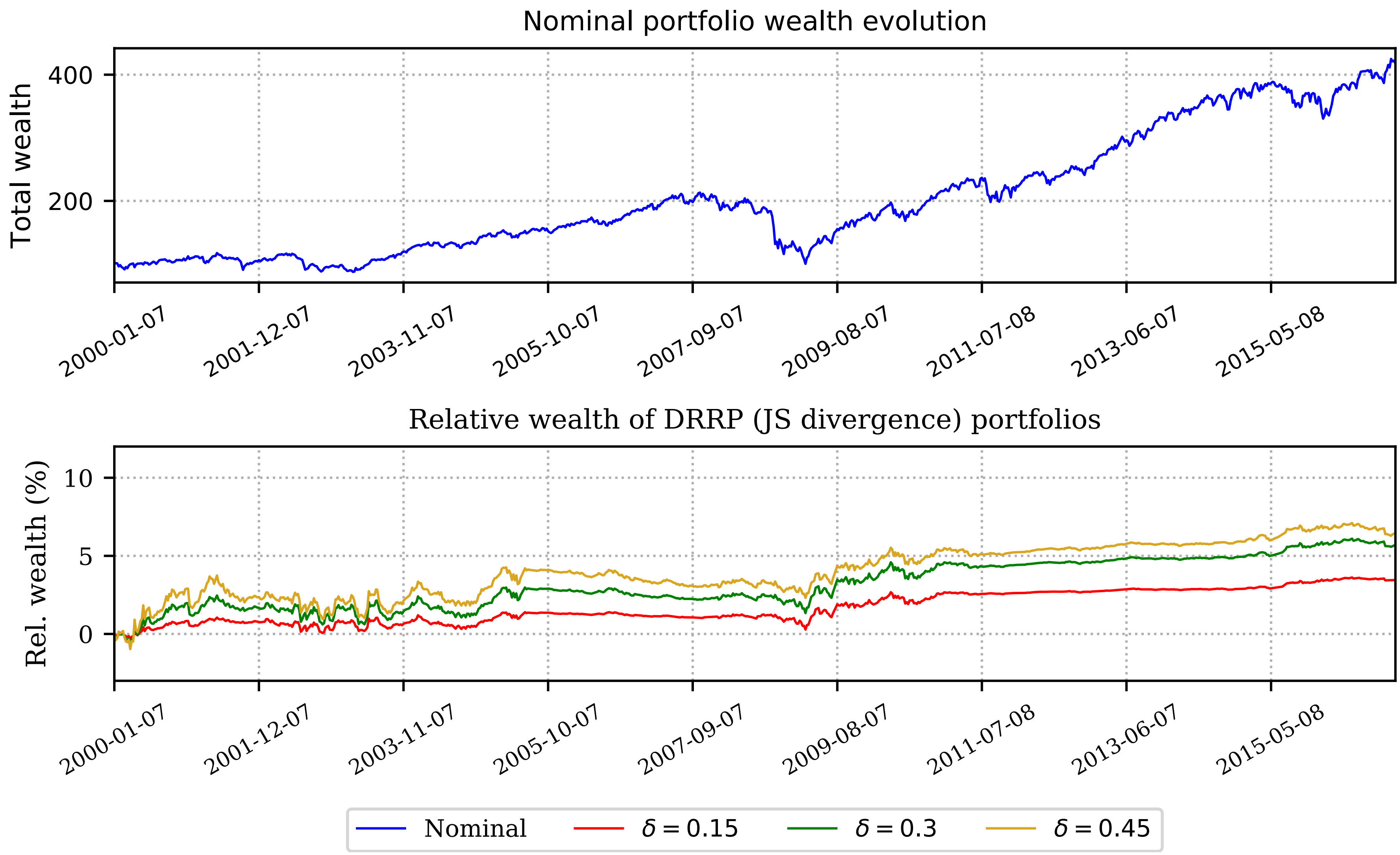
	$n = 200, T = 5,000, \delta = 0.3$					
	JS		Hellinger		TV	
	PGDA	PGA–SCP	PGDA	PGA–SCP	PGDA	PGA–SCP
Time (s)	157	79.9	569	99.6	235	128
Iterations	59	33	132	28	48	37
Var. $(\times 10^4)$	**3.97**	**9.06**	3.67	9.98	0.71	12.9

- The PGDA algorithm converges to a sub-optimal solution.

