Bayesian Distribution Regression*

Weige Huang† Emmanuel S. Tsyawo‡

Second Draft: July, 2018

Abstract

This paper introduces a Bayesian version of distribution regression. Different from using standard distribution regression where one is able to obtain only the entire distribution, we are able to estimate the distribution for the entire distribution of the outcome variable using Bayesian distribution regression. More specifically, our method helps estimate the distribution for any point on the distribution of the outcome variable but not limited to only a point as using standard distribution regression, which makes straightforward inference on the points on the distribution. It is also straightforward to conduct inference on any distributional statistics like quantiles and variance of interest. We develop asymptotic theorems for our estimators. Combined with counterfactual analysis, we are able to compute counterfactual distributional effects and conduct inferences on these effects. Our application of the method to the Fama-French five-factor model demonstrates substantial heterogeneity in the impact of the market return on the distribution of the portfolio return.

JEL classification: C01, C11, C53, G12

Keywords: Distribution Regression, Counterfactual analysis, Monte Carlo, Metropolis-Hastings algorithm

---

*We thank Brantly Callaway, Oleg Rytchkov, Charles E. Swanson for valuable comments. All errors are our own.

†PhD Student, Department of Economics, Temple University. Email: weige.huang@temple.edu

‡PhD Student, Department of Economics, Temple University. Email: estsyawo@temple.edu

1 Introduction

There are two popular approaches to estimate the entire distribution of the outcome variable, which are quantile and distribution regression. Though Foresi and Peracchi (1995) introduced distribution regression, Chernozhukov, Fernández-Val, and Melly (2013a) recently increased its popularity in applied economics. For instance, Richey and Rosburg (2016) and Callaway and Huang (2017) use distribution regression to study intergenerational income mobility. Though using standard distribution regression is able to obtain the entire distribution of an outcome variable conditional on other covariates, one ought to rely on the bootstrap in order to perform inference on the distribution itself or any distributional statistics of interest (say quantile or mean counterfactual effects). To the best of our knowledge, there is no study that permits direct inference by using results obtained from estimating the distribution regression. Inspired by Yu and Moyeed (2001), Schennach (2005) and Lancaster and Jae Jun (2010) who combine bayesian techniques and quantile regression, we develop a bayesian distribution regression method, which leverages the likelihood in a distribution regression framework in order to arrive at a bayesian version of distribution regression.[1] With this method, we are able to obtain the entire distribution of the outcome variable at any distributional statistic (e.g., mean, median, mode or any quantile) and its distribution.

On the relevance of our method, we note that direct inference from the results of bayesian distribution regression. More specifically, we are able to obtain not only the distribution of the outcome variable as obtained by the standard (i.e., frequentist) distribution regression method but the distribution thereof. For non-bayesian distribution regression, one usually employs the bootstrap in order to do the inference whereas ours provides the entire distribution over which inference is performed using bayesian techniques. In addition, the asymptotic (normal) approximation of the posterior distribution obtains as a closed form function of the modes at different points of the outcome. This feature of the approximated posterior enables

[1] We notice Law, Sutherland, Sejdinovic, and Flaxman (2017) invent a method also termed bayesian distribution regression. However, it is in machine learning field and is very different from our method.

the derivation of joint distributions (and inference as a result) of the outcome, counterfactual, distribution and quantile treatment effects at arbitrarily many points of the outcome. Also, where counterfactual or partial effects are of interest, we are able to obtain their distribution or interesting statistics thereof, thus paving the way for tests, not only of means or statistics but of entire distributions.

Though our framework easily lends itself to treatment effect analysis, it ought to be borne in mind that our results do not have the causal interpretation unless the condition of unconfoundedness is satisfied (e.g., Rosenbaum and Rubin (1983), Heckman and Robb (1985), Imbens (2004)). In the absence of any such credible hurdle, the method described in this paper is easily applied to program evaluation and counterfactual effect analysis. It is straightforward to conduct inference because we obtain entire distributions of the effects and using the quantiles as confidence intervals is valid because the efficiency within the MLE framework which is based on the condition of generalised information equality holds (see Chernozhukov and Hong (2003) and Chernozhukov (Fall 2007)).

We apply the method to the Fama and French (2015) five-factor model. We estimate the entire (counterfactual) distribution of the portfolio returns and its confidence intervals. We also study the counterfactual distributional and quantile effects of the changes in the stock market returns on the portfolio returns. The counterfatual analysis is of interesting because it enables the portfolio managers to evaluate the impacts of the market returns on the portfolio they construct in variant scenarios. We show that the effects (measured as the mean or median of the effect) of the market return on the distribution of the monthly portfolio return and the distribution of the effects exhibit considerable heterogeneity.

The rest of the paper is organized as follows. In Section 2, we present the Bayesian distribution regression model and define the (counterfactual) distribution and quantiles of the outcome and treatment effects.[2] In Section 3, we describe the computational algorithm for carrying out bayesian distribution regression. Section 4 presents the asymptotic theory

[2] We use the term *treatment effects* in this paper to denote distribution and quantile treatment effects (Δ^{DE} and Δ^{QE} respectively.)

and joint inference at several points of the distribution of the outcome using the normal approximation of the posterior distribution. We apply the method on the Fama and French (2015) five-factor asset pricing model in Section 5 and section 6 concludes.

2 The model

In this paper, the focus is to develop a bayesian approach to distribution regression. A key ingredient for this task is the likelihood function needed, in addition to the prior distribution, for obtaining the posterior distribution of parameters. We consider only two prior distributions: the multivariate normal and the uniform distribution.

2.1 The likelihood function

Observed data are iid samples $(y_i, \mathbf{x}_i)$ where the dependent variable y is a continuous and the $N \times k$ matrix $\mathbf{x}$ include a treatment variable t and other covariates X. A threshold value $y_o \in \mathcal{Y}$ where $\mathcal{Y} \subset \mathbb{R}$ denotes the support of y enables us to define a binary variable $\check{y}_i^o = \mathbf{1}\{y_i \leq y_o\}$ that equals one if $y_i \leq y_o$ and zero otherwise. For a threshold y_o, the likelihood of an observation i is given by

$$p(\check{y}_i^o|\boldsymbol{\theta}_o) = \Lambda(\mathbf{x}_i, \boldsymbol{\theta}_o)^{\check{y}_i^o}(1 - \Lambda(\mathbf{x}_i, \boldsymbol{\theta}_o))^{1-\check{y}_i^o} = \frac{\exp(\mathbf{x}_i\boldsymbol{\theta}_o\mathbf{1}\{y_i \leq y_o\})}{1 + \exp(\mathbf{x}_i\boldsymbol{\theta}_o)} \tag{2.1}$$

where $\boldsymbol{\theta}_o$ is a $k \times 1$ vector of unknown parameters and $\Lambda(v) = (1 + \exp(-v))^{-1}$ is the logistic link function.[3] The logistic link is mostly preferred because of its analytical form. Alternatively, one can use the normal distribution as link function. Chernozhukov, Fernández-Val, and Melly (2013b, section 3.1.2) notes any link function can approximate the conditional distribution arbitrarily well by using sufficiently rich transformations of $\mathbf{x}$, for example, polynomials, b-splines and tensor products. The joint likelihood at a fixed y_o and vector of

[3] See Koenker and Yoon (2009) for a thorough study of link functions for binary response models.

parameters $\boldsymbol{\theta}_o$ is given by

$$p(\breve{y}^o|\boldsymbol{\theta}_o) = \prod_{i=1}^{N} p(\breve{y}_i^o|\boldsymbol{\theta}_o) = \prod_{i=1}^{N} \frac{\exp(\mathbf{x}_i\boldsymbol{\theta}_o\mathbf{1}\{y_i \leq y_o\})}{1+\exp(\mathbf{x}_i\boldsymbol{\theta}_o)} \tag{2.2}$$

where the vector $\mathbf{x}_i$ includes all covariates t_i and X_i. Distribution regression proceeds by estimating the mode of (2.2) at several thresholds y_o in $\bar{\mathcal{Y}}$ that cover the support of y fairly well. The choice of the finite subset $\bar{\mathcal{Y}}$ in $\mathcal{Y}$ for a continuous y needs to satisfy the condition that the Hausdorf distance between $\bar{\mathcal{Y}}$ and $\mathcal{Y}$ is approaching zero at a rate faster than $1/\sqrt{N}$ (see Chernozhukov, Fernandez-Val, and Weidner (2018, remark 2)). A common choice of prior distribution of $\boldsymbol{\theta}_o$ is the normal prior, $\theta_j \sim N(\mu_j, \sigma_j)$, $j = 1, ..., k$ with $\mu_j = 0$ and $\sigma_j \in [10, 100]$. The resulting prior distribution is

$$p(\boldsymbol{\theta}_o) = \prod_{j=1}^{k} f(\theta_j|\mu_j, \sigma_j) = \prod_{j=1}^{k} \frac{1}{\sqrt{2\pi}\sigma_j} \exp\Big\{-\frac{1}{2}\Big(\frac{\theta_j-\mu_j}{\sigma_j}\Big)\Big\} \tag{2.3}$$

An alternative specification of the prior distribution is the improper uniform prior $p(\theta_j) \propto 1$ for each parameter θ_j. Using (2.2) and (2.3) above, the following posterior distribution of $\boldsymbol{\theta}_o$ obtains

$$p(\boldsymbol{\theta}_o|\breve{y}^o) = \frac{p(\breve{y}^o|\boldsymbol{\theta}_o)p(\boldsymbol{\theta}_o)}{p(\breve{y}^o)} \propto p(\breve{y}^o|\boldsymbol{\theta}_o)p(\boldsymbol{\theta}_o) \tag{2.4}$$

where the proportionality follows because $p(\breve{y}^o)$ does not depend on $\boldsymbol{\theta}_o$. Observe the dependence of $\boldsymbol{\theta}_o$ on y_o via $\breve{y}^o$. In fact, for fairly distinct values[4] of $y_{0,g} \in \bar{\mathcal{Y}}$, $g = 1, ..., G$, the posterior distributions $\{p(\boldsymbol{\theta}_{o,g}|\breve{y}^{o,g})\}_{g=1}^{G}$ are distinct.

