Redistribution of Equity Returns After The Minimum Wage Policy

A Thesis Submitted in Partial Fulfillment of
Bachelor of Science in (Honor) Statistics at the University of Michigan

Mingxuan Ge

Supervisor: Prof. Gunsilius

Co-adviser: Prof. Ionides

Final Version: April 19, 2022

Abstract

This paper identifies short run causal effects of labor-market risk on financial markets by analyzing the impact on an unexpected minimum wage policy change had on the equity returns of companies in London Stock Exchange Markets on July 8, 2015. This announcement was a drastic change in the UK minimum wage system and underlying common intuitions predict that the increase of labor costs will reduce firms' profitability, especially in labor-intensive industries, and destroy investors' confidences on future performances of certain firms. A Subsequent increase of risks and uncertainty of investments would affect the corresponding equity returns, at least in the short run. Existing analyses mainly focus on Granger causality but do not provide estimates of the causal effect of the policy change. For the causal analysis, we apply the recently introduced distributional synthetic controls (DSCs). This allows us to replicate the complete distributions of distributions of equity returns and study the whole counterfactual process, rather than superficial effects on discounted future cash flows or risk-premium as a form of normalized expected returns. Current econometric methods lack the ability to study causal effects in a scenario like ours where we care about entire return distributions. We find that the minimum wage hike actually caused an increase in returns for most stocks, via the form of risk compensation.

Keywords: distributional synthetic controls; equity return distributions; minimum wage policy

I would like to express sincere gratitude to Professor Gunsilius for his guidance and for how much he has deepened my knowledge and interests in econometrics and statistics, throughout this thesis and his Econ 671-679 PhD sequence.

Contents

1 Introduction

1.1 Background Information of The Minimum Wage Policy

Most European countries have not pursed a National Minimum Wage policy aggressively in the 20th century, and the United Kingdoms are no exception. Since then United Kingdom enacted the National Minimum Wage Act of 1998, the first national level was introduced in the following year 1999 and started to draw attention among social scientists to study its potential impacts on society. The Main goals of such a policy include but are not limited to reduce income inequality, curb potential exploitation of low-skilled workers, and guarantee a sustainable living wage for the poor. In that National Minimum Wage Act of 1998, £3.60 per hour for workers older than 21 years of age, together with a youth development rate at £3.00 per hour for 18–21-year-olds was implemented. After the Liberal Democrats was in charge of the government for a full 5 year period, between 2010 and 2015, stimulating a faster increase of national minimum wage level in that period, the Conservative Party won the election in May 2015.

The Chancellor of the Exchequer, George Osborne, as the new leader in the Conservative party who are traditionally opposed to increases of minimum wage, announced what he termed the "National Living Wage" at the end of his budget speech on 8 July 2015, which started at £7.2 per hour in April 2016 and was projected to rise to at least £9 per hour by April 2020. The unexpected claim aims to serve as an emergent budget call to offset the reduction of corporate tax credits published in Her Majesty's Treasury's document on July 1st. The following Figure 1 illustrates that this claim leads to the most significant increase of national minimum wage level since 1998 and it caused an immediate increase of 50 pence from initially projected level in the end of 2015, roughly 7.5%.

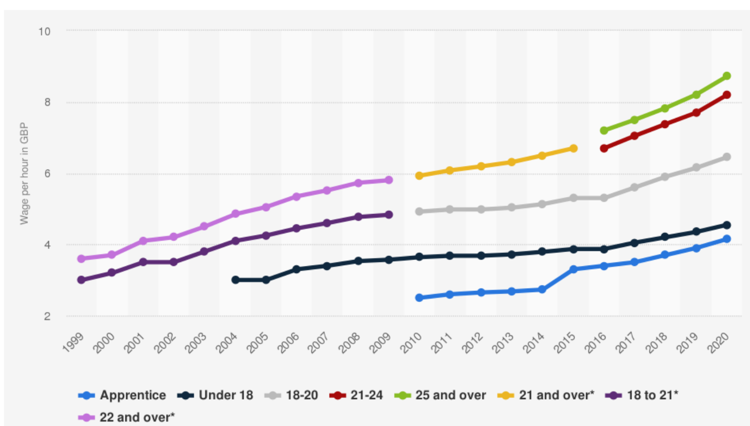


Figure 1: National Minimum Wage in the United Kingdom from 1999 to 2020 [NMW]

1.2 Literature Review

Before presenting literature review on statistical methodology, we shall also discuss about the research question itself in the relevant economics literature. The economic effects of imposing a minimum wage policy have been debated and widely studied among Labor Economists. Key questions are whether it is really an effective public policy to solve societal issues like unemployment rates and income inequality, presented in both short run and long run. There has been a trend of studies gradually shifting from survey data ([Brown et al., 1982], [Card, 1992]) to more micro-based on fast food restaurants [Card and Krueger, 1993], and then recent theoretical work on European scenarios ([Neumark and Wascher, 2006], [Riley and Bondibene, 2017], [Christl et al., 2018]).

Moreover, political effects also spread impacts out on firms and industrial organizations, and those losses of firm profitability via increased labor input costs accrue over time and may pass negative effects back to consumers and society, limiting the effectiveness of the policy itself. On the firm-Level, [Draca et al., 2011] have demonstrated that wage gains from minimum wage map into firms' profit reductions and bigger falls in margins in industries are those with relatively high market power. Within a broader context of Labor Economics, [Chodorow-Reich et al., 2021] used a local labor market analysis to show that a 20 percent in stock valuations drives up aggregate labor bill at least by 1.7 percent and improves local employment and payroll in nontradable industries and in total, with no effect on employment in tradable industries. Other researchers also illustrate that stock prices of firms in immigrant-intensive industries increased dramatically (70 percent - 170 percent) after the influx of low-skilled immigrants, and they show both evidences by United States studies of the Immigration Act of 1990 and 1999 Temporary Protected Status order. Moreover, this outperformence of the entire market (α arbitrage) in stock returns of observed firms lasted at least six months. Thus, empirical evidences do indicate that risks from labor markets may impact asset paths and it's worthwhile to investigate further on movements of equity values in financial markets.

Nonetheless, main papers with analysis on minimum wage policy, such as one paper studied in United States, the veto of a minimum wage raise by President Bush, show no significant effects on stock markets with labor market risks. One plausible explanation based on the Efficient Market Hypothesis [Fama, 2021] is that as the endogeneity exist in most empirical economics work, stock markets also have similar issues, namely the degree of efficiency in stock markets. If the market is efficient and when the event does not convey the new information, the effect may have already been absorbed. It thus requires an unexpected shift in policy to deliver new information, similar to an natural experiment, and to detect causal effects. Thus, it explains why UK Minimum Wage on 2015 deserves particular study since it was announced by the Conservative party who used to discourage the surge of national minimum wage level. This treatment is plausibly exogenous, not chosen by the firm, and uncorrelated with both observed or unobserved firm characteristics. [Bell and Machin, 2018] examined the exactly same event as we do in this paper and showed evident falls in the stock market value of low wage firms, where the decrease is in comparable magnitude of traditional studies on potential falls of firms' profitability in response to labor input costs. They adopted Fama-French 4 Factor Model to compute weighted cumulative abnormal returns on stock markets before and after the announcement of the policy. Those analysis of labor market risks are in fact vital to understand the long-run adjustment since equity values are foundations for large companies to gather investment funds, but we still lack strong basis to claim the true intervention or causal effect on reduction of returns if we only identify the risk-premium. This is important since the idea of risk-premium in asset pricing models only predict what an asset is expected to yield in excess of the risk-free rate of return, rather than the true philosophical impact of the inborn risk of the policy.

