Bayesian Distribution Regression*

Weige Huang[†] Emmanuel S. Tsyawo[‡]

First Draft: October 2017

This Draft: August 2018

Abstract

This paper introduces a Bayesian version of distribution regression that enables inference on estimated distributions, quantiles, distributional effects, etc. Combined with counterfactual analysis, we are able to compute distribution and quantile effects and conduct inference thereon. The Bayesian asymptotic theory we develop extend these gains to computational time and tractability of the estimated distributions. 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 ours.

[†]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

Distribution regression and quantile regression are popular approaches for the estimation of distributions, quantiles, and distributional effects[1] on the outcome variable. Though Foresi and Peracchi (1995) introduced distribution regression, Chernozhukov, Fernández-Val, and Melly (2013a) recently increased its popularity in applied economics. Examples of works that employ distribution regression include Doorley and Sierminska (2012), Shepherd et al. (2013), Wüthrich (2015), Han, Lutz, and Sand (2016), Richey and Rosburg (2016), Chernozhukov, Fernandez-Val, Melly, and Wüthrich (2016) Dube (2017), Callaway and Huang (2017), Chernozhukov, Fernández-Val, Han, and Kowalski (2018), and Chernozhukov, Fernandez-Val, and Weidner (2018). 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 for inference on the distribution itself or any distributional statistic, say quantile or mean counterfactual effects. Our approach enables direct inference using Bayesian techniques[2] that only requires parameter estimates and data.

Inspired by Yu and Moyeed (2001), Schennach (2005) and Lancaster and Jae Jun (2010) who develop Bayesian quantile regression based on working likelihoods, we develop a Bayesian distribution regression (BDR) method, which leverages the likelihood in a distribution regression framework in order to arrive at a Bayesian version of distribution regression.[3] With this method, we are able to perform inference on any distributional statistic or functional.

On the relevance of our method, we note that direct inference results from Bayesian distribution regression. More specifically, we are able to obtain not only the distribution of the outcome variable as an estimand, as obtained by the standard (i.e., frequentist) distribution regression, but the (posterior) density thereof. For non-Bayesian distribution regression, one usually employs the bootstrap in order to perform inference whereas ours provides the

[1] Distributional effects considered in this paper include distribution effects and quantile effects.

[2] We use Markov Chain Monte Carlo techniques and asymptotic Bayesian approximations.

[3] We notice Law, Sutherland, Sejdinovic, and Flaxman (2017) developed a machine learning method termed Bayesian distribution regression. It is, however, very different from our method.

entire distribution over which inference is performed using Bayesian posterior inference. 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 the derivation of joint distributions (and inference as a result) of the outcome, counterfactual, distribution and quantile treatment effects among other estimands of interest at arbitrarily many points of the outcome. Also, where counterfactual or partial effects are of interest, we are able to obtain their distribution thus paving the way for tests, not only of means, median, standard deviations, etc. but of entire distributions.

Bayesian distribution regression can be viewed as an alternative to Bayesian quantile regression.[4] Beyond the advantages of doing Bayesian inference[5] on distributions estimated using distribution regression, Bayesian distribution regression that we propose enables us to carry over to the Bayesian framework, the gains in using distribution regression in general. As noted by Chernozhukov, Fernández-Val, and Melly (2013a), distribution regression (DR), unlike quantile regression, does not require smooth conditional distributions; it handles discrete, continuous or mixed outcome variables fairly well.[6] DR allows heterogeneity in the impact of covariates on the outcome at different points of the distribution (Chernozhukov, Fernandez-Val, Melly, and Wüthrich, 2016). Chernozhukov, Fernández-Val, and Melly (2013a, Remark 3.1) shows that distribution regression involves simpler steps[7] in computing the distribution but not the quantile function. Bayesian distribution regression unifies the convenience of distribution regression and Bayesian inference.

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

[4]See Yang and Wang (2017) for a review of Bayesian quantile regression methods derived from different types of working likelihoods.

[5]Bayesian inference is conditional on the data and is exact. It allows the incorporation of prior information in a logical way that follows from Bayes' theorem. For a thorough treatment of the advantages and disadvantages of Bayesian analysis, see Berger (2013, sections 4.1 and 4.12).