2.2 The conditional distribution

Given the above posterior distribution (2.4) of $\boldsymbol{\theta}_o$, it is straighforward to use Markov Chain Monte Carlo (MCMC) methods[5] to obtain draws of $\boldsymbol{\theta}_o$. At a fixed threshold y_o, the

[4]Observe that because only $\breve{y}_i^o = \mathbf{1}\{y_i \leq y_o\}$ varies with y_o, the change in the posterior $p(\boldsymbol{\theta}_o|\breve{y}^o)$ comes from at least one observation's binary $\breve{y}_i^o \in \{0, 1\}$ switching value.

[5]Two options, the Independence Metropolis-Hastings and the Random Walk Metropolis-Hastings algorithms are available in our package R `bayesdistreg`.

unconditional distribution (with respect to $\breve{y}^o$) of y^o_θ is

$$p(y^o_\theta|\boldsymbol{\theta}_o) = \int p(y^o_\theta|\breve{y}^o, \boldsymbol{\theta}_o)p(\breve{y}^o|\boldsymbol{\theta}_o)d\breve{y}^o = \int p(y^o_\theta|\breve{y}^o, \boldsymbol{\theta}_o)p(\boldsymbol{\theta}_o|\breve{y}^o)p(\breve{y}^o)/p(\boldsymbol{\theta}_o)d\breve{y}^o \tag{2.5}$$

and[6]

$$y^o_\theta = \int_{\boldsymbol{X}} F_Y(y_o|\mathbf{x}, \boldsymbol{\theta}_o)p(\mathbf{x})d\mathbf{x} = F_Y(y_o|\boldsymbol{\theta}_o) \quad (y^o_\theta \in (0,1)) \tag{2.6}$$

where $\boldsymbol{X}$ is the support of all covariates X_i, t_i. In practice y^o_θ is computed as $N^{-1}\sum_{i=1}^N \Lambda(\mathbf{x}_i\boldsymbol{\theta}_o)$.[7]

Remark 1. *It is straightforward to extend our method to estimate the counterfactual distributions and compute counterfactual effects. The importance of estimating counterfactual distributions for policy analysis (Stock (1989) and Heckman and Vytlacil (2007)) lies in its ability to uncover heterogeneity in the impact of covariates on the distribution (and by extension the quantile) of the outcome.*

One can obtain the counterfactual distribution of the outcome by replacing t_i with t^c_i and obtain $y^c_\theta = N^{-1}\sum_{i=1}^N \Lambda(t^c_i, X_i, \boldsymbol{\theta}_o)$, where y^c_θ is the counterfactual of y_θ. The counterfactual expression for (2.5) obtains by simply replacing y^o_θ with y^c_θ.

The counterfactual distribution effect at a threshold $y_o \in \mathcal{Y}$ ($\Delta^{DE}_{y_o}$) given by

$$\Delta^{DE}_{y_o} = y^o_\theta - y^{o,c}_\theta = F_Y(y_o|\boldsymbol{\theta}_o) - F_{Y^c}(y_o|\boldsymbol{\theta}_o) \tag{2.7}$$

has the following distribution conditional on $\boldsymbol{\theta}_o$

$$p(\Delta^{DE}_{y_o}|\boldsymbol{\theta}_o) = \int p(\Delta^{DE}_{y_o}|\breve{y}^o, \boldsymbol{\theta}_o)p(\breve{y}^o|\boldsymbol{\theta}_o)d\breve{y}^o = \int p(\Delta^{DE}_{y_o}|\breve{y}^o, \boldsymbol{\theta}_o)p(\boldsymbol{\theta}_o|\breve{y}^o)p(\breve{y}^o)/p(\boldsymbol{\theta}_o)d\breve{y}^o \tag{2.8}$$

Notice that the term after the second equality has the posterior distribution $p(\boldsymbol{\theta}_o|\breve{y}^o)$. A uni-

[6]To implement the distribution regression estimator, we estimate a series of bayesian logit models on a mesh of threshold values. We first specify a series of thresholds y_o and then run each $\mathbf{1}(y \leq y_o)$ on X and t using bayesian logit regression, where y is the observed outcome.

[7]Since the distribution $F_Y(y_o|\boldsymbol{\theta}_o)$ in (2.6) may be non-monotone in y_o, we apply the monotonisation method of Chernozhukov, Fernández-Val, and Galichon (2010) based on rearrangement. In practice, we may think of rearrangement as sorting (Chernozhukov, Fernández-Val, and Galichon (2010), p. 1098).

form prior distribution on $\boldsymbol{\theta}_o$ allows for further simplification of $p(\Delta_{y_o}^{DE}|\boldsymbol{\theta}_o)$ because $p(\boldsymbol{\theta}_o|\check{y}^o) \propto p(\check{y}^o|\boldsymbol{\theta}_o)$. The counterfactual quantile effect at the τ'th quantile of y is given by

$$\begin{aligned} \Delta_\tau^{QE} &= F_Y^{-1}(\tau|\boldsymbol{\theta}_o) - F_{Y^c}^{-1}(\tau|\boldsymbol{\theta}_o) \\ \text{where } F_Y^{-1}(\tau|\boldsymbol{\theta}_o) &= \inf\{y \in \mathcal{Y} : F_Y(y_o|\boldsymbol{\theta}_o) \geq \tau\} \end{aligned} \tag{2.9}$$

is the left inverse of $F_Y(y_o|\boldsymbol{\theta}_o)$, $\tau \in (0,1)$ (see Chernozhukov, Fernández-Val, and Melly (2013b, Appendix A)). $F_{Y^c}^{-1}(\tau|\boldsymbol{\theta}_o)$ is defined analogously. The distribution of the counterfactual quantile effect does not obtain as a direct product of distribution regression at a single index y_o but rather after inverting the entire distribution on $\mathcal{Y}$. This is shown in the next section.

3 Algorithm

In this section, we present an algorithm for the estimation of bayesian distribution regression. This not only makes the practical understanding of it easier but also facilitates computations. To aid computation and applicability, we provide an R package `bayesdistreg`.[8]

Algorithm 1 (Bayesian distribution)**.**

1. *Obtain a grid of threshold values $y_{0,g}$, $g = 1, ..., G$.*
2. *For each $g = 1, ..., G$*
   (a) *Obtain the likelihood function $p(\check{y}^{o,g}|\boldsymbol{\theta}_{o,g})$ using (2.2) where $\check{y}_i^{o,g} = 1\{y_i \leq y_{0,g}\}$.*
   (b) *Multiply the likelihood $p(\check{y}^{o,g}|\boldsymbol{\theta}_{o,g})$ by the prior to obtain the posterior $p(\boldsymbol{\theta}_{o,g}|\check{y}^{o,g}) \propto p(\check{y}^{o,g}|\boldsymbol{\theta}_{o,g})p(\boldsymbol{\theta}_{o,g})$.*
   (c) *Simulate M draws of $\boldsymbol{\theta}_{o,g}$ from the posterior $p(\boldsymbol{\theta}_o|\check{y}^o)$.[9] For each $m = 1, ..., M$*

[8] This package is freely made available for the open source software R. Install using the command `devtools::install_github("estsyawo/bayesdistreg")`.

[9] In the package `bayesdistreg`, the functions `IndepMH` and `RWMH` implement the independence and random-walk chain Metropolis-Hastings algorithms. For a general treatment of posterior simulation techniques, see Gelman, Carlin, Stern, and Rubin (1995, chapter 11)

i. *Make a draw* $\boldsymbol{\theta}^m_{o,g}$ *from the posterior.*

ii. *Compute* $y_{\theta,m} = F_Y(y_{o,g}|\boldsymbol{\theta}^m_{o,g}) = N^{-1}\sum_{i=1}^N \Lambda(\mathbf{x}_i, \boldsymbol{\theta}^m_{o,g})$, *and its counterfactual* $y^c_{\theta,m}$. *and the m'th draw of distribution effect at* $y_{o,g}$, $\Delta^{DE}_{y_{o,g},m} = y_{\theta,m} - y^c_{\theta,m}$

(d) *end m*

(e) *Sort the computed* $y_{\theta,m}$ *and* $y^c_{\theta,m}$ *in ascending order.*

3. *end g*

The $G \times M$ matrices $\boldsymbol{P}^o_\theta$, $\boldsymbol{P}^{o,c}_\theta$ and $\boldsymbol{\Delta}^{DE}$ that obtain constitute draws of y^o_θ, $y^{o,c}_\theta$ and $\Delta^{DE}_{y_o}$ at thresholds $\{y_{o,g}\}_{g=1}^G \subset \bar{\mathcal{Y}}$. In the following additional steps, we show how to obtain the distribution of the quantiles.

Algorithm 2 (Inverse conditional distribution)**.**

1. *Monotonise each column of* $\boldsymbol{P}^o_\theta$ *and* $\boldsymbol{P}^{o,c}_\theta$.[10]

2. *Invert each column of* $\boldsymbol{P}^o_\theta$ *and* $\boldsymbol{P}^{o,c}_\theta$ *and the vector* y_{0g} *by taking the left inverse (see equation (2.9)) to obtain the distribution of the quantiles* $\boldsymbol{Q}^o_\theta$ *and* $\boldsymbol{Q}^{o,c}_\theta$.