In contrast to the work of [Bell and Machin, 2018], instead of relying on asset pricing to estimate expected returns which are aggregate outcomes, this paper seeks for dynamic link, namely causal effects of risks from this unexpected announcement on stock markets. To obtain one category of labor-intensive industries for the sake of finding potentially most significant effects, we choose to specify public listed Real Estate companies on London Stock Exchange Market. Therefore, we try to detect the causal effect of the UK's NWM Policy on July 8th 2015 into equity returns around that period, i.e. How much decrease of returns are due to this announcement of the policy? This paper only focuses on the short run, since stock markets are so volatile that long-run effects may even not exist.

2 Methodology

In this section, we provide an overview of relevant literature on past major approaches and present their deficiency for implementation on panel data into the finance literature. One widely used statistical technique is Difference-in-Differences (DID), which compares the average change over time in the outcome variable for the treatment group and the average change over time for the control units, with an assumption of parallel trend of time series. This assumption is usually too strong to be proper for financial data and it is highly likely that there even does not exist a parallel trend in stock markets ever. Another group is the classical methods of Synthetic Controls (SCs) [Abadie et al., 2015], which construct a weighted combination of groups as control units to match characteristics of the target unit, assuming that the optimal combination of the control units hold "optimal" in the post-intervention periods. One of the main advantage is the feasibility of allowing potential time-varying unobserved confounders, and in asset pricing models, poor prediction power of expected risk-premium in extreme events are highly likely due to those (omitted) latent variables. Synthetic control methods is indeed important for financial studies and some researchers have adopted them in the literature ([Fremeth et al., 2016] [Opatrny, 2021]) for similar organization and financial markets studies as ours. Nonetheless, this group of synthetic controls lacks the ability to separate aggregate scalar or vector-valued quantities, limiting the sense of studying causal effects into scenarios like our return distributions. These facts motivate our study of exploring the newly introduced Distributional Synthetic Control (DSCs) and its estimator [Gunsilius, 2020]. Though still depending on the extrapolation step over time, we can replicate complete distributions of distributions of equity returns and study the whole counterfactual process, rather than aggregate effects on discounted future cash flows or risk premium as a form of normalized expected returns. Our analysis based on individual-level data therefore provides a geometrically faithful estimate of the entire counterfactual quantile function of the treated unit. With DSCs, we acquire optimal weights of Wasserstein Barycenter to replicate a synthetic post-intervention distribution. The difference between the synthetic unit and actual unit will be the causal effect.

Remark. *Because of unexpectedness (randomness) within the announcement, the intervention of the national minimum wage policy is similar as a treatment in experimental research studies, so we use the word treatment and intervention interchangeably subject to context.*

2.1 Prediction on Wasserstein space

As the statistical approach used in DSCs relies on the theory of barycenters in 2-Wasserstein space [Agueh and Carlier, 2011], we start to briefly introduce relevant mathematical background and terminologies.

Definition 1. *The* 2*-Wasserstein distance* $W_2(P_X, P_Y)$ *between two probability measures* P_X *and* P_Y *supported on* $\mathbb{R}^d$ *is defined as*

$$W_2^2(P_X, P_Y) := \inf_{\gamma \in \Pi(P_X, P_Y)} \left\{ \int_{\mathscr{X} \times \mathscr{Y}} |x - y|^2 \, d\gamma(x, y) \right\},$$

where $|\cdot|$ *is the Euclidean norm,* $\Pi(P_X, P_Y)$ *denotes the set of all joint probability distributions* γ *on* $\mathbb{R}^d \times \mathbb{R}^d$ *for which* P_X *and* P_Y *are the marginals, and where both* P_X *and* P_Y *have finite second moments, i.e.*

$$\int |x|^2 \, dP_X(x), \int |y|^2 \, dP_Y(y) < +\infty.$$

Definition 2. *The Wasserstein space is defined as the set* $\mathscr{P}_2(\mathbb{R}^d)$ *of probability measures on* $\mathbb{R}^d$ *with finite second moments, equipped with the Wasserstein distance.*

Definition 3. *The barycenter of a set of probability measures* $\{P_j\}_j$ *in the* 2*-Wasserstein space is defined as the element* P^* *which solves the following optimization problem [Agueh and Carlier, 2011].*

$$\min_{P \in \mathscr{P}_2(\mathbb{R}^d)} \sum_{j=1}^{J} \frac{\lambda_j}{2} W_2^2(P, P_j),$$

where the weights $\lambda_1, \ldots, \lambda_J$ *are non-negative and sum to unity, i.e.* $\sum_{j=1}^{J} \lambda_j = 1$.

Based on these concepts, the theoretical approach will be as follows. We assume we observe target return distributions P_t of the United Kingdom real estate sector and a set $\{P_j\}_{j=1,\ldots,J}$ of control return distributions for some real estate sectors in other countries which we will use to predict the return distributions of the target unit. The idea then is to find the optimal weights $\lambda_1^*, \ldots, \lambda_J^*$ for which the resulting barycenter is the closest possible to the target distribution P_t as measured in the Wasserstein distance [Bonneel et al., 2016], i.e.

$$\lambda^* = \operatorname*{argmin}_{\lambda_1, \ldots, \lambda_J \in \mathbb{R}^J} W_2^2(P_t, P^*(\lambda_1, \ldots, \lambda_J))$$

where $P^*(\lambda_1, \ldots, \lambda_J)$ is the barycenter of the set of control probability measures $\{P_j\}_j$, and in this paper, we will only consider univariate return distributions.

2.2 The Basic Causal Model

Let Y_{jt} denote a stochastic process of the aggregate outcome of interest- daily returns with intervention, and $Y_{jt,N}$ be the counterpart without intervention, for units $j \in J$ and time periods $t \in T$. Then we introduce a basic nonseparable and nonlinear model [Athey and Imbens, 2006], which relies on the idea of matching observables and unobservables, in our causal framework.

$$Y_{jt,N} = h(t, U_{jt})$$

where we postulate the existence of some measure-preserving functions h and the latent random variable U_{jt} represents the unobservable characteristics of unit j.