[6]In their simulation exercise in appendix SB of the supplemental material, Chernozhukov, Fernández-Val, and Melly (2013a) show that quantile regression is only more accurate with continuous conditional distribution and performs worse in the presence of mass points.

[7]Distribution regression involves a convenient functional form that involves neither inversion or trimming. See Chernozhukov, Fernández-Val, and Melly (2013a, Remark 3.1).

of unconfoundedness is satisfied (e.g., Rosenbaum and Rubin (1983), Heckman and Robb (1985), Imbens (2004)). In the absence of any such hurdle, the method described in this paper can be 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 in the maximum likelihood 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 bands. We also study the counterfactual distribution and quantile effects of the changes in the stock market returns on portfolio returns. Counterfactual analysis is important because it enables portfolio managers to evaluate the impact of market returns on portfolios they construct. We show that the effects (measured as the mean or median) 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, define the (counterfactual) distribution and quantiles of the outcome and treatment effects[8] and outline the estimation algorithm. Tools of Bayesian inference developed in Section 3 enable simultaneous inference on conditional distributions, distribution effects and quantile effects. Section 4 presents the asymptotic theory and joint inference at several points of the distribution of the outcome using the normal approximation of the posterior distribution. For an empirical application, we study the Fama and French (2015) five-factor asset pricing model in section 5 using our method and conclude in section 6.

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

2 The model

Our 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. To keep the analysis as close as possible to frequentist results of the semi-parametric distribution regression, we assume non-informative uniform prior distributions on the parameters.

2.1 Likelihood and posterior

Observed data are iid samples $(y_i, \mathbf{x}_i)$ where the dependent variable y is continuous and the $N \times k$ matrix $\mathbf{x}$ includes a treatment variable t (in the case of treatment effect analysis) 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 $\breve{y}_i^o = \mathbb{1}\{y_i \leq y_o\}$ that equals one if $y_i \leq y_o$ and zero otherwise. Distribution regression involves running $\breve{y}^o$ at several thresholds $y_o \in \mathcal{Y}$. For a threshold value y_o, the likelihood of an observation i is given by

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

where $\boldsymbol{\theta}_o$ is a $k \times 1$ vector of unknown parameters, $\mathit{\Lambda}(v) = (1 + \exp(-v))^{-1}$ is the logistic link function and the conditional distribution is $F_Y(y_o|\mathbf{x}) = \mathit{\Lambda}(\mathbf{x}\boldsymbol{\theta}_o)$.[9] The logistic link is mostly preferred because of its analytical form. Alternatively, one can use the normal distribution as link function. As 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 parameters $\boldsymbol{\theta}_o$ is given by

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

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

For notational ease, we suppress covariates $\mathbf{x}$. Distribution regression proceeds by maximising (eq. (2.2)) at several thresholds y_o in $\bar{\mathcal{Y}} \in \mathcal{Y}$ 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)). Common choices of prior distribution of $\boldsymbol{\theta}_o$ include the normal prior, $\theta_j \sim N(\mu_j, \sigma_j)$, $j = 1, ..., k$ and the improper uniform prior $p(\theta_j) \propto 1$ for each θ_j. For simplicity, we use the uniform prior in this study. Using the likelihood (eq. (2.2)) above, the following posterior distribution of $\boldsymbol{\theta}_o$ obtains

$$p(\boldsymbol{\theta}_o|\check{y}^o) = \frac{p(\check{y}^o|\boldsymbol{\theta}_o)p(\boldsymbol{\theta}_o)}{p(\check{y}^o)} \propto p(\check{y}^o|\boldsymbol{\theta}_o) \tag{2.3}$$

where the proportionality follows because $p(\check{y}^o)$ does not depend on $\boldsymbol{\theta}_o$ and the non-informativeness of the prior density $p(\boldsymbol{\theta}_o)$. Observe the dependence of $\boldsymbol{\theta}_o$ on y_o via $\check{y}^o$ in $\check{y}_i^o = \mathbb{1}\{y_i \leq y_o\}$. In fact, for fairly distinct values[10] of $y_{0,g} \in \bar{\mathcal{Y}}$, $g = 1, ..., G$, the posterior distributions $\{p(\boldsymbol{\theta}_{o,g}|\check{y}^{o,g})\}_{g=1}^G$ are distinct.