3. *Compute the counterfactual quantile treatment effect* $\boldsymbol{\Delta}^{QE} = \boldsymbol{Q}^o_\theta - \boldsymbol{Q}^{o,c}_\theta$

Observe that taking the left inverse can result in discrete mass points if G is small and y is not continuously distributed enough. In our package `bayesdistreg`, we implement spline interpolations in order to ensure smoothness and continuity in the quantile functions $F_Y^{-1}(\tau|\boldsymbol{\theta}_o)$ that are generated. This has the advantage of not restricting the quantile functions to the discrete values $y_{0g} \in \mathcal{Y}$.

4 Asymptotic theory

Large sample theory in Bayesian analysis is often not crucial for inference since Bayesian analysis provides distributions on statistics of interest and inference thereon is direct. However, large sample results are useful and computational convenient approximations. Some

[10] For the avoidance of further notation, we retain the same notation for monotonised $\boldsymbol{P}^o_\theta$ and $\boldsymbol{P}^{o,c}_\theta$.

applications have used the normal approximations of posterior distributions especially when these are relatively more tractable. Rubin and Schenker (1987), Agresti and Coull (1998) and Clogg et al. (1991) are among works that use approximations of posterior distributions for bayesian inference. In order to strengthen the point-wise results obtained in the preceding two sections, we make the following assumption.

Assumption 1 ($\boldsymbol{\theta}_o$ as a function continuous in y_o)**.**
The mode of the posterior distribution $p(\boldsymbol{\theta}_o|\breve{y}^o)$ is a continuous function in y_o, i.e. $\boldsymbol{\theta}_o = \vartheta(y_o)$ for a function $\vartheta : \mathbb{R} \to \mathbb{R}^k$ with $\vartheta(y_o) = \arg\max_{\boldsymbol{\theta}_o} p(\boldsymbol{\theta}_o|\breve{y}^o) = \arg\max_{\boldsymbol{\theta}_o} p(\boldsymbol{\theta}_o|\mathbf{1}\{y \leq y_o\})$.

The goal of assumption (1) is to reduce the infinite-dimensional problem to a set of finite maximisation problems.[11]

Theorem 1 (Asymptotic normality of the posterior distribution)**.** *Under assumption (1), the posterior distribution (2.4) is asymptotically normal with mean vector and covariance matrix continuous in $y_o \in \mathcal{Y}$. The asymptotic distribution is*

$$p(\boldsymbol{\theta}_o|\breve{y}^o) \approx \mathcal{N}(\hat{\boldsymbol{\theta}}_o, [\mathcal{I}(\hat{\boldsymbol{\theta}}_o)]^{-1}) \tag{4.1}$$

where $\mathcal{I}(\hat{\boldsymbol{\theta}}_o) = \sum_{i=1}^{N} \left(\frac{\exp(\mathbf{x}_i\hat{\boldsymbol{\theta}}_o)}{(1+\exp(\mathbf{x}_i\hat{\boldsymbol{\theta}}_o))^2}\right)\mathbf{x}_i\mathbf{x}_i'$ and $\hat{\boldsymbol{\theta}}_o = \arg\max_{\boldsymbol{\theta}_o} p(\boldsymbol{\theta}_o|\mathbf{1}\{y \leq y_o\})$

Proof.

The log of the posterior obtains as

$$L(\boldsymbol{\theta}_o|\breve{y}^o) = \log p(\boldsymbol{\theta}_o|\breve{y}^o) = \sum_{i=1}^{N} \mathbf{x}_i\boldsymbol{\theta}_o I(y_i \leq y_o) - \sum_{i=1}^{N} \log(1 + \exp(\mathbf{x}_i\boldsymbol{\theta}_o)) \tag{4.2}$$

such that $y_o \in \mathcal{Y}$. The score function of $L(\boldsymbol{\theta}_o|\breve{y}^o)$ is given by

$$\mathbf{s}(\boldsymbol{\theta}_o) = \nabla_{\boldsymbol{\theta}_o} L(\boldsymbol{\theta}_o|\breve{y}^o) = \sum_{i=1}^{N} \left(I(y_i \leq y_o) - \frac{\exp(\mathbf{x}_i\boldsymbol{\theta}_o)}{1 + \exp(\mathbf{x}_i\boldsymbol{\theta}_o)}\right)\mathbf{x}_i' \tag{4.3}$$

[11] Example of works which employ splines for approximating infinite dimensional parameters include Wei and Carroll (2009) and Arellano and Bonhomme (2016).

Taking a second derivative with respect to $\boldsymbol{\theta}_o$,

$$\frac{d^2}{d\boldsymbol{\theta}_o} L(\boldsymbol{\theta}_o|\breve{y}^o) = -\sum_{i=1}^{N} \Big(\frac{\exp(\mathbf{x}_i \boldsymbol{\theta}_o)}{(1+\exp(\mathbf{x}_i \boldsymbol{\theta}_o))^2} \Big) \mathbf{x}_i \mathbf{x}_i' \tag{4.4}$$

obtains as the hessian matrix. Notice that the expression in (4.4) above is only dependent in on y_o via $\boldsymbol{\theta}_o$. Taking the Taylor expansion of $L(\boldsymbol{\theta}_o|\breve{y}^o)$ around the mode $\hat{\boldsymbol{\theta}}_o$ gives

$$L(\boldsymbol{\theta}_o|\breve{y}^o) = L(\hat{\boldsymbol{\theta}}_o|\breve{y}^o) + \frac{1}{2}(\boldsymbol{\theta}_o - \hat{\boldsymbol{\theta}}_o)' \Big[\frac{d^2}{d\boldsymbol{\theta}_o} L(\boldsymbol{\theta}_o|\breve{y}^o) \Big] \Big|_{\boldsymbol{\theta}_o = \hat{\boldsymbol{\theta}}_o} (\boldsymbol{\theta}_o - \hat{\boldsymbol{\theta}}_o) + (s.o.) \tag{4.5}$$

where $(s.o.)$ are negligible smaller order terms. The first term is constant and the second is proportional to the logarithm of the multivariate normal density of $\boldsymbol{\theta}_o$ with

$$p(\boldsymbol{\theta}_o|\breve{y}^o) \approx \mathcal{N}(\hat{\boldsymbol{\theta}}_o, [\mathcal{I}(\hat{\boldsymbol{\theta}}_o)]^{-1}) \tag{4.6}$$

where $\mathcal{I}(\hat{\boldsymbol{\theta}}_o) = -\frac{d^2}{d\boldsymbol{\theta}_o} L(\boldsymbol{\theta}_o|\breve{y}^o)$ is the information matrix. By assumption (1), $\boldsymbol{\theta}_o$ is continuous in y_o and by extension, the (asymptotic) posterior distribution (4.6) is a continuous function in y_o. □

The continuity of the approximated posterior (4.6) in y_o has the analytical and computational convenience. The mean and variance terms $I(\boldsymbol{\theta}_o)^{-1}$ only vary by y_o through $\boldsymbol{\theta}_o = \vartheta(y_o)$. For a finite subset $\{y_{o,g}\}_{g=1}^{G} \subset \bar{\mathcal{Y}}$, the set of corresponding posterior distributions obtains as $\{p(\boldsymbol{\theta}_{o,g}|\breve{y}^{o,g})\}_{g=1}^{G}$. Assumption (1) enables us to strengthen this result further for a $y_o \in \mathcal{Y}$ and the posterior at any point $y_o \in \mathcal{Y}$ obtains as a function thereof. The approximated posterior (4.6) can be used in computing $p(y_\theta^o|\boldsymbol{\theta}_o)$ in (2.5) and $p(\Delta_{y_o}^{DE}|\boldsymbol{\theta}_o)$ in (2.8) and this process is computationally faster than the MCMC because the approximated (4.6) is in a closed form.

In the following, corollaries, we push the asymptotic arguments further in order to obtain closed-form expressions for the (joint) distributions of the outcome distributions and the distribution treatment effects.

Theorem 2 (Asymptotic distribution of $y^o_\theta = F_Y(y_o|\boldsymbol{\theta}_o)$).

The asymptotic distribution of $F_Y(y_o|\boldsymbol{\theta}_o)$ following from theorem (1) is normal with both mean and variance continuous in y_o.

$$\hat{F}_Y(y_o|\hat{\boldsymbol{\theta}}_o) \sim \mathcal{N}(F_{y_o}, \mathcal{V}_{F_{y_o}}) \tag{4.7}$$

where $F_{y_o} = \int F_{Y|X}(y_o|\mathbf{x})dF(\mathbf{x}) = \int \Lambda(\mathbf{x}\boldsymbol{\theta}_o)dF(\mathbf{x})$, $\mathcal{V}_{F_{y_o}} = E[\boldsymbol{\lambda}^{\mathbf{x}}_i(\boldsymbol{\theta}_o)'[I(\boldsymbol{\theta}_o)]^{-1}\boldsymbol{\lambda}^{\mathbf{x}}_i(\boldsymbol{\theta}_o)]$ *and* $\boldsymbol{\lambda}^{\mathbf{x}}_i(\boldsymbol{\theta}_o) = \Lambda'(\mathbf{x}_i\boldsymbol{\theta}_o)\mathbf{x}'_i$.