In other words, the distribution μ_t of Y_t is the *pushforward* of ν_t by h_t, where ν_t is the (time-variant) distribution of U_t, written as $\mu_t = (h_t)\#\nu_t$. One major advantage to adopt DSCs in analysis of volatile financial markets is that the causal model could allow the latent random variable U to change over time. Then the core research question is to estimate the treatment (causal) effects:

$$\tau_t = Y_{jt} - Y_{jt,N}$$

2.3 Implementation of Distributional Synthetic Controls

Next, we extend the outcome-daily returns into outcome-return distributions and formalize some notations aligned with Gunsilius [2020]. The implementation of distributional synthetic controls is essentially a time-period-by-time-period prediction approach on the Wasserstein space.

The quantity of interest is $F_{Y_{jt}}$, the probability law of Y_{jt}, where in the section *Application to Real Data*, we have $j \in J$ and we observe $J+1$ units cross periods $1, 2, \ldots, T$, in which the first unit $(j = 1)$ is regarded as the target unit (UK) and rest j units are control units. Along the time series, the target unit is exposed to the intervention starting from periods $T_0 + 1, ..., T$. In accordance with the literature, we call $t \leq T_0$ the pre-intervention- or pre-treatment periods and $t > T_0$ the post-intervention- or post-treatment periods. That said, in our setting, T_0 is the date of the unexpected announcement, on July 8th 2015. Our first unit is real estate sector in United Kingdom, and rest countries of stock exchange markets are control units.

The method regresses quantile functions on whole quantile functions, i.e the explanatory variables are functional quantities, and the quantile function is defined as:

$$F^{-1}(q) := \inf\{x \in \mathbb{R} : q \leq F(x)\}, \text{ for } q \in [0,1]$$

Then it proceeds with computing quantile functions of target unit and control units, and then computing the optimal weights λ^* in Wasserstein space as follows:

$$\vec{\lambda}_t^* = \underset{\vec{\lambda} \in \Delta^{J-1}}{\operatorname{argmin}} \int_0^1 |\sum_{j=2}^{J+1} \lambda_j F_{Y_{jt}}^{-1}(q) - F_{Y_{1t}}^{-1}(q)|^2 \, dq, \quad \text{where } \Delta^{J-1} := \text{a unit simplex} \tag{1}$$

Empirically, we randomly draw a standard uniform distribution with sample size M, denoted by U_M, and solve the approximation of equation (1):

$$\vec{\lambda}_t^{\,*} = \underset{\vec{\lambda} \in \Delta^{J-1}}{\operatorname{argmin}} \frac{1}{M} \sum_{m=1}^{M} |\sum_{j=2}^{J+1} \lambda_{jt} F_{Y_{jt}}^{-1}(U_m) - F_{Y_{1t}}^{-1}(U_m)|^2 \tag{2}$$

To summarize, we acquire optimal weights from equation (2) and average the optimal weights equally for all pre-intervention period(s), i.e

$$\vec{\lambda}^* = \sum_{t=1}^{T_0} w_t \vec{\lambda}_t^*, \quad \text{where}$$

$$w_t = \frac{1}{T_0}$$

Finally, in all post-intervention period(s), we obtain the counterfactual quantile distribution for the target unit without intervention and the DSC Estimator is shown in equation (3):

$$F^{-1}_{Y_{1t,N}}(q) = \sum_{j=2}^{J+1} \lambda_j F^{-1}_{Y_{jt}}(q)$$

$$\textbf{DSC Estimator}: \widehat{\tau_{1t}} = F^{-1}_{Y_{1t}}(q) - F^{-1}_{Y_{1t,N}}(q) \quad \forall q \in [0,1] \tag{3}$$

Note that equation (1) is a convex optimization problem which has a unique solution of weights. As M goes to infinity, optimal weights derived from the approximation equation converge to the optimal weights in the integral expression. Besides, one may relax $\{w_t\}_{t \leq T_0}$ to be some weights in the unit simplex and get rid of influential time periods [Arkhangelsky et al., 2019]. This is, however, not required in our setting because of shorter time periods.

3 Simulation Studies

In this section, relating with our financial applications, we study the power of distributional control methods under varying sample sizes. As we focus on stock prices in panel data case, we perform simulations tests based on multi-dimensional geometric Brownian motion model.

Definition 4. *A d-dimensional geometric Brownian Motion GBM($\mu, \sum$) is a continuous stochastic process $Y = (Y_1, \cdots, Y_d)^T$ which satisfies the following system of stochastic differential equations.*

$$\begin{cases} dY_i(t) & = \mu_i Y_i(t)dt + \sigma_i Y_i(t) dW_i(t), \quad i = 1, \cdots, d \\ Y(0) & = (Y_1(0), \cdots, Y_d(0))^T \end{cases}$$

- *where each W_i is a standard one-dimensional Brownian motion on $(\Omega, \mathscr{F}, P)$ with natural filtration $\{\mathscr{F}_t\}_{0 \leq t \leq \infty}$, and $W_i(t)$ and $W_j(t)$ have correlation ρ_{ij}*
- *where μ is a deterministic $d \times 1$ drift vector $(\mu_1, \cdots, \mu_d)$, σ is the corresponding deterministic $d \times 1$ volatility vector, and $\sum_{ij}$ is the $d \times d$ covariance matrix $\sigma_i \sigma_j \rho_{ij}$*

Using the fact that the covariance matrix is always positive semi-definite, we apply the Cholesky decomposition to the $\sum$ and then we could rewrite the definition as

$$\frac{dY_i(t)}{Y_i(t)} = \mu_i dt + \sum_{j=1}^{d} A_{ij} dW_j(t) \quad i = 1, \ldots, d$$
$$\text{for any matrix } A \text{ such that } AA^T = \sum$$

Next, given d-dimensional assets in each (target/control) unit with k time periods, we let $Y(t) \in \mathbb{R}^d \times \mathbb{R}^k$ denote the time series of units' price levels. As it has both stationary and independent increments, the above characterization leads to the following algorithm for simulating stock prices in the form of geometric Brownian motion model with constant parameters. For the reason that parameter estimations is not the focus in this paper, we let μ and σ be deterministic over time, which we interpret them as the return and volatility or standard deviation parameters.