2.2 Conditional distribution and distributional effects

Given the above posterior distribution (eq. (2.3)) of $\boldsymbol{\theta}_o$, it is straightforward to use Markov Chain Monte Carlo (MCMC) methods[11] to obtain draws of $\boldsymbol{\theta}_o$. At a fixed threshold y_o, the unconditional distribution (with respect to $\check{y}^o$) of y_θ^o is

$$p(y_\theta^o) = \int p(y_\theta^o|\check{y}^o, \boldsymbol{\theta}_o)p(\boldsymbol{\theta}_o|\check{y}^o)p(d\check{y}^o)d\check{y}^o d\boldsymbol{\theta}_o \tag{2.4}$$

and

$$y_\theta^o = \int_{\boldsymbol{X}} F_Y(y_o|\mathbf{x})p(\mathbf{x})d\mathbf{x} = F_Y(y_o) \quad (y_\theta^o \in (0,1)) \tag{2.5}$$

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

[11]Two options, the Independence Metropolis-Hastings and the Random Walk Metropolis-Hastings algorithms, are available in our R package `bayesdistreg`. For a thorough treatment of methods for posterior simulation and computation, see Gelman, Carlin, Stern, and Rubin (1995, ch. 11 and 13)

where $\boldsymbol{X}$ is the support of all covariates $\mathbf{x}$. In practice, y^o_θ is computed as $N^{-1}\sum_{i=1}^{N}\Lambda(\mathbf{x}_i\boldsymbol{\theta}_o)$.[12]

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.

Remark 1. *One can obtain the counterfactual distribution of the outcome by replacing t_i with t^c_i in $\mathbf{x}$ and obtain $\hat{y}^c_\theta = N^{-1}\sum_{i=1}^{N}\Lambda(t^c_i, X_i, \boldsymbol{\theta}_o)$, where $\hat{y}^c_\theta$ is an estimate of y^c_θ is the counterfactual of y_θ. The counterfactual expression for (eq. (2.4)) obtains by simply replacing y^o_θ with y^c_θ.*

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

$$\Delta^{DE}_{y_o} = y^o_\theta - y^{o,c}_\theta = F_Y(y_o) - F_{Y^c}(y_o) \tag{2.6}$$

has the following distribution

$$p(\Delta^{DE}_{y_o}) = \int p(\Delta^{DE}_{y_o}|\breve{y}^o, \boldsymbol{\theta}_o)p(\boldsymbol{\theta}_o|\breve{y}^o)p(d\breve{y}^o)d\breve{y}^o d\boldsymbol{\theta}_o \tag{2.7}$$

The counterfactual quantile effect at the τ'th quantile of y is given by

$$\begin{gathered}\Delta^{QE}_{\tau} = F^{-1}_Y(\tau) - F^{-1}_{Y^c}(\tau)\\ \text{where } F^{-1}_Y(\tau) = \inf\{y \in \mathcal{Y} : F_Y(y_o) \geq \tau\}\end{gathered} \tag{2.8}$$

is the left inverse of $F_Y(y_o)$, $\tau \in (0, 1)$ (see Chernozhukov, Fernández-Val, and Melly (2013b, Appendix A)). $F^{-1}_{Y^c}(\tau)$ 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

[12] Since the distribution $F_Y(y_o|\boldsymbol{\theta}_o)$ in (eq. (2.5)) 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).

rather after inverting the entire distribution on $\mathcal{Y}$.

2.3 Estimation

We present an algorithm for the estimation of Bayesian distribution regression (BDR). 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`.[13]

Algorithm 1 (BDR algorithm)**.**

1. *Obtain a grid of threshold values* $y_{0,g},\ g = 1, ..., G$.