Proof. The above results follow from the delta method[12] and noting the exchangeability of the derivative and the integral which holds under general regular conditions.

$$\begin{aligned}&\sqrt{N}(\hat{F}_Y(y_o|\hat{\boldsymbol{\theta}}_o) - F_Y(y_o|\boldsymbol{\theta}_o)) = \sqrt{N}(\hat{F}_Y(y_o|\hat{\boldsymbol{\theta}}_o) - \hat{F}_Y(y_o|\boldsymbol{\theta}_o) + \hat{F}_Y(y_o|\boldsymbol{\theta}_o) - F_Y(y_o|\boldsymbol{\theta}_o))\\ &= \sqrt{N}(N^{-1}\sum_{i=1}^{N}\Lambda(\mathbf{x}_i\hat{\boldsymbol{\theta}}_o) - N^{-1}\sum_{i=1}^{N}\Lambda(\mathbf{x}_i,\boldsymbol{\theta}_o)) + \sqrt{N}(N^{-1}\sum_{i=1}^{N}\Lambda(\mathbf{x}_i,\boldsymbol{\theta}_o) - F_Y(y_o|\boldsymbol{\theta}_o))\end{aligned} \tag{4.8}$$

The second term converges to zero in probability. Applying the delta method (see Van der Vaart (1998, Chapter 3)) to the first term, we have

$$\sqrt{N}(\hat{F}_Y(y_o|\hat{\boldsymbol{\theta}}_o) - F_Y(y_o|\boldsymbol{\theta}_o)) = N^{-1}\sum_{i=1}^{N}\Lambda'(\mathbf{x}_i\boldsymbol{\theta}_o)\mathbf{x}'_i\sqrt{N}(\hat{\boldsymbol{\theta}}_o - \boldsymbol{\theta}_o) + o_p(1) \tag{4.9}$$

Applying the central limit theorem,

$$\sqrt{N}(\hat{F}_Y(y_o|\hat{\boldsymbol{\theta}}_o) - F_Y(y_o|\boldsymbol{\theta}_o)) \xrightarrow{d} \mathcal{N}(\mathbf{0}, N\mathcal{V}_{F_{y_o}}) \tag{4.10}$$

where $\mathcal{V}_{F_{y_o}} = E[\boldsymbol{\lambda}^{\mathbf{x}}_i(\boldsymbol{\theta}_o)'[I(\boldsymbol{\theta}_o)]^{-1}\boldsymbol{\lambda}^{\mathbf{x}}_i(\boldsymbol{\theta}_o)]$ and $\boldsymbol{\lambda}^{\mathbf{x}}_i(\boldsymbol{\theta}_o) = \Lambda'(\mathbf{x}_i\boldsymbol{\theta}_o)\mathbf{x}'_i$. □

Corollary 1 (Joint asymptotic distribution of $\{\hat{F}_Y(y_{o,g}|\hat{\boldsymbol{\theta}}_{o,g})\}_{g=1}^G$).

Extending results from theorem (2) to the joint distribution at several indices $\{y_{o,g}\}_{g=1}^G \subset \bar{\mathcal{Y}}$, the joint asymptotic distribution of $\hat{\mathbf{F}}_Y = [\hat{F}_Y(y_{o,1}|\hat{\boldsymbol{\theta}}_{o,1}), ..., \hat{F}_Y(y_{o,G}|\hat{\boldsymbol{\theta}}_{o,G})]'$ is joint normally

[12] See (Van der Vaart (1998), chapter 3).

distributed with

$$\hat{\mathbf{F}}_Y \sim \mathcal{N}(\mathbf{F}_Y, \mathbf{\Omega}_{F_{y_o}}) \tag{4.11}$$

where $\mathbf{F}_Y = [F_Y(y_{o,1}|\boldsymbol{\theta}_{o,1}), ..., F_Y(y_{o,G}|\boldsymbol{\theta}_{o,G})]'$,
$\mathbf{\Omega}_{F_{y_o}}$*'s* (g,g)*'th element* $\mathcal{V}_{F_{y_{o,g}}} = E[\boldsymbol{\lambda}_i^{\mathbf{x}}(\boldsymbol{\theta}_{o,g})'[I(\boldsymbol{\theta}_{o,g})]^{-1}\boldsymbol{\lambda}_i^{\mathbf{x}}(\boldsymbol{\theta}_{o,g})]$*, and* (g,h)*'th element* $\mathcal{V}_{F_{y_{g,h}}} = E[\boldsymbol{\lambda}_i^{\mathbf{x}}(\boldsymbol{\theta}_{o,g})'[I(\boldsymbol{\theta}_{o,g})]^{-1}I_i(\boldsymbol{\theta}_{g,h})[I(\boldsymbol{\theta}_{o,g})]^{-1}\boldsymbol{\lambda}_i^{\mathbf{x}}(\boldsymbol{\theta}_{o,g})]$*,* $I_i(\boldsymbol{\theta}_{g,h}) = N^{-1}\mathbf{s}_i(\boldsymbol{\theta}_g)\mathbf{s}_i(\boldsymbol{\theta}_h)'$*. Suppose assumption (1) holds, the above result extends to arbitrarily many indices* $\{y_{o,g}\}_{g=1}^G \subset \mathcal{Y}$ *(or even a continuum).*

Proof. From the score function of the posterior by $\mathbf{s}(\boldsymbol{\theta}_o)$ (4.3), the influence function representation of $\hat{\boldsymbol{\theta}}_o$ (see Wooldridge (2010, equations 12.15 - 12.17)) obtains as

$$\sqrt{N}(\hat{\boldsymbol{\theta}}_o - \boldsymbol{\theta}_o) = [I(\boldsymbol{\theta}_o)]^{-1}N^{-1/2}\sum_{i=1}^N \mathbf{s}_i(\boldsymbol{\theta}_o) + o_p(1) \tag{4.12}$$

Expanding (4.9) using (4.12) obtains

$$\sqrt{N}(\hat{F}_Y(y_{o,g}|\hat{\boldsymbol{\theta}}_{o,g}) - F_Y(y_{o,g}|\boldsymbol{\theta}_{o,g})) = N^{-1}\sum_{i=1}^N \Lambda'(\mathbf{x}_i\boldsymbol{\theta}_{o,g})\mathbf{x}_i'[I(\boldsymbol{\theta}_{o,g})]^{-1}N^{-1/2}\mathbf{s}_i(\boldsymbol{\theta}_{o,g}) + o_p(1) \tag{4.13}$$

Applying the multivariate central limit theorem (see Van der Vaart (1998, Section 2.18)) to $\sqrt{N}[\hat{\mathbf{F}}_Y - \mathbf{F}_Y]' = \sqrt{N}[(\hat{F}_Y(y_{o,1}|\hat{\boldsymbol{\theta}}_{o,1}) - F_Y(y_{o,1}|\boldsymbol{\theta}_{o,1})), ..., (\hat{F}_Y(y_{o,G}|\hat{\boldsymbol{\theta}}_{o,G}) - F_Y(y_{o,G}|\boldsymbol{\theta}_{o,G}))]'$ using the representation in (4.13),

$$\sqrt{N}[\hat{\mathbf{F}}_Y - \mathbf{F}_Y]' \xrightarrow{d} \mathcal{N}(\mathbf{0}, N\mathbf{\Omega}_{F_y}) \tag{4.14}$$

where $\mathbf{\Omega}_{F_{y_o}}$ comprises the following elements: (g,g)'th element $\mathcal{V}_{F_{y_{o,g}}} = E[\boldsymbol{\lambda}_i^{\mathbf{x}}(\boldsymbol{\theta}_{o,g})'[I(\boldsymbol{\theta}_{o,g})]^{-1}\boldsymbol{\lambda}_i^{\mathbf{x}}(\boldsymbol{\theta}_{o,g})]$, (g,h)'th element $\mathcal{V}_{F_{y_{g,h}}} = E[\boldsymbol{\lambda}_i^{\mathbf{x}}(\boldsymbol{\theta}_{o,g})'[I(\boldsymbol{\theta}_{o,g})]^{-1}I_i(\boldsymbol{\theta}_{g,h})[I(\boldsymbol{\theta}_{o,g})]^{-1}\boldsymbol{\lambda}_i^{\mathbf{x}}(\boldsymbol{\theta}_{o,g})]$, and $I_i(\boldsymbol{\theta}_{g,h}) = N^{-1}\mathbf{s}_i(\boldsymbol{\theta}_g)\mathbf{s}_i(\boldsymbol{\theta}_h)'$. □

Theorem 3 (Asymptotic distribution of $\hat{\Delta}_{y_o}^{DE}$)**.**

The distribution effect at a threshold $y_o \in \mathcal{Y}$

$$\hat{\Delta}_{y_o}^{DE} = \hat{F}_Y(y_o) - \hat{F}_Y^c(y_o) \tag{4.15}$$

is normally distributed, $\Delta_{y_o}^{DE} \sim \mathcal{N}(\hat{\Delta}_{y_o}, \mathcal{V}_{\hat{\Delta}_{y_o}})$ *where* $\hat{\Delta} = \Delta(\hat{\boldsymbol{\theta}}_o) = \int(\Lambda(\mathbf{x}\hat{\boldsymbol{\theta}}_o) - \Lambda(\mathbf{x}\boldsymbol{\alpha}'\hat{\boldsymbol{\theta}}_o))dF(\mathbf{x})$, $\mathcal{V}_{\hat{\Delta}_{y_o}} = \Delta'(\hat{\boldsymbol{\theta}}_o)[\mathcal{I}(\hat{\boldsymbol{\theta}}_o)]^{-1}\Delta'(\hat{\boldsymbol{\theta}}_o)'$, $\Delta'(\hat{\boldsymbol{\theta}}_o) = \int(\Lambda'(\mathbf{x}\hat{\boldsymbol{\theta}}_o) - \Lambda'(\mathbf{x}\boldsymbol{\alpha}'\hat{\boldsymbol{\theta}}_o)\boldsymbol{\alpha})\mathbf{x}'dF(\mathbf{x})$ *and* $\boldsymbol{\alpha}$ *is a* $k \times k$ *diagonal matrix that multiplicatively creates a counterfactual of* $\mathbf{x}$*. Further, the distribution of* $\Delta_{y_o}^{DE}$ *obtains as a continuous function of* $y_o \in \mathcal{Y}$*.*

Proof.