Algorithm 1 Simulations

Input: d: number of stocks in each unit; k: number of time periods; $(\min_\mu, \max_\mu)$: the range of return parameter; $(\min_\sigma, \max_\sigma)$: the range of standard deviation parameter

Output: $Y(t) \in \mathbb{R}^d \times \mathbb{R}^k$: a simulated time series of unit's prices

Function `Generator` $(d, k, \min_\mu, \max_\mu, \min_\sigma, \max_\sigma)$ **:**

  initialize a $d \times k$ matrix as $S(0)$

  uniformly randomize $\mu = \{\mu_1, \dots, \mu_d\}$ and $\sigma = \{\sigma_1, \dots, \sigma_d\}$ by $\min_\mu, \max_\mu$ and $\min_\sigma, \max_\sigma$

  randomly initialize a $d \times d$ correlation matrix ρ and set $\sum = diag(\sigma)\ \rho\ diag(\sigma)$

  apply Cholesky factorization to $\sum$ and record it as A;

  randomly generate a matrix Z with k i.i.d pairs of $N(0,1)$

  for *each stock* $i = 1, \dots, d$ **do**

    for *each time period* $m = 1, \dots, k-1$ **do**

      compute $Y_i(t_{m+1}) = Y_i(t_m) \exp\left((\mu_i - \frac{1}{2}\sigma_i^2)(t_{m+1} - t_m) + \sqrt{t_{m+1} - t_m} \sum_{j=1}^{d} A_{ij} Z_{i,m+1} \right)$

    end

  end

  return $S(t)$

End Function

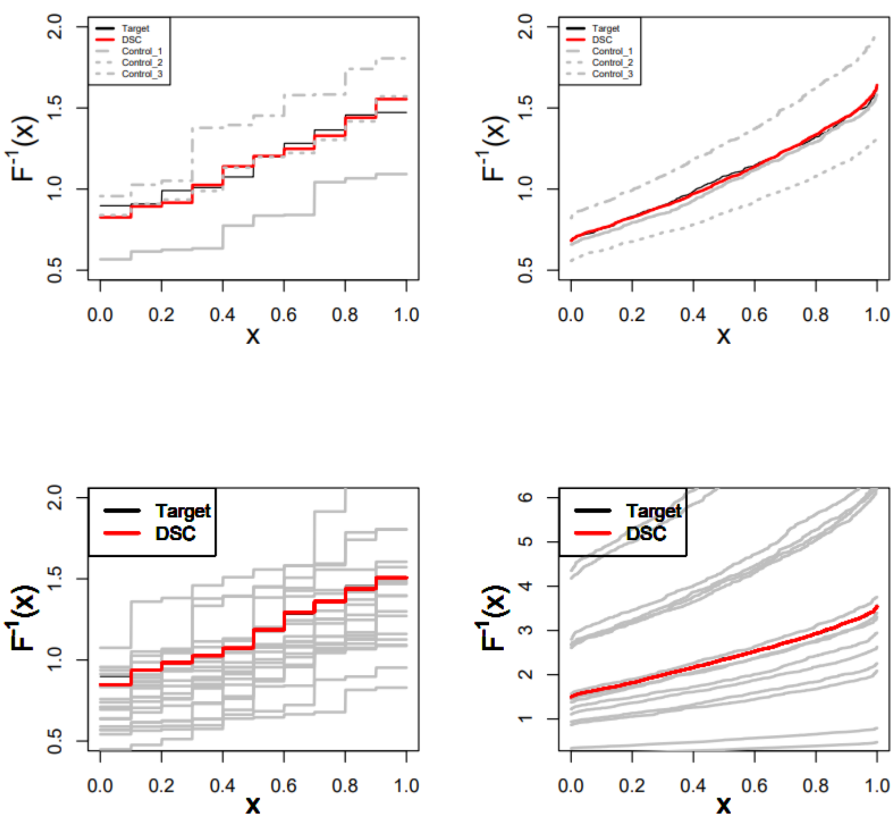


Figure 2: Comparison results of 3 control units and 20 control units. Row-wise comparison is for sample size of 10 and 500. The proportion of parameter of the simulation algorithm: 10,3,-0.2,0.2,0,0.1.

There are several clarifications of the simulation algorithm and the above figure that we hope to address to. Firstly, our outcome of interest, denoted by Y, is for daily returns or daily return distributions, but for simplicity, we choose to start from the same stock price 1 for all stocks in all units so that we can get rid of extra calculations of returns. Secondly, we choose to simulate three periods but only display the last one. Lastly, we do not impose any intervention throughout this parametric simulations and therefore do not conduct any causality analysis in this section. The main goal of this section is to have the following observations: (1) DSCs works well even with small group of control units, as long as they are meaningful chosen (2) Larger group of control units and larger sample size in each control unit both increases prediction powers.

4 Application to Real Data

4.1 Data Description

According to the Global Industry Classification Standard (CIGS), 11 different stock sectors have been classified, which are *Energy, Materials, Industrials, Utilities, Healthcare, Financials, Consumer Discretionary, Consumer Staples, Information Technology, Communication Services, Real Estate*. For both the target unit (United Kingdom) and control units, our focused sector is the *Real Estate* sector, and some estate focused engineering constructions companies in the *Industrials* sector. These two sectors are relatively centralized, with limited interactions (cross-ownership) with other sectors, but also very diversified in terms of company sizes and equity values. This fact minimizes outsider impacts and helps our later analysis of redistribution of return distributions.

Previous literature, for instance, Bell and Machin [2018] investigates Intra-Day returns and claims effects stabilizing around 2% after the first day, July 8th, and lasting for next few days. While revisiting this week's data in above sectors specifically, in contrast to their work in a high-frequency viewpoint, we analyze the daily return distributions throughout the whole week (five weekdays) from July 6th to July 10th. Our pre-treatment periods therefore are July 6th and July 7th and post-treatment periods are July 8th, July 9th & July 10th. We select and report nine control units: United States, Canada, Australia, Singapore, Japan, European(Euronext), Hong Kong, Indian & Germany. For simplifications, in figures, they are shorted as US, CA, AU, SI, JP, EU, HK, ID & GM respectively. We define stock returns based on adjusted closing price over two consecutive periods, including information after accounting for any corporate actions such as dividends and undergoing stock split. Adopting closing prices rather than open prices incorporates information released by firms after the closure of stock exchange markets. More analysis and selection procedures are attached in the Data section of Appendix.

Definition 5. *Adjusted Periodic Stock Return:*

$$r_{i,j,t} = \frac{Y_{i,j,t} - Y_{i,j,t-1}}{Y_{i,j,t-1}}$$

where Y indicates the adjusted stock prices, j indicates the unit index, i indicates the stock index within the unit, t is the time index in recorded daily unit, and r is the return acquired from two consecutive trading days, i.e not necessarily over two days in calendars due to the gap of weekends.

Given this definition, we display kernel density (return distribution) plots of our target sector, United Kingdom, before and after the announcement of the minimum wage policy. Recall that this unanticipated increase of national minimum wages is to offset an overly intense policy on July 1st 2015 when Her Majesty's Treasury 2015 announced a reduction of corporate tax credit from 20% to 19% in 2017 and 18% in 2018. Therefore we need to assume that the London Stock Exchange Market is at least in the weak-form efficiency [Fama, 2021], i.e all market information are reflected in historical prices and therefore last week's information does not impact on this week's prices anymore. This assumption is indeed reasonable because Figure 3 illustrates that the return distribution follows a Gaussian distribution with mean 0.