2. *For each* $g = 1, ..., G$

   (a) *Obtain the likelihood function* $p(\breve{y}^{o,g}|\boldsymbol{\theta}_{o,g})$ *using eq.* (2.2) *where* $\breve{y}_i^{o,g} = 1\{y_i \leq y_{0,g}\}$ *and the posterior distribution* $p(\boldsymbol{\theta}_{o,g}|\breve{y}^{o,g})$ *therefrom using eq.* (2.3)*.*

   (b) *For each* $m = 1, ..., M$*'th MCMC simulation*

      i. *Make a draw* $\boldsymbol{\theta}_{o,g}^m$ *from the posterior* $p(\boldsymbol{\theta}_{o,g}|\breve{y}^{o,g})$*.*[14]

      ii. *Compute* $\hat{y}_{\theta,m} = N^{-1}\sum_{i=1}^{N} \Lambda(\mathbf{x}_i, \boldsymbol{\theta}_{o,g}^m)$

   (c) *end* m

3. *end* g

Remark 2 (Computing counterfactual distributions and distribution effects)**.**
For counterfactual analysis, step (item 2(b)ii) can be modified to compute $y_{\theta,m}^c$ *and* $\Delta_{y_{o,g},m}^{DE} = y_{\theta,m} - y_{\theta,m}^c$ *at* $y_{o,g}$*. The* $G \times M$ *matrices* $\boldsymbol{P}_\theta^o$*,* $\boldsymbol{P}_\theta^{o,c}$ *and* $\boldsymbol{\Delta}_\theta^{DE}$ *that obtain from the BDR (algorithm 1) constitute draws of* y_θ^o*,* $y_\theta^{o,c}$ *and* $\Delta_{y_o}^{DE}$ *at thresholds* $\{y_{o,g}\}_{g=1}^G \subseteq \bar{\mathcal{Y}}$*.*

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

[14] 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) and Albert (2009, chapter 6).

3 Bayesian Inference

Bayesian analysis offers much flexibility in summarising posterior inference. Though summaries of location (mean, median, mode(s)) and variation (standard deviation, interquartile range, etc.) are desirable and practical, summaries on posterior uncertainty are quite important (Gelman, Carlin, Stern, and Rubin, 1995, section 2.3). In this paper, we focus on confidence bands that comprise $100(1-\alpha)\%$ central intervals of posterior probability (or posterior intervals) which are directly interpretable as the $\alpha/2$ and $1-\alpha/2$ quantiles of the posterior density. An alternative to posterior intervals is the $100(1-\alpha)\%$ region of highest posterior density (HPD) but this has the drawback of not being necessarily invariant to transformations (Berger, 2013, section 4.3.2, example 6). Confidence bands enable the testing of hypotheses by and beyond those intended by the researcher.

Definition 1 (Bayesian confidence bands of $\mathbf{F}$)**.**
Let $\mathbb{D}$ collect non-decreasing functions that map $\bar{\mathcal{Y}}$ into $[0,1]$[15] *and let $\mathbf{F} \in \mathbb{D}$ be a target distribution.*[16] *A confidence band $I = [L, U]$ collects intervals $I(y) = [L(y), U(y)]$, $L(y) \leq F(y) \leq U(y)\ \forall\ y \in \bar{\mathcal{Y}}$. If I covers $\mathbf{F}$ with probability at least $(1-\alpha)$, $I = [L, U]$ is a confidence band of $\mathbf{F}$ of level $(1-\alpha)$.*

The above definition follows definition 1 in Chernozhukov, Fernandez-Val, Melly, and Wüthrich (2016, section 2.1). Note that in our Bayesian framework, the confidence bands obtain by taking posterior intervals of $F_Y(y_o)$ (or of its counterfactual $F_{Y^c}(y_o)$) at each $y_{o,g}$, $g = 1, ..., G$. Note that the confidence band I may not be symmetric about $\mathbf{F}$. This occurs because Bayesian inference is exact and is based on the posterior distribution which may not be symmetric about $\mathbf{F}$ in small samples. Because the intervals of $\mathbf{F}$ using BDR are pointwise, they are generally conservative in joint coverage.

In the following theorem, we obtain adjustments to pointwise confidence bands from

[15]In the case of $\mathcal{Y}$, we consider it as a closed interval in the extended real line $\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$. See Chernozhukov, Fernandez-Val, Melly, and Wüthrich (2016, section 2.1)

[16]Target distributions are estimands of location like the mean, median or mode at $\{y_{o,g}\}_{g=1}^{G} \subseteq \bar{\mathcal{Y}}$.