$$\begin{aligned}
\sqrt{N}(\Delta_{y_o}^{DE} - \Delta_{y_o}^{DE}) &= \sqrt{N}((\hat{F}_Y(y_o) - \hat{F}_Y^c(y_o)) - (F_Y(y_o) - F_Y^c(y_o))) \\
&= \sqrt{N}(\hat{F}_Y(y_o) - F_Y(y_o)) - \sqrt{N}(\hat{F}_Y^c(y_o) - F_Y^c(y_o)) \\
&= N^{-1}\sum_{i=1}^{N}(\Lambda'(\mathbf{x}_i\boldsymbol{\theta}_o) - \Lambda'(\mathbf{x}_i\boldsymbol{\alpha}'\boldsymbol{\theta}_o)\boldsymbol{\alpha})\mathbf{x}_i'[I(\boldsymbol{\theta}_o)]^{-1}N^{-1/2}\mathbf{s}_i(\boldsymbol{\theta}_o) + o_p(1) \\
&= N^{-1}\sum_{i=1}^{N}\bar{\boldsymbol{\lambda}}_i^{\mathbf{x}}(\boldsymbol{\theta}_o)[I(\boldsymbol{\theta}_o)]^{-1}N^{-1/2}\mathbf{s}_i(\boldsymbol{\theta}_o) + o_p(1)
\end{aligned} \tag{4.16}$$

From the above influence function representation, it follows from the CLT that

$$\sqrt{N}(\hat{\Delta}_{y_o}^{DE} - \Delta_{y_o}^{DE}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, N\mathcal{V}_{\Delta_o^{DE}}) \tag{4.17}$$

where $\mathcal{V}_{\Delta_o^{DE}} = E[\bar{\boldsymbol{\lambda}}_i^{\mathbf{x}}(\boldsymbol{\theta}_o)[I(\boldsymbol{\theta}_o)]^{-1}\bar{\boldsymbol{\lambda}}_i^{\mathbf{x}}(\boldsymbol{\theta}_o)']$

$\square$

Corollary 2 (Joint asymptotic distribution of)**.**
Extending results from theorem (3) to the joint distribution at several indices $\{y_{o,g}\}_{g=1}^{G} \subset \bar{\mathcal{Y}}$*, the joint asymptotic distribution of* $\hat{\boldsymbol{\Delta}}^{DE} = [\hat{\Delta}_{y_{o,1}}^{DE}, ..., \hat{\Delta}_{y_{o,G}}^{DE}]'$ *is joint normally distributed with*

$$\hat{\boldsymbol{\Delta}}^{DE} \xrightarrow{d} \mathcal{N}(\boldsymbol{\Delta}^{DE}, \boldsymbol{\Omega}_{\Delta}) \tag{4.18}$$

where the (g,h) *element of* $\boldsymbol{\Omega}_\Delta$ *is* $E[\bar{\boldsymbol{\lambda}}_i^{\mathbf{x}}(\boldsymbol{\theta}_{o,g})'[I(\boldsymbol{\theta}_{o,g})]^{-1}I_i(\boldsymbol{\theta}_{g,h})[I(\boldsymbol{\theta}_{o,g})]^{-1}\bar{\boldsymbol{\lambda}}_i^{\mathbf{x}}(\boldsymbol{\theta}_{o,g})]$

Proof. This result obtains by applying the multivariate central limit theorem (see Van der Vaart (1998, Section 2.18)) to $\sqrt{N}(\hat{\boldsymbol{\Delta}}^{DE} - \boldsymbol{\Delta}^{DE}) = \sqrt{N}[(\hat{\Delta}^{DE}_{y_{o,1}} - \Delta^{DE}_{y_{o,1}}), ..., (\hat{\Delta}^{DE}_{y_{o,G}} - \Delta^{DE}_{y_{o,G}})]'$. Using the representation in (4.16),

$$\sqrt{N}(\hat{\boldsymbol{\Delta}}^{DE} - \boldsymbol{\Delta}^{DE}) \xrightarrow{d} \mathcal{N}(\boldsymbol{\Delta}^{DE}, N\boldsymbol{\Omega}_\Delta) \tag{4.19}$$

□

5 An application: counterfactual effects on portfolio returns

In this section, we apply our approach to Fama and French (2015) five-factor model to estimate the (counterfactual) distribution and quantiles of portfolio returns. We also estimate the counterfactual quantile and distributional effects of the changes in the stock market return on the portfolio returns.

5.1 Fama-French five-factor model

Much evidence has been found that average stock returns are related to the entire stock market performance (Sharpe (1964), Lintner (1965), Breeden (1979)). Stock returns are also related to the stock size, value, investment and profitability ratio.[13] Inspired by the evidence of Novy-Marx (2013) and Titman, Wei, and Xie (2004), Fama and French (2015) establish

[13]Size is measured by market capitalization, price times shares outstanding. Value is book-to-market equity ratio, B/M. See Banz (1981), Basu (1983) and Rosenberg, Reid, and Lanstein (1985) for evidence on size and value; Aharoni, Grundy, and Zeng (2013) for investment; Novy-Marx (2013) for profitability. See Breeden, Gibbons, and Litzenberger (1989), Reinganum (1981), Haugen and Baker (1996), Cohen, Gompers, and Vuolteenaho (2002), Fairfield, Whisenant, and Yohn (2003), Titman, Wei, and Xie (2004),Fama and French (2006), Fama and French (2008), Hou, Xue, and Zhang (2015), Fama and French (2016) and Fama and French (2017) for more evidence.

the five-factor model by adding profitability and investment factors to the Fama and French (1993) three-factor model.

Fama and French's five-factor model is as follows,

$$R_{it} = a_i + b_i Mkt_t + s_i SMB_t + h_i HML_t + r_i RMW_t + c_i CMA_t + e_{it} \tag{5.1}$$

In Equation 5.1, R_{it} is the excess return on portfolio i for period t, which is equal to the return minus the riskfree return.[14] Mkt_t is the excess return on the value-weighted market portfolio, which is equal to the market return minus the riskfree return. SMB_t is the Size factor, which is the return on a diversified portfolio of small stocks minus the return on a diversified portfolio of big stocks. HML_t denotes the value factor, which is the difference between the returns of high and low B/M stocks. RMW_t represents the profitability factor, which is the difference between the returns on diversified portfolios of stocks with robust and weak profitability. CMA_t denotes the investment factor, which is the difference between the returns on diversified portfolios of the stocks of low and high (conservative and aggressive) investment firms and e_{it} is the error term.

5.2 Data and results

5.2.1 Data

We use Fama and French (2015) US dataset, which can be downloaded from Kenneth R. French's website. We use monthly dataset which ranges from July 1963 to Feb 2018, 656 months. The factors chosen in this paper are constructed from 2×3 sorts on Size and book-to-market equity ratio or profitability or investment in the United States. The dependent variable is the monthly excess returns on portfolios formed from 2×3 sorts on Size and book-to-market equity ratio or profitability or investment.[15] We choose value-weighted portfolio returns. Thus, we have 18 portfolios in all. To save space, we show the

[14] For monthly data, the riskfree rate is one-month Treasury bill rate.

[15] See Fama and French (2015) for the detail about constructions of portfolio returns and factors.

results on one of the 18 portfolios, which is the one whose return is best explained by the five-factor model, i.e. with highest adjusted R^2. With this criterion, we choose the portfolio of small stocks with lowest profitability. Table 1 shows summary statistics for the portfolio examined and five factors. The average monthly excess return on the portfolio is 0.615%, which is economically significant although statistically insignificant. The factor returns are all also economically significant but statistically insignificant.

Table 1: Summary statistics

	Portfolio	Mkt	SMB	HML	RMW	CMA
Mean	0.615	0.532	0.246	0.341	0.249	0.282
SD	6.519	4.389	3.033	2.815	2.217	2.007

Note: The table shows summary statistics for the monthly excess returns on the portofolio examined in this paper and five factors formed from 2×3 sorts.
Sources: Kenneth R. French's website

Table 2: Regression results

Independent Variable	Constant	Mkt	SMB	HML	RMW	CMA
Coefficient	-0.033	1.021	0.972	-0.058	-0.503	0.033
Standard Error	0.028	0.007	0.009	0.013	0.013	0.019

Note: The table shows the regression results by running the Fama and French's five-factor model.
Sources: Kenneth R. French's website

5.2.2 Results

Table 2 shows the results by running the Fama and French's five-factor model. We can see the coefficients on all factors, which are 1.021 for the market factor, 0.972 for SMB, -0.058 for HML, -0.503 for RMW, and 0.033 for CMA. Notice that the coefficients on the market, size, value and profitability factors are significant. However, the coefficients on constant and investment factor are small and statistically insignificant. The coefficient on the market factor is very close to 1, thus we expect the mean return on the portfolio will comove perfectly with the market return conditional on the other factors. Using bayesian distribution regression, we are able to examine how the quantiles and the distribution of returns respond to changes in the market return.[16]

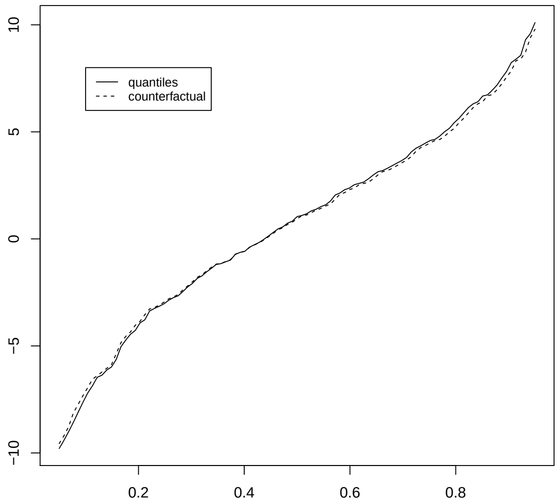
Figure 1 shows the (counterfactual) quantiles of the portfolio returns, in which the counterfactual quantiles are the quantiles that would prevail when the stock market declines by 5%. The top left panel plots the quantiles of the portfolio return and its 95% confidence interval. The bottom right panel plots the counterfactual quantiles of the portfolio return and 95% confidence interval.[17] The bottom right panel plots the quantile effects of 5% decline in the market on the portfolio returns. In the Figure, the horizontal axis shows the quantiles (5%-95%) and the vertical axis shows the quantile returns or quantile effects. The bottom right panel 1d shows the quantile effects for the lower or higher quantiles are statistically and economically significant, although these effects from at $\tau \in [0.31, 0.46]$ include 0 in the 95% confidence band. For instance, the effects of 5% decline in the stock market on the 20% percentile is -.16% per month, i.e., about -1.92% per year.