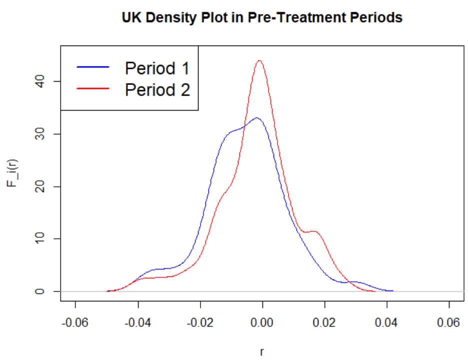


Figure 3: UK Density in Pre-Treatment Periods

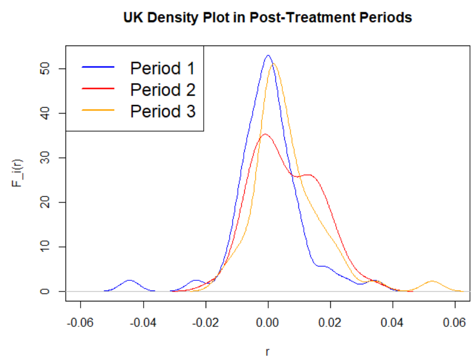


Figure 4: UK Density in Post-Treatment Periods

4.2 Analysis on Pre-treatment periods

In this section, we acquire the optimal weights λ^* and provide with relevant pre-treatment plots according to the Section: Implementation of Distributional synthetic controls.

Table of Optimal Weights	
	Weights
United States	0.043597806
Canada	0.048290756
Australia	0.059187076
Singapore	0.471751566
Japan	0.038356249
Europe	0.291433064
Hong Kong	0.040878438
India	0.006504984

DSCs has the explanatory power of weights on control units. Because of similar structures in real estate sector and closely related economy, it is intuitive that Europe and Singapore have higher weights, 0.29 and 0.47 respectively. One plausible explanation for Singapore to have the highest is that its real estate sector is the smallest among the rest and less diversified. We could also see that for developing countries like India, they are assigned with an almost trivial weight. Overall, this table indicates that our results are both statistically and economically meaningful.

Given those optimal weights of control units, we then construct the distributionally synthetic control unit over the pre-treatment periods and display following figures.

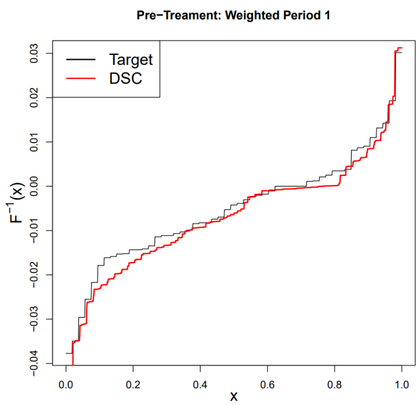


Figure 5: UK Pre-treatment: Period 1

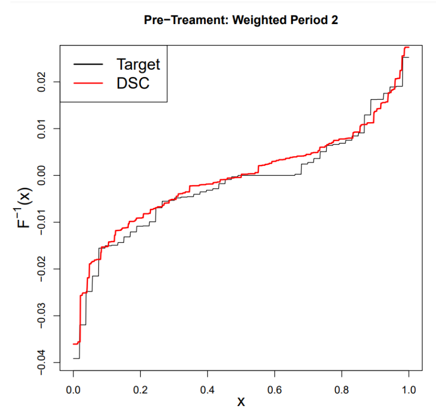


Figure 6: UK Pre-treatment: Period 2

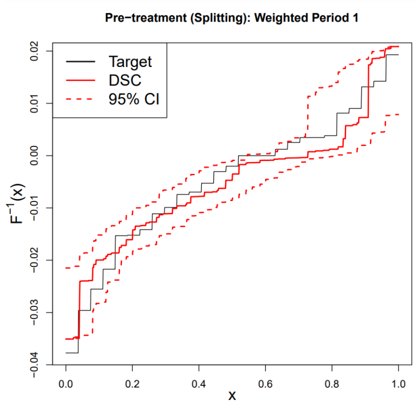
4.2.1 Uniform Confidence Intervals

Next, we acquire the uniform confidence intervals by 50-50 sample splitting on the pre-treatment periods. That said, optimal weights of the control units are computed on the first half data, and we use those weights to estimate the barycenter and the confidence intervals on the second half of the data. The procedure is conducted via a standard bootstrap.

For illustrative purpose, we provide two groups of weights for comparison in the following table, but for future analysis in later sections, we insist on regular group of weights, unless elsewhere notified. Note that though specific values change, the pattern holds similarly, in which Singapore and Europe still have highest weights, and India has the lowest.

Table of Optimal Weights		
	Regular	Splitting Case
United States	0.043597806	1.150637e-02
Canada	0.048290756	3.713365e-02
Australia	0.059187076	-1.969413e-06
Singapore	0.471751566	6.831869e-01
Japan	0.038356249	5.182494e-02
Europe	0.291433064	1.853997e-01
Hong Kong	0.040878438	2.657071e-02
India	0.006504984	4.381927e-03

Then we display figures with respect to return distributions of the distributionally synthetic control unit and the target unit with 95 % confidence intervals over pre-treatment periods.



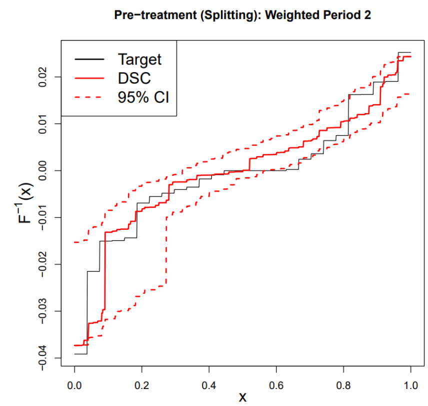
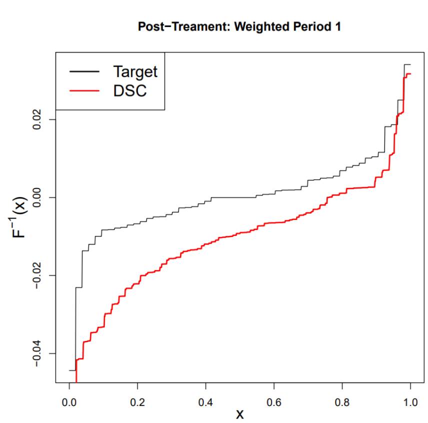


Figure 7: UK Pre-treatment: Period 1(95 % CI) Figure 8: UK Pre-treatment:Period 2(95 % CI)

4.3 Analysis on Post-treatment periods

Other than standard regularity assumptions, we also require that the optimal weights derived in the pre-treatment periods stay optimal after the treatment (intervention). In other words, the structure of DSCs is an extrapolation over time. Then we have the following three figures.

Before conducting any statistical analysis, these figures show that the potential causal effects seems to only exist in the first post-treatment period, and perhaps also the second one. This is consistent with the assumption concerning weak-form of efficiency in London Stock Exchange Market that we made when we analyzed UK's density plot of return distributions in pre-treatment periods and claimed that the overly intense policy with respect to reduction of tax credits on July 1st should be absorbed within a week. This result further shows that investors respond to an announcement with drastic impacts much faster, and the effects diverge away only in one or two periods. It also partially explains why the focus of *Minimum Wages and Firm value* [Bell and Machin, 2018] is on high-frequency data and Intra-Day returns, though not in true philosophical causality like ours.