Out-of-sample experimental setup

- Data:
  - 30 industry portfolios (from Kenneth French's data library).
  - Weekly asset return data from 1998 until 2016.
  - Two years worth of data for calibration.

Numerical experiments



Numerical experiments

Summary of financial performance between 2000–2016

	Nom.	JS			Hellinger			TV		
$\delta =$		0.15	0.3	0.45	0.15	0.3	0.45	0.15	0.3	0.45
Ann. Ex. Return (%)	6.64	6.84	6.96	7.01	6.85	6.97	7.02	6.95	6.98	7.01
Ann. Volatility (%)	17.0	17.2	17.2	17.2	17.2	17.2	17.2	17.2	17.2	17.2
Sharpe Ratio (%)	**39.0**	39.8	**40.5**	40.7	39.9	**40.5**	40.8	40.4	**40.6**	40.7
Avg. Turnover (%)	10.0	12.6	15.1	16.7	12.8	15.2	16.8	16.0	17.2	18.0

- Robust portfolios have a higher Sharpe ratio than the nominal.
- Performance is similar between robust portfolios with the same δ.

Summary of financial performance between 2000–2016

	Nom.	JS			Hellinger			TV		
$\delta =$		0.15	0.3	0.45	0.15	0.3	0.45	0.15	0.3	0.45
Ann. Ex. Return (%)	6.64	6.84	6.96	7.01	6.85	6.97	7.02	6.95	6.98	7.01
Ann. Volatility (%)	17.0	17.2	17.2	17.2	17.2	17.2	17.2	17.2	17.2	17.2
Sharpe Ratio (%)	39.0	**39.8**	**40.5**	**40.7**	39.9	40.5	40.8	40.4	40.6	40.7
Avg. Turnover (%)	10.0	12.6	15.1	16.7	12.8	15.2	16.8	16.0	17.2	18.0

- The Sharpe ratio increases with robustness

Numerical experiments

Summary of financial performance between 2000–2016

	Nom.	JS			Hellinger			TV		
$\delta =$		0.15	0.3	0.45	0.15	0.3	0.45	0.15	0.3	0.45
Ann. Ex. Return (%)	6.64	6.84	6.96	7.01	6.85	6.97	7.02	6.95	6.98	7.01
Ann. Volatility (%)	17.0	17.2	17.2	17.2	17.2	17.2	17.2	17.2	17.2	17.2
Sharpe Ratio (%)	39.0	39.8	40.5	40.7	39.9	40.5	40.8	40.4	40.6	40.7
Avg. Turnover (%)	**10.0**	**12.6**	**15.1**	**16.7**	12.8	15.2	16.8	16.0	17.2	18.0

- Not surprisingly, turnover also increases with robustness

Distributionally robust risk parity

Findings and contribution

- Introduce distributional robustness to risk parity through a discrete probability distribution.

Distributionally robust risk parity

Findings and contribution

- Introduce distributional robustness to risk parity through a discrete probability distribution.
- The ambiguity set is determined by the user.
  - The user can choose their preferred statistical distance measure.
  - The level of robustness is set by the confidence level.

Distributionally robust risk parity

Findings and contribution

- Introduce distributional robustness to risk parity through a discrete probability distribution.
- The ambiguity set is determined by the user.
  - The user can choose their preferred statistical distance measure.
  - The level of robustness is set by the confidence level.
- Introduce an algorithm for constrained convex–concave minimax problems.
  - The algorithm can tackle other portfolio selection problems.
  - It may also generalize to other similar minimax problems.

Distributionally robust risk parity

Findings and contribution

- Introduce distributional robustness to risk parity through a discrete probability distribution.
- The ambiguity set is determined by the user.
  - The user can choose their preferred statistical distance measure.
  - The level of robustness is set by the confidence level.
- Introduce an algorithm for constrained convex–concave minimax problems.
  - The algorithm can tackle other portfolio selection problems.
  - It may also generalize to other similar minimax problems.
- For more information:

  Costa, G. and Kwon, R. H. (2020). Data-driven distributionally robust risk parity portfolio optimization. *Available at SSRN 3709680*

References

[1] Costa, G. and Kwon, R. H. (2020). Data-driven distributionally robust risk parity portfolio optimization. *Available at SSRN 3709680*.