Figure 2 shows the effects (in percentages) of the counterfactual changes in the stock market (from a decline by 20% to an increase by 20%) on the quantiles of the portfolio return. Panel 2a shows that the 5th quantile of portfolio returns increases as the market return increases. More specifically, the 5th quantile of the return on the selected portfolio decreases as the market return declines and increases as the market return increases.[18] However, Panel 2b and 2c show the opposite applies to the median and 95th quantile portfolio return. For instance, the 5th quantile return would go down by 0.22% if the market return decreases by 5%. However, the median and 95th quantile portfolio return would go up by 0.09% and 0.34% if the market declines by 5%. These results imply the heterogeneous impact of the market return on the distribution of the portfolio return.

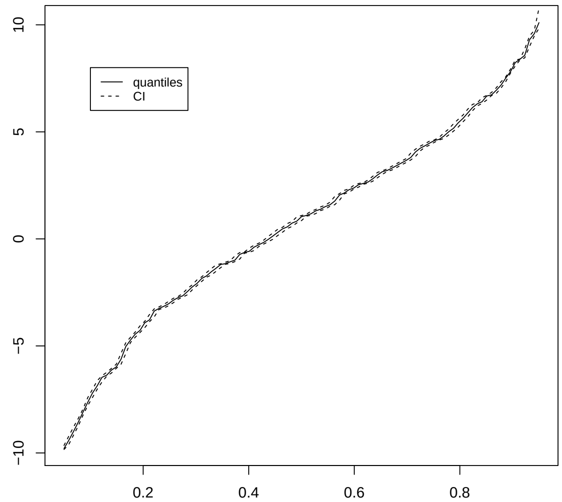
[16]For the effects on portfolio return, it is more understandable to interpret from the perspective of quantile effects. Therefore, we focus on the quantile effects in the main text. The distribution effects results are also provided in the Appendix A.1.

[17]We do not plot the distribution of a distribution which is a three dimensional figure. Instead, we plot 95% confidence interval, which is simply the 5th and 95th quantile of the distribution of the distribution.

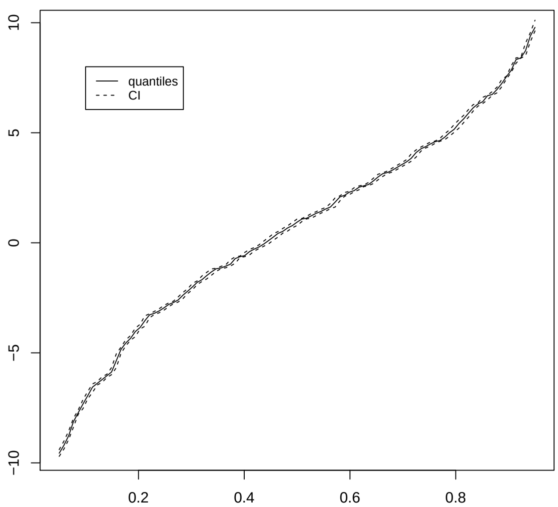
[18]Note that 0.8 denotes the market decline by 20%.



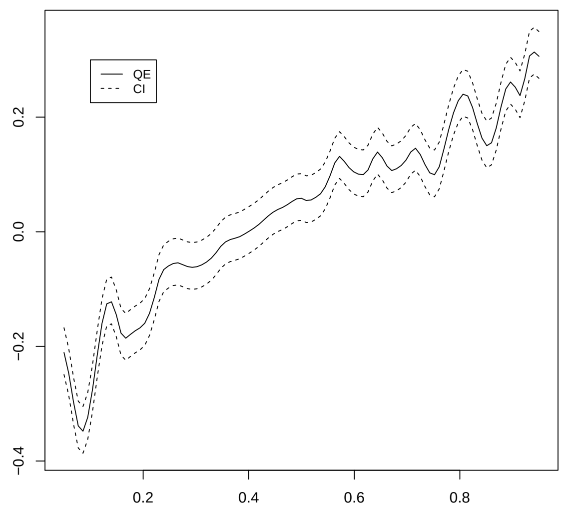
(a) Quantiles and counterfactual



(b) Quantiles



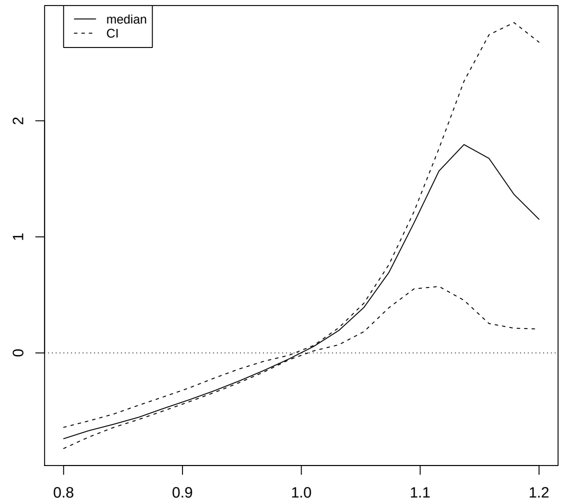
(c) Couterfactual quantiles



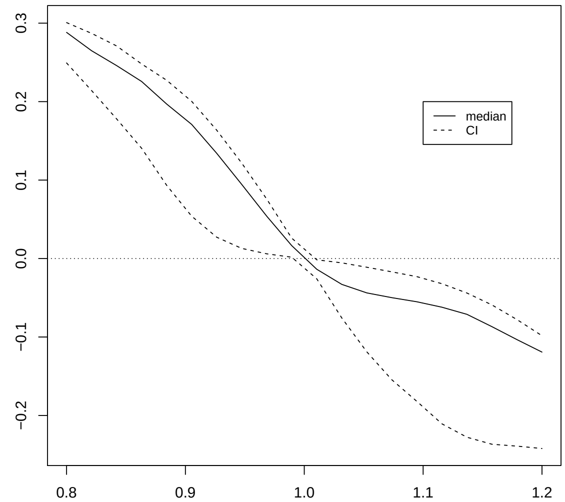
(d) Quantile effects

Figure 1: (Counterfactual) quantiles and quantile effects

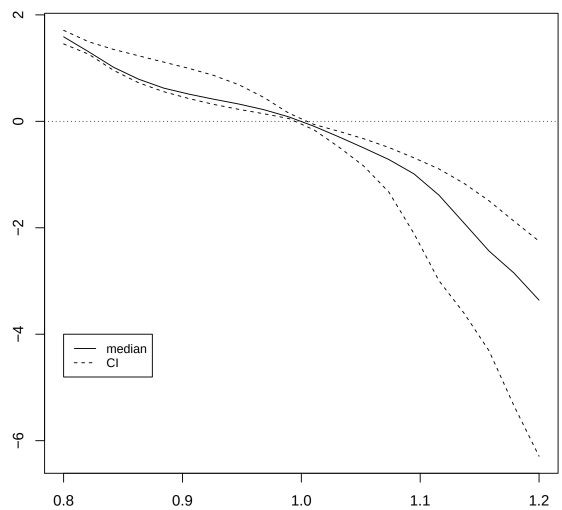
Notes: The top left panel plots the (counterfactual) quantiles of the portfolio return, in which the counterfactual quantiles are the quantiles that would prevail when the stock market declines by 5%. The top left panel plots the quantiles of the portfolio return and its 95% confidence interval. The bottom right panel plots the counterfactual quantiles of the portfolio return and 95% confidence interval. The bottom right panel plots the quantile effects of 5% decline in the market on the portfolio returns.
Sources: Kenneth R. French's website



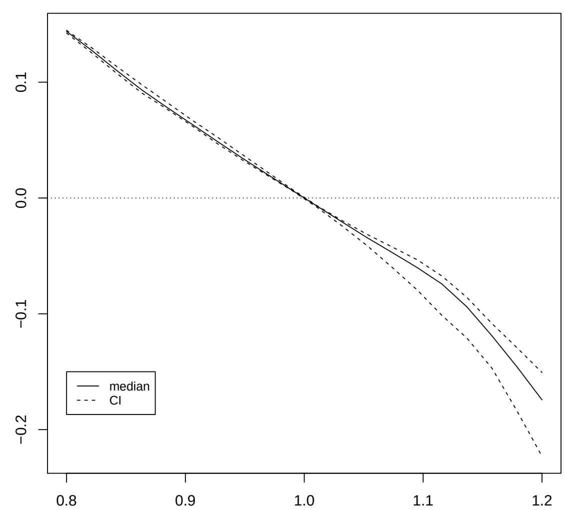
(a) 5%



(b) Median



(c) 95%



(d) Mean effects

Figure 2: Quantile treatment effects

Notes: The figure plots the quantile treatment effects of the stock market changes on the quantiles of portfolio returns. The top left panel plots the quantile effect of market on the 5% quantile of portfolio returns and 95% confidence interval. The top right panel plots the quantile effect on the median of portfolio returns and 95% confidence interval. The bottom left panel plots the quantile effect on the 95% quantile of portfolio returns and 95% confidence interval and the bottom right plots the effects on the mean.
Sources: Kenneth R. French's website

6 Conclusion

In sum, we show the Bayesian approach to distribution regression by leveraging the logit likelihood function at a grid of points on the support of the outcome. The method helps obtain the distribution of a distribution of an outcome variable from which inference follows. With the entire distribution, any distributional statistic, say quantiles, can be computed directly. Combined with counterfactual analysis, the counterfactual distribution or quantiles are obtained and counterfactual effects can be computed. The asymptotic approximation of the posterior distribution enables us to obtain a joint normal distribution of the parameters points on the support of the outcome variable. Approximation offers computational and analytical convenience to our bayesian framework for distribution regression. Under continuity assumptions, we strengthen this result further by obtaining joint normal distributions at arbitrarily many points on the support of the dependent variable beyond the points at which the original estimation is done. In our empirical application, we demonstrated the heterogeneity in the distribution of the effect of the market return on the portfolio return.