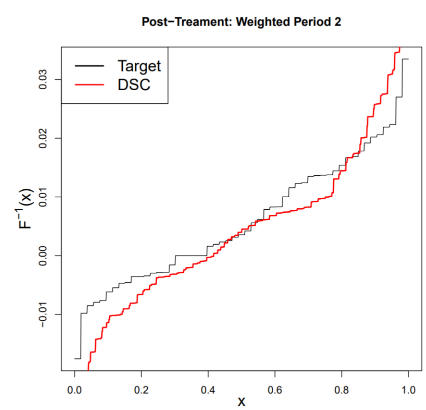


Figure 9: UK Post-treatment Period 1 Figure 10: UK Post-treatment Period 2

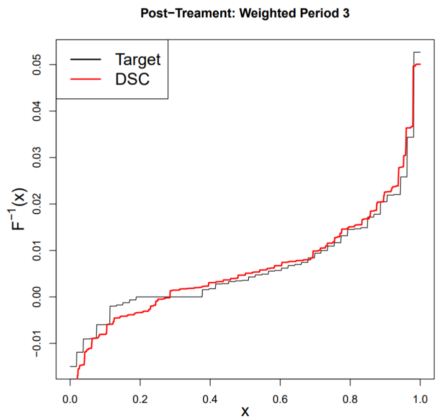


Figure 11: UK Post-treatment Period 3

4.4 Permutation inference for causal effects

Though we mimic an experimental study, the classical randomization inference when the intervention is randomly assigned might still be improbable in our setting. To identify if a causal relationship exists, we proceed with the following pseudo-algorithm and summarize a table concerning Wasserstein distances based on placebo tests. ([Abadie et al., 2015], [Gunsilius, 2020]).

We obtain permutation distributions by iteratively reassigning the treatment into control units and estimate the placebo effects. We conclude that there *might* be a causal relationship, subject to our causal model, between the new national minimum wage policy and differences of returns distributions, if the magnitude of the effect (non-negative values of Wasserstein distances) on target unit is extreme relative to rest permutation distributions.

Algorithm 2 Placebo permutation tests

for *each unit* $k = 1, \dots, J+1$ **do**
  for *each time period* $t = 1, \dots, T_0$ **do**
    obtain and record the optimal weights $\lambda^*_{t,k}$ using (1)
  end
  compute the overall optimal weights $\lambda^*_k = \sum_{t \le 1} w_t \lambda^*_{t,k}$
  for *each time period* $t = T_0 + 1, \dots, T$ **do**
    compute 2-Wasserstein barycenter using the weights λ^*_k to obtain $F^{-1}_{Y_{kt,N}}$
    record the 2-Wasserstein distances $d_{kt} := (\int_0^1 |F^{-1}_{Y_{kt}}(q) - F^{-1}_{Y_{kt,N}}(q)|^2 dq)^{\frac{1}{2}}$
  end
end

4.4.1 Comparison of Wasserstein Distances

Table of Wasserstein Distances Based on the Test			
Country Name or Area Name	Period 1	Period 2	Period 3
Hong Kong	0.06495521	0.1276433	0.03148504
UK*(target)	0.01527052	0.017276913	0.002702466
Japan	0.01452354	0.01497696	0.02148388
United States	0.01326735	0.01133127	0.009728711
Singapore	0.01311423	0.009610633	0.003978411
Australia	0.01215706	0.0156904	0.008803297
India	0.01081684	0.00868945	0.01374697
Europe	0.004649775	0.008963285	0.005866267
Canada	0.003274371	0.0122054	0.007116168

We observe that the target unit (United Kingdom) has the most extreme Wasserstein distance relative to other control units, if we exclude the *outlier* Hong Kong, which only takes 3 to 5 percent weights in the synthetic control unit. The significance of the effects on Hong Kong is, however, not negligible. One plausible explanation is that due to the 2015-2016 Chinese stock market turbulence starting from 12 June 2015, the stock market had fallen 30 percent over three weeks, impacting more than half public listed companies, by 8–9 July 2015. This drastically impacted Asian stock exchange markets and we strongly believe that it was independent of our intervention in United Kingdom. Then, if we assume that there was indeed devastating strike, DSCs should not be able to replicate Hong Kong sector well given other control units in pre-treatments, which we will show in the next section.

4.4.2 Poor Matching for Hong Kong in the Pre-Treatment Periods

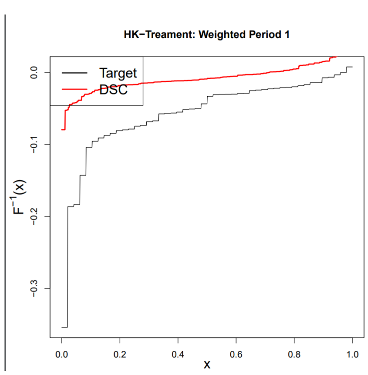


Figure 12: HK Pre-Treatment Period 1

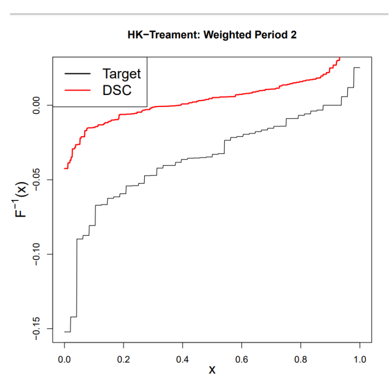


Figure 13: HK Pre-Treatment Period 2

4.4.3 Summary

In this section, we summarize findings of causality analysis. First of all, our DSCs still replicate the target unit-United Kingdom well after the intervention since the synthetic control unit holds geometrically faithful shape. This is a basis of our second point.

Secondly, after excluding the outlier (Hong Kong) and taking potentially existing effects into other Asian units (Japan, Singapore, India), the Wasserstein distance of United Kingdom is most extreme relative to rest permutation distributions and therefore we conclude with a causal relationship, subject to our causal model. This then delivers some financial meanings. According to the density plots of post-treatment periods (Figure 4), period 1's return distribution roughly follows a symmetric distribution centering around 0. Nonetheless, due to the existence of the corresponding causal effect, investors were compensated for bearing extra risks and uncertainty from this unexpected increase of national minimum wage policy. That said, the whole distributions of equity returns in the UK's real estate sector would be left-skewed to some extent without the intervention. Furthermore, the difference of returns between the DSC line and Target line is not equally distributed, and firms with more negative returns were compensated more, i.e there exists a larger gap. The difference gradually gets smaller when the returns are becoming more positive. Put differently, equities with (extreme) negative returns or (extreme) positive returns, namely returns on two tails, are relatively more riskier to invest. Notice that the phrase of more positive returns only refers to a relative scale, and the highest return in the post-treatment period 1 is merely a risk-free rate of return, which is 2.1% [RF]. This explains that the gap around 0.02 of the y-axis in that period is nearly invisible. Though causal effects are gradually absorbed after the first period, one might still be able to detect a similar pattern of risk compensation on two ends over the next two periods. This idea is very similar as the standard risk-premium concept in asset pricing models [Bell and Machin, 2018], but via our DSCs methods, we could present it in the form of true philosophical causality.