Bayesian posterior intervals in algorithm 1 to obtain confidence bands with simultaneous coverage.

Theorem 1 (Simultaneous Bayesian confidence bands of $\mathbf{F}$)**.**
Let $\mathbf{F} \in \mathbb{D}$ be a target distribution function that obtains pointwise from BDR algorithm 1. $\mathbf{F}$ is simultaneously covered by $I^* = [L^*, U^*] = [\mathbf{F} - c^*_{1-\alpha}, \mathbf{F} + c^*_{1-\alpha}]$ with probability at least $(1-\alpha)$ where $c^*_{1-\alpha} = \max_y(q_{1-\alpha}(|F(y) - \hat{F}(y)|))$ and $q_{1-\alpha}(\cdot)$ denotes the $(1-\alpha)$'th quantile function.

Proof. See appendix B.2. □

The simultaneous confidence bands are easy to construct using Bayesian distribution regression (BDR) because the posterior density of $F(y)$ obtains at each $y \in \bar{\mathcal{Y}}$.

Remark 3 (Constructing Simultaneous Bayesian confidence bands of $\mathbf{F}$)**.**
Obtain $\mathbf{F} = [F(y_1), ..., F(y_G)]$ pointwise. For each $y_g \in \bar{\mathcal{Y}}$, compute the corresponding vector $|F(y_g) - y^{o,g}_\theta|$ where $y^{o,g}_\theta$ denotes all draws from the corresponding $p(y^{o,g}_\theta)$ in eq. (2.4)*.*

Having established simultaneous coverage of confidence bands of the BDR in theorem 1, we use results in Chernozhukov, Fernandez-Val, Melly, and Wüthrich (2016) and Chernozhukov, Fernandez-Val, and Weidner (2018) to extend inference to quantile functions and distributional effects.

Lemma 1 (Confidence bands of quantile functions $\mathbf{F}^{-1}$)**.**
For a distribution function $\mathbf{F} \in \mathbb{D}$ covered by distribution band functions $I = [L, U]$ with probability $(1-\alpha)$, the quantile function $\mathbf{F}^{-1}$ is covered by $I^{-1}(\tau) = [U^{-1}(\tau), L^{-1}(\tau)]$, $\tau \in [0, 1]$ with probability $(1-\alpha)$.

Proof. The proof follows from Chernozhukov, Fernandez-Val, Melly, and Wüthrich (2016, theorem 1). □

As noted in Chernozhukov, Fernandez-Val, Melly, and Wüthrich (2016), Scheffe, Tukey, et al. (1945) is the first to show that pointwise confidence intervals of quantiles obtained by inverting confidence intervals of distributions are valid but are conservative for discrete outcome variables. The result in the above lemma simplifies inference on quantiles obtained from the inversion of distributions. The following lemma extends results of lemma 1 to quantile effects.

Lemma 2 (Confidence bands of distribution and quantile effects).
For quantile functions and $\mathbf{F}^{-1}$ *covered by* $I^{-1} = [U^{-1}, L^{-1}]$ *with probability at least* $(1 - \alpha)$*, (with those of the counterfactual quantile function defined analogously), the quantile effect function* $\boldsymbol{\Delta}^{QE}$ *is covered by* $I_{QE} = [U^{-1}, L^{-1}] - [U^{-1,c}, L^{-1,c}]$ *with probability at least* $(1 - \alpha)$ *where the minus operator denotes a point-wise Minkowski difference.*

Proof. The proof follows from Chernozhukov, Fernandez-Val, Melly, and Wüthrich (2016, theorem 2) □

4 Asymptotic Inference

Large sample theory in Bayesian analysis is often not crucial for inference since Bayesian analysis provides distributions for direct inference. However, large sample results are useful and computationally convenient approximations. Some 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.[17]

Assumption 1.