References

[1] Agresti, Alan and Brent A Coull. “Approximate is better than ?exact? for interval estimation of binomial proportions”. *The American Statistician* 52.2 (1998), pp. 119–126.

[2] Aharoni, Gil, Bruce Grundy, and Qi Zeng. “Stock returns and the Miller Modigliani valuation formula: Revisiting the Fama French analysis”. *Journal of Financial Economics* 110.2 (2013), pp. 347–357.

[3] Arellano, Manuel and Stéphane Bonhomme. “Nonlinear panel data estimation via quantile regressions”. *The Econometrics Journal* 19.3 (2016).

[4] Banz, Rolf W. “The relationship between return and market value of common stocks”. *Journal of financial economics* 9.1 (1981), pp. 3–18.

[5] Basu, Sanjoy. “The relationship between earnings’ yield, market value and return for NYSE common stocks: Further evidence”. *Journal of financial economics* 12.1 (1983), pp. 129–156.

[6] Breeden, Douglas T. “An intertemporal asset pricing model with stochastic consumption and investment opportunities”. *Journal of financial Economics* 7.3 (1979), pp. 265–296.

[7] Breeden, Douglas T, Michael R Gibbons, and Robert H Litzenberger. “Empirical tests of the consumption-oriented CAPM”. *The Journal of Finance* 44.2 (1989), pp. 231–262.

[8] Callaway, Brantly and Weige Huang. “Counterfactual Analysis of Intergenerational Income Mobility”. Working paper. 2017.

[9] Chernozhukov, Victor. “course materials for 14.385 Nonlinear Econometric Analysis”. MIT OpenCourseWare (http://ocw.mit.edu), Massachusetts Institute of Technology. Downloaded on [05 10 2017]. Fall 2007.

[10] Chernozhukov, Victor, Iván Fernández-Val, and Alfred Galichon. "Quantile and probability curves without crossing". *Econometrica* 78.3 (2010), pp. 1093–1125.

[11] Chernozhukov, Victor, Iván Fernández-Val, and Blaise Melly. "Inference on counterfactual distributions". *Econometrica* 81.6 (2013), pp. 2205–2268.

[12] Chernozhukov, Victor, Iván Fernández-Val, and Blaise Melly. "Inference on counterfactual distributions". *Econometrica* 81.6 (2013), pp. 2205–2268.

[13] Chernozhukov, Victor, Ivan Fernandez-Val, and Martin Weidner. "Network and panel quantile effects via distribution regression". *arXiv preprint arXiv:1803.08154* (2018).

[14] Chernozhukov, Victor and Han Hong. "An MCMC approach to classical estimation". *Journal of Econometrics* 115.2 (2003), pp. 293–346.

[15] Clogg, Clifford C, Donald B Rubin, Nathaniel Schenker, Bradley Schultz, and Lynn Weidman. "Multiple imputation of industry and occupation codes in census public-use samples using Bayesian logistic regression". *Journal of the American Statistical Association* 86.413 (1991), pp. 68–78.

[16] Cohen, Randolph B, Paul A Gompers, and Tuomo Vuolteenaho. "Who underreacts to cash-flow news? Evidence from trading between individuals and institutions". *Journal of financial Economics* 66.2 (2002), pp. 409–462.

[17] Fairfield, Patricia M, J Scott Whisenant, and Teri Lombardi Yohn. "Accrued earnings and growth: Implications for future profitability and market mispricing". *The accounting review* 78.1 (2003), pp. 353–371.

[18] Fama, Eugene F and Kenneth R French. "Common risk factors in the returns on stocks and bonds". *Journal of financial economics* 33.1 (1993), pp. 3–56.

[19] Fama, Eugene F and Kenneth R French. "Profitability, investment and average returns". *Journal of Financial Economics* 82.3 (2006), pp. 491–518.

[20] Fama, Eugene F and Kenneth R French. "Dissecting anomalies". *The Journal of Finance* 63.4 (2008), pp. 1653–1678.

[21] Fama, Eugene F and Kenneth R French. "A five-factor asset pricing model". *Journal of Financial Economics* 116 (2015), pp. 1–22.

[22] Fama, Eugene F and Kenneth R French. "Dissecting anomalies with a five-factor model". *The Review of Financial Studies* 29.1 (2016), pp. 69–103.

[23] Fama, Eugene F and Kenneth R French. "International tests of a five-factor asset pricing model". *Journal of Financial Economics* 123.3 (2017), pp. 441–463.

[24] Foresi, Silverio and Franco Peracchi. "The conditional distribution of excess returns: An empirical analysis". *Journal of the American Statistical Association* 90.430 (1995), pp. 451–466.

[25] Gelman, Andrew, John B Carlin, Hal S Stern, and Donald B Rubin. *Bayesian data analysis*. Chapman and Hall/CRC, 1995.

[26] Haugen, Robert A and Nardin L Baker. "Commonality in the determinants of expected stock returns". *Journal of Financial Economics* 41.3 (1996), pp. 401–439.

[27] Heckman, James J and Richard Robb. "Alternative methods for evaluating the impact of interventions: An overview". *Journal of econometrics* 30.1 (1985), pp. 239–267.

[28] Heckman, James J and Edward J Vytlacil. "Econometric evaluation of social programs, part I: Causal models, structural models and econometric policy evaluation". *Handbook of econometrics* 6 (2007), pp. 4779–4874.

[29] Hou, Kewei, Chen Xue, and Lu Zhang. "Digesting anomalies: An investment approach". *The Review of Financial Studies* 28.3 (2015), pp. 650–705.

[30] Imbens, Guido W. "Nonparametric estimation of average treatment effects under exogeneity: A review". *Review of Economics and statistics* 86.1 (2004), pp. 4–29.

[31] Koenker, Roger and Jungmo Yoon. “Parametric links for binary choice models: A Fisherian–Bayesian colloquy”. *Journal of Econometrics* 152.2 (2009), pp. 120–130.

[32] Lancaster, Tony and Sung Jae Jun. “Bayesian quantile regression methods”. *Journal of Applied Econometrics* 25.2 (2010), pp. 287–307.

[33] Law, Ho Chung Leon, Dougal J Sutherland, Dino Sejdinovic, and Seth Flaxman. “Bayesian Distribution Regression”. *stat* 1050 (2017), p. 11.

[34] Lintner, John. “The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets”. *The review of economics and statistics* (1965), pp. 13–37.

[35] Novy-Marx, Robert. “The other side of value: The gross profitability premium”. *Journal of Financial Economics* 108.1 (2013), pp. 1–28.

[36] Reinganum, Marc R. “A new empirical perspective on the CAPM”. *Journal of financial and quantitative analysis* 16.4 (1981), pp. 439–462.

[37] Richey, Jeremiah and Alicia Rosburg. “Understanding intergenerational economic mobility by decomposing joint distributions” (2016).

[38] Rosenbaum, Paul R. and Donald B. Rubin. “The Central Role of the Propensity Score in Observational Studies for Causal Effects”. *Biometrika* 70.1 (1983), pp. 41–55. ISSN: 00063444. URL: http://www.jstor.org/stable/2335942.

[39] Rosenberg, Barr, Kenneth Reid, and Ronald Lanstein. “Persuasive evidence of market inefficiency”. *The Journal of Portfolio Management* 11.3 (1985), pp. 9–16.

[40] Rubin, Donald B and Nathaniel Schenker. “Logit-based interval estimation for binomial data using the Jeffreys prior”. *Sociological methodology* (1987), pp. 131–144.

[41] Schennach, Susanne M. “Bayesian exponentially tilted empirical likelihood”. *Biometrika* 92.1 (2005), pp. 31–46.

[42] Sharpe, William F. "Capital asset prices: A theory of market equilibrium under conditions of risk". *The journal of finance* 19.3 (1964), pp. 425–442.

[43] Stock, James H. "Nonparametric policy analysis". *Journal of the American Statistical Association* 84.406 (1989), pp. 567–575.

[44] Titman, Sheridan, KC John Wei, and Feixue Xie. "Capital investments and stock returns". *Journal of financial and Quantitative Analysis* 39.4 (2004), pp. 677–700.

[45] Van der Vaart, Aad W. *Asymptotic statistics*. Vol. 3. Cambridge university press, 1998.

[46] Wei, Ying and Raymond J Carroll. "Quantile regression with measurement error". *Journal of the American Statistical Association* 104.487 (2009), pp. 1129–1143.

[47] Wooldridge, Jeffrey M. *Econometric analysis of cross section and panel data*. MIT press, 2010.

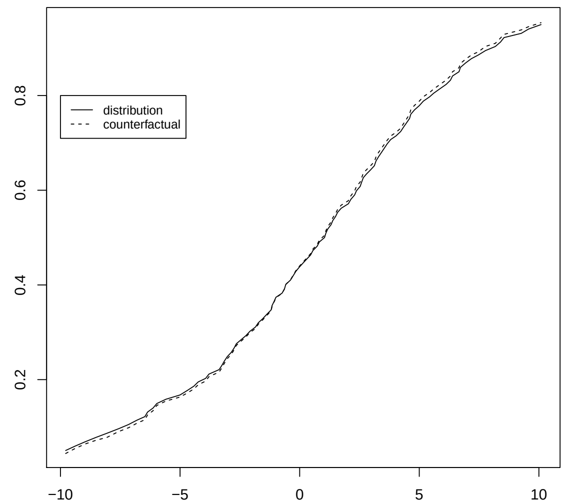
[48] Yu, Keming and Rana A Moyeed. "Bayesian quantile regression". *Statistics & Probability Letters* 54.4 (2001), pp. 437–447.