5 Conclusion

In this paper, we conduct a simple simulation analysis relating financial applications into distributional synthetic controls and study some properties of its prediction power. Then we investigate short-run causal effects of the unexpected UK National Minimum Wage Policy. Assume that optimal weights derived in the pre-treatment periods remain unchanged after the intervention and equities are auto-correlated but not cross-related, we study the counterfactual process of the complete distributions of equity returns' distributions. We find that the minimum wage hike actually caused an increase in returns for most stocks in the distributional sense, via the form of risk compensation in philosophical causality rather than the counterpart of risk-premium in Asset Pricing Models.

Needless to say, this paper has several limitations and future work on those aspects might be very useful. Per the data, a better proxy of the selection criteria for sectors and incorporation of larger sample sizes is very vital for more comprehensive and detailed conclusions. For instance, it will help us acquire a more precise uniform confidence intervals on pre-treatment periods after 50-50 sample splitting bootstrap method, therefore increasing prediction powers. Besides, a further analysis of more detailed categories of stocks developed in [Bell and Machin, 2018] will provide insights into the link between firm profitability by direct increase of costs in labor inputs and return distributions. Per the methodology, current DSCs is a time-period-by-time-period prediction approach on Wasserstein space, and an enrichment into multivariate approach using joint distribution over all time periods may adjust business cycle fluctuations and explain dynamic questions like how fast can a firm can adjust changes from labors to technology or capital.

6 Appendix

6.1 Data

The data is from Yahoo finance and also available by the quantmod package in R. We would like to reiterate several common remarks that one might be interested to ask.

- Since the existing literature lacks a precise and consistent selection criterion for stocks into real estate sectors, we rely on the categorization from one third-party, *Financial Knowledge and Information Portal*, which ranks stocks with market capitalization. Note that Fknol's classification and thus our paper does include Real Estate Investment Trust (REIT) companies. As long as major investments of those companies are related to real estate, their return distributions should be affected and deserved concerns since they may reallocate holding portfolios after the announcement, and investors will capture signals from these series of actions, adjusting investments in REITs' stocks. This paper tries to incorporate as many assets as possible, for the sake of reflecting a comprehensive and authentic return distributions of sectors, but excludes some with low trading volume, especially those with no trading over given periods at all.
- Control units are determined with following schemes. We firstly rank major stock exchange groups by their market capitalization in global shares, and identify countries by developed stock markets and emerging stock markets. That said, different stock exchange markets within one country will be combined, for example, NASDAQ and NYSE will be put into one control unit since they are both classified as developed stock markets within United States. Since Stocks are not non-traded goods, to isolate displacement effects and subsequent endogenous issues, one important remark is that each control unit does not include companies that go public in London Stock Exchange Market or have main services or over half of employees in the United Kingdom.
- More comprehensive Budget 2015 Measures could be found in Appendix of [Bell and Machin, 2018] or the original document of the parliament.

6.2 Parametric Estimation in Simulation Studies

For pedagogical practices and illustrations of some arguments in the main section, this section complements the parametric estimation analysis for Geometric Brownian Model (GBM) in Section 3. To simulate asset paths with GBM, we need to select an appropriate parametric set of $(\mu, \sigma) \in \Theta$, a (finite-dimensional) bivariate parameter space. Then we investigate two standard sets of parametric estimators (Maximum Likelihood Estimator and Successive Percent-Returns Estimator). For simplicity, we assume that the expected returns are independent of the value of the asset price series and they are Gaussian distributed.

6.2.1 Maximum Likelihood Estimator (MLE)

Maximum Likelihood Estimator is a method of estimating the parameters for the assumed probability density function f_θ ($\theta \in \Theta$) that will maximize the likelihood of observed data sample $(x_1, x_2, ..., x_n)$ for the random vector $(X_1, X_2, ..., X_n)$.

Given the i.i.d assumptions of random variables, the log-likelihood ratio is written as follows:

$$L(\theta) = \sum_{i=1}^{n} \ln f_{\theta}(x_i)$$

Next, differentiating the density function with respect to each parameter and setting the resulting derivative equal to zero, the (ultimate) drift parameter will be adjusted with volatility. We also develop some supporting financial meanings behind this adjustment later. Given $x_i = \ln S_{t_i} - \ln S_{t_{i-1}}$, let $m_i = \sum_{i=1}^{n} \frac{x_i}{n}$, and then we have the maximum likelihood estimator:

$$\textbf{Maximum Likelihood Estimator:} \quad \hat{\mu} = \hat{m} + \frac{1}{2}\sigma^2 \quad \hat{v} = \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{n-1}$$

Note that the Markov property allows us to write the likelihood along a time series of observations as a product of transition probabilities between each pair of two adjacent time steps for non-i.i.d stochastic processes. More details can be found out in [Brigo et al., 2007]

6.2.2 Successive Percent-Returns Estimator (SPRE)

Other than the maximum likelihood estimator, we also consider ratios of successive percent differences between daily observations. Given $x_i = \ln S_{t_i} - \ln S_{t_{i-1}}$,

$$\textbf{Successive Percent-Returns Estimator:} \quad \hat{\mu} = \sum_{i=1}^{n} \frac{x_i - x_{i-1}}{x_{i-1}} \quad \hat{v} = \frac{1}{n-1} \sum_{i=1}^{n} \left(\frac{x_i - x_{i-1}}{x_{i-1}} - \hat{\mu}\right)^2$$

6.2.3 Simulation and Analysis

An exactly same but univariate Monte carlo algorithm is implemented to generate simulation analysis. We let the true drift parameter as 0.03 and true variance parameter as 0.01. Given parameters, we specify steps (trading days) to be simulated as 90, and initial price is set to be 100.

A significance level of 0.05 will be used, i.e we will investigate if the known parameters was included in the 95 % confidence interval on each proposed set of parametric estimations. Next, we perform two hypothesis testings and analyze relevant operating characteristics.

$$H_0 : \mu = 0.03, \; H_a : \mu \neq 0.3$$
$$H_{0'} : \sigma^2 = 0.01, \; H_{a'} : \sigma^2 \neq 0.01$$

Table of Confidence Interval		
Student's t-test	Drift CI	Volatility CI
MLE	(0.02945804, 0.03079222)	(0.009982998, 0.010169430)
SPRE	(0.02989954, 0.03127553)	(0.01064713, 0.01085628)

This table shows that we may reject the null hypothesis with respect to volatility parameter for Successive Percent Return because it seems that there is an upwards biased trend. Before drawing any conclusions, we analyze the bias of those estimators for drift and volatility parameter,

$$bias = E[\hat{\theta} - \theta], \text{ where } \theta = \mu \text{ or } \sigma$$

We generate several samples with different size k of estimates.