(a) For each $y_o \in \bar{\mathcal{Y}}$*,* $\boldsymbol{\theta}_o$ *is defined on a corresponding compact set* $\boldsymbol{\Theta}_o \subset \mathbb{R}^k$*.*

(b) For each $y_o \in \bar{\mathcal{Y}}$*,* $\Lambda : \boldsymbol{\Theta}_o \to \mathbb{R}$ *is continuously differentiable on* $\boldsymbol{\Theta}_o \subset \mathbb{R}^k$*.*

(c) For each $y_o \in \bar{\mathcal{Y}}$*,* $\boldsymbol{\theta}_o$ *is in the interior of* $\boldsymbol{\Theta}_o$*.*

[17]For example, Agresti and Coull (1998) shows that tests based on the score function perform better than exact ones in terms of coverage probabilities of the confidence interval. A Bayesian inference based on the score retains the advantage in a frequentist sense.

Theorem 2 (Convergence of the posterior distribution)**.** - *Gelman, Carlin, Stern, and Rubin (1995, p. 587)*
Suppose assumption 1(a) holds in a neighbourhood $\mathcal{A}_o$ with non-zero prior probability, then $P(\boldsymbol{\theta}_o \in \mathcal{A}_o|\check{y}^o) \to 1$ as $N \to \infty$ where $\boldsymbol{\theta}_o$ minimises the Kullback-Leibler information $KL(\boldsymbol{\theta}_o) = \int \log\left(\frac{p(\check{y}^o)}{p(\check{y}^o|\boldsymbol{\theta}_o)}\right)p(\check{y}^o)d\check{y}^o$.

Proof. See Gelman, Carlin, Stern, and Rubin (1995, Appendix B) for a proof. □

Theorem 2 shows that the collection of posterior distributions $\{p(\boldsymbol{\theta}_{o,g}|\check{y}^{o,g})\}_{g=1}^{G}$ converge pointwise to the true posteriors. This result is important because it shows that the respective modes $\boldsymbol{\theta}_{o,g}$ are consistent estimates of the true values. In the next theorem, we show pointwise asymptotic normality of the posterior distributions.

Theorem 3 (Asymptotic normality of the posterior distribution)**.** - *Gelman, Carlin, Stern, and Rubin (1995, p. 587)*
Under standard regularity assumptions[18]*, the posterior distribution eq.* (2.3) *is asymptotically normal with mean vector and covariance matrix continuous in $y_o \in \bar{\mathcal{Y}}$. The asymptotic distribution is*

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

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 $\hat{\boldsymbol{\theta}}_o = \arg\max_{\boldsymbol{\theta}_o} p(\boldsymbol{\theta}_o|\mathbb{1}\{y \leq y_o\})$ for $y_o \in \mathcal{Y}$.

Proof. This proof is fairly standard. See appendix B.2 for proof. □

Remark 4 (Robust variance alternative to $\mathcal{I}(\hat{\boldsymbol{\theta}}_o)^{-1}$)**.**
Distribution regression is a semi-parametric estimation method which uses a series of binary response models to approximate the conditional distribution without making an assumption about the correct specification of the binary response model. In that case, one may want to

[18]See theorems 12.3 and 13.2 in Wooldridge (2010) for regularity conditions and asymptotic normality results for M- and Maximum likelihood estimators.

replace $[\mathcal{I}(\hat{\boldsymbol{\theta}}_o)]^{-1}$ *with a robust variance matrix*

$$\boldsymbol{V}_\theta = \Big(\sum_{i=1}^{N} \mathbf{H}_i(\boldsymbol{\theta}_o)\Big)^{-1}\Big(\sum_{i=1}^{N} \mathbf{s}_i(\boldsymbol{\theta}_o)\mathbf{s}_i(\boldsymbol{\theta}_o)'\Big)\Big(\sum_{i=1}^{N} \mathbf{H}_i(\boldsymbol{\theta}_o)\Big)^{-1}$$

where $\mathbf{H}_i(\boldsymbol{\theta}_o) = \frac{d^2}{d\boldsymbol{\theta}_o} L(\boldsymbol{\theta}_o|\check{y}_i^o) = -\Big(\frac{\exp(\mathbf{x}_i\boldsymbol{\theta}_o)}{(1+\exp(\mathbf{x}_i\boldsymbol{\theta}_o))^2}\Big)\mathbf{x}_i\mathbf{x}_i'$ *and* $\mathbf{s}_i(\boldsymbol{\theta}_o) = \nabla_{\boldsymbol{\theta}_o} L(\boldsymbol{\theta}_o|\check{y}_i^o) = \Big(\mathbb{1}\{y_i \leq y_o\} - \frac{\exp(\mathbf{x}_i\boldsymbol{\theta}_o)}{1+\exp(\mathbf{x}_i\boldsymbol{\theta}_o)}\Big)\mathbf{x}_i'$

For a finite set $\{y_{o,g}\}_{g=1}^{G} \subseteq \bar{\mathcal{Y}}$, the set of corresponding posterior distributions obtains as $\{p(\boldsymbol{\theta}_{o,g}|\check{y}^{o,g})\}_{g=1}^{G}$. The approximated posterior eq. (B.6) can be used in computing $p(y_\theta^o|\boldsymbol{\theta}_o)$ in eq. (2.4) and $p(\Delta_{y_o}^{DE}|\boldsymbol{\theta}_o)$ in eq. (2.7) and this process is computationally faster than MCMC because the approximated eq. (B.6) comes in a closed form.

In the following, theorem and 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 effects.

Theorem 4 (Asymptotic distribution of $y_\theta^o = F_Y(y_o|\boldsymbol{\theta}_o)$)**.**
Let assumption 1(b) & assumption 1(c) hold, using results in theorems 2 and 3, the asymptotic distribution of $\hat{F}_Y(y_o|\boldsymbol{\theta}_o)$ *is normal.*

$$\hat{F}_Y(y_o|\hat{\boldsymbol{\theta}}_o) \sim \mathcal{N}(F_{y_o}, \mathcal{V}_{F_{y_o}}) \tag{4.2}$$

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

Proof. See appendix B.2. □

Theorem 4 applies directly to the counterfactual distribution noting that $F_{y_o}^c = \int F_{Y|X}^c(y_o|\mathbf{x})dF(\mathbf{x}) = \int \mathit{\Lambda}(\mathbf{x}\boldsymbol{\alpha}'\boldsymbol{\theta}_o)dF(\mathbf{x})$ where $\boldsymbol{\alpha}$ is a $k \times k$ diagonal matrix that multiplicatively creates a counterfactual of $\mathbf{x}$. This applicability is assumed in the next corollaries. In the following corollary, results in theorem 4 are extended to the joint distribution at thresholds $\{y_{o,g}\}_{g=1}^{G} \subseteq \bar{\mathcal{Y}}$.

Corollary 1 (Joint asymptotic distribution of $\{\hat{F}_Y(y_{o,g}|\hat{\boldsymbol{\theta}}_{o,g})\}_{g=1}^G$)**.**
Extending results from theorem 4 to the joint distribution at several indices $\{y_{o,g}\}_{g=1}^G \subseteq \bar{\mathcal{Y}}$, *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 distributed with*

$$\hat{\mathbf{F}}_Y \sim \mathcal{N}(\mathbf{F}_Y, \boldsymbol{\Omega}_{F_{y_o}}) \tag{4.3}$$

where $\mathbf{F}_Y = [F_Y(y_{o,1}|\boldsymbol{\theta}_{o,1}), ..., F_Y(y_{o,G}|\boldsymbol{\theta}_{o,G})]'$,
$\boldsymbol{\Omega}_{F_{y_o}}$*'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* $\mathcal{V}_{F_{y_{g,h}}} = E[\boldsymbol{\lambda}_i^{\mathbf{x}}(\boldsymbol{\theta}_{o,g})'[I(\boldsymbol{\theta}_{o,g})]^{-1}I_i(\boldsymbol{\theta}_{g,h})[I(\boldsymbol{\theta}_{o,h})]^{-1}\boldsymbol{\lambda}_i^{\mathbf{x}}(\boldsymbol{\theta}_{o,h})]$, $I_i(\boldsymbol{\theta}_{g,h}) = N^{-1}\mathbf{s}_i(\boldsymbol{\theta}_g)\mathbf{s}_i(\boldsymbol{\theta}_h)'$.

Proof. See appendix B.2. □

The