A Appendix

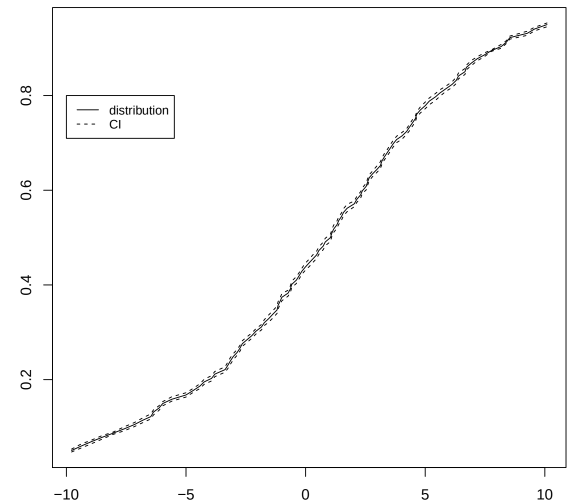
A.1 Distribution effects

Figure 3 shows the (counterfactual) distribution of the portfolio return and the distribution effects of a market decline by 5%. Panel 3a plots the distribution and its conterfactual. Panel 3b plots the distribution and its 95% confidence band. Panel 3c plots the counterfactual distribution and its 95% pointwise confidence band. Panel 3d plots the counterfactual distribution effects of a 5% decline in the market return (in percentages). For instance, we can see that the probability of the portfolio return being less -5% would decrease by 0.45% if the market return decreases by 5%.

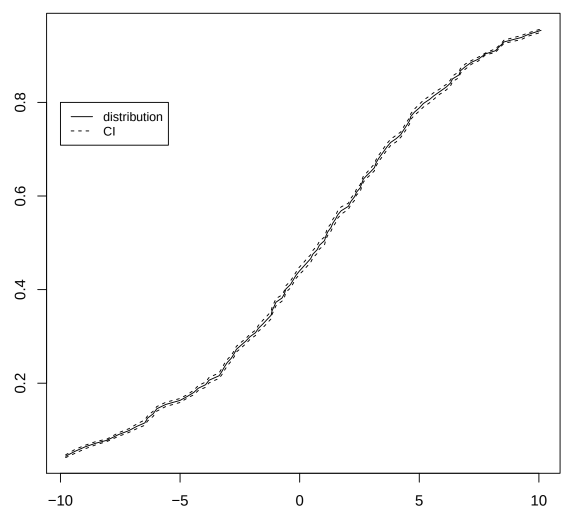
Figure 4 shows the counterfactual distribution effects (in percentages) of changes in the market returns (from a decline by 20% to an increase by 20%). Panel 4a shows that the probability of the portfolio returns being less than 5th percentile of the portfolio return, which is -9.79% per month, is 1.2% when the market return declines by 10%. However, Panel 4c shows that the probability of the portfolio returns being less than 95th percentile of the portfolio return, which is 10.09% per month, is about -0.76% when the market return declines by 10%. Although much less significant, Panel 4b shows that the probability of the portfolio returns being less than median of the portfolio return, which is about 1.04% per month, is -0.96% when the market return declines by 10%. The opposite roughly applies in the case of 10% increase in the market return. These results show the asymmetric impacts of the market returns on the distribution of the portfolio return.



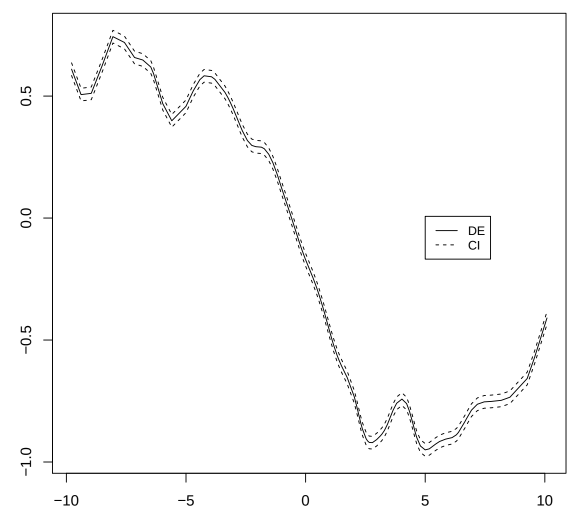
(a) Distibution and counterfactual



(b) Distribution



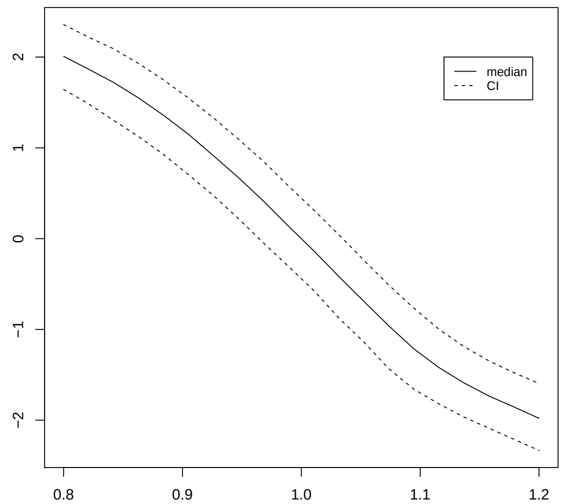
(c) Couterfactual distribution



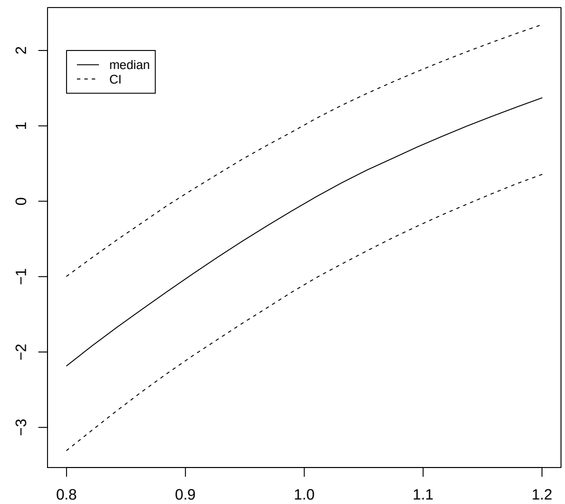
(d) Distribution effects

Figure 3: (Counterfactual) distribution and distribution effects

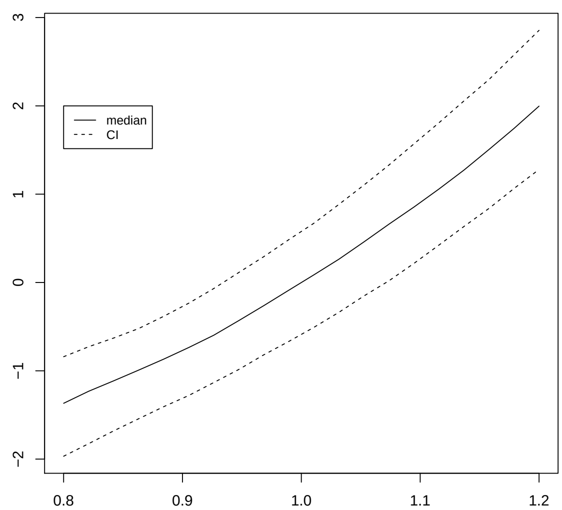
Notes: The top left panel plots the (counterfactual) distribution of the portfolio return, in which the counterfactual distribution is the distribution that would prevail when the stock market declines by 5%. The top right panel plots the distribution of the portfolio return and its 95% confidence interval. The bottom left panel plots the counterfactual distribution of the portfolio return and 95% confidence interval. The bottom right panel plots the distribution effects (in percentage) of 5% decline in the stock market and 95% confidence interval.
Sources: Kenneth R. French's website



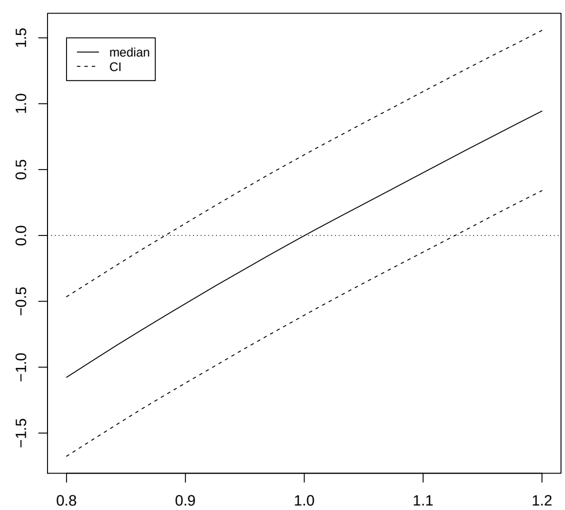
(a) 5%



(b) Median



(c) 95%



(d) Mean effects

Figure 4: Distribution effects

Notes: The figure plots the distribution effects (in percentage) of the stock market changes on the quantiles of portfolio returns. The top left panel plots the distribution effect of market on the 5% quantile of portfolio returns and 95% confidence interval. The top right panel plots the distribution effect on the median of portfolio returns and 95% confidence interval. The bottom left panel plots the distribution effect on the 95% quantile of portfolio returns and the bottom right plots the distribution effect on the mean.
Sources: Kenneth R. French's website