Bias of Drift Parameter		
Sample Size	MLE	SPRE
k =100	0.0006435225	0.001143452
k =500	0.0005790974	0.00105818
k =1000	0.0001251313	0.0005875374
k =2000	0.0001746285	0.00063649

Bias of Volatility Parameter		
Sample Size	MLE	SPRE
k=100	0.0002026305	0.0009096269
k=500	3.124084e-05	0.0007112985
k=1000	7.621392e-05	0.0007517006
k=2000	-2.210805e-05	0.0006508395

As the sample size increases, for both estimators, the bias gradually reduces. Though the absolute scale of bias is not large, relative scale to true parameters will be significant. Therefore, we could then confirm the initial conjecture that Successive Percent Return estimator indeed has an upward biased trend of volatility parameter. Nonetheless, it can be the case that a procedure we claim a valid confidence interval has confidence coefficient $c < 1-\alpha$, where α is our significance level. We are then supposed to evaluate our procedures to exclude this scenario.

For a parameter θ and the random interval $[A,B]$, the confidence coefficient is defined as:

$$c = P(A \leq \theta, B \leq \theta)$$

If $c \geq 1-\alpha$, then we claim that $[A,B]$ is a valid $(1-\alpha)*100$ confidence interval. The procedure is to generate k times of samples of size k, and compute a 95 % confidence interval using the student t.test for each, and see how many samples contain the truth. The proportion which contains the truth is our estimates of the confidence coefficient. Since the confidence coefficient is a probability, with Monte carlo methods, we can express our uncertainty via a binomial proportions test based interval presented in 95 % proportion. The binomial test is an exact test of the statistical significance of deviations from a theoretically expected distribution of observations into two categories.

Confidence Coefficient of Drift Parameter		
Sample Size	MLE	SPRE
k=100	(0.87397, 0.97767)	(0.86108, 0.97139)
k=500	(0.92707, 0.96738)	(0.80917, 0.87469)
k=1000	(0.93912, 0.96613)	(0.72712, 0.78137)
k=2000	(0.93004, 0.95367)	(0.63140, 0.67545)

Confidence Coefficient of Volatility Parameter		
Sample Size	MLE	SPRE
k=100	(0.88717, 0.98357)	(0.00025, 0.054460)
k=500	(0.94856, 0.98160)	(0, 0.00735)
k=1000	(0.93237, 0.96092)	(0, 0.00368)
k=2000	(0.94771, 0.96591)	(0, 0.00184)

Then we draw the conclusion that Successive Percent Return estimator leads to an upward biased predictions of the volatility population parameter because it never includes 0.95. Furthermore, as the sample size k increases, even in the scenario of drift parameter, SPRE gradually deviates from including 0.95. In fact, these findings based on simulation results are extremely important and consistent with financial applications, because they illustrate that investors should never separate considerations of risks (volatility) and returns (drift). To get extra returns higher than risk-free rate, investors need to bear higher risks and risk-premium is the corresponding compensation. This is exactly connected with our inferential results in the main sections of the paper, though we present an alternative of risk-premium.

One last advantage of MLE is that empirically, the adoption of log version on MLE induces significant computational efficiency, especially when the sample size is large.

Running Time		
Sample Size	MLE	SPRE
k=100	4.448142 secs	4.681202 secs
k=500	1.22242 mins	1.278513 mins
k=1000	3.319433 mins	5.234686 mins
k=2000	15.96918 mins	20.80921 mins

References

Uk: National minimum wage 1999-2020 data source. URL `https://www.statista.com/statistics/280483/national-minimum-wage-in-the-uk/`.

Average risk free rate (rf) on investments in the united kingdom (uk) from 2015 to 2021. URL `https://www.statista.com/statistics/885750/average-risk-free-rate-united-kingdom/`.

A. Abadie, A. Diamond, and J. Hainmueller. Comparative politics and the synthetic control method. *American Journal of Political Science*, 59(2):495–510, 2015.

M. Agueh and G. Carlier. Barycenters in the wasserstein space. *SIAM Journal on Mathematical Analysis*, 43(2):904–924, 2011.

D. Arkhangelsky, S. Athey, D. A. Hirshberg, G. W. Imbens, and S. Wager. Synthetic difference in differences. Technical report, National Bureau of Economic Research, 2019.

S. Athey and G. W. Imbens. Identification and inference in nonlinear difference-in-differences models. *Econometrica*, 74(2):431–497, 2006.

B. Bell and S. Machin. Minimum wages and firm value. *Journal of Labor Economics*, 36(1):159–195, 2018.

N. Bonneel, G. Peyré, and M. Cuturi. Wasserstein barycentric coordinates: histogram regression using optimal transport. *ACM Trans. Graph.*, 35(4):71–1, 2016.

D. Brigo, A. Dalessandro, M. Neugebauer, and F. Triki. A stochastic processes toolkit for risk management. *Available at SSRN 1109160*, 2007.

C. Brown, C. Gilroy, and A. Kohen. The effect of the minimum wage on employment and unemployment: a survey. Technical report, National Bureau of Economic Research, 1982.

D. Card. Using regional variation in wages to measure the effects of the federal minimum wage. *Ilr Review*, 46(1):22–37, 1992.

D. Card and A. B. Krueger. Minimum wages and employment: A case study of the fast food industry in new jersey and pennsylvania, 1993.

G. Chodorow-Reich, P. T. Nenov, and A. Simsek. Stock market wealth and the real economy: A local labor market approach. *American Economic Review*, 111(5):1613–57, 2021.

M. Christl, M. Köppl-Turyna, and D. Kucsera. Revisiting the employment effects of minimum wages in europe. *German Economic Review*, 19(4):426–465, 2018.

M. Draca, S. Machin, and J. Van Reenen. Minimum wages and firm profitability. *American economic journal: applied economics*, 3(1):129–51, 2011.

E. F. Fama. Efficient capital markets a review of theory and empirical work. *The Fama Portfolio*, pages 76–121, 2021.

A. R. Fremeth, G. L. Holburn, and B. K. Richter. Bridging qualitative and quantitative methods in organizational research: Applications of synthetic control methodology in the us automobile industry. *Organization Science*, 27(2):462–482, 2016.

F. Gunsilius. Distributional synthetic controls. *arXiv preprint arXiv:2001.06118*, 2020.

D. Neumark and W. Wascher. Minimum wages and employment: A review of evidence from the new minimum wage research, 2006.

M. Opatrny. The impact of the brexit vote on uk financial markets: a synthetic control method approach. *Empirica*, 48(2):559–587, 2021.

R. Riley and C. R. Bondibene. Raising the standard: Minimum wages and firm productivity. *Labour Economics*, 44:27–50, 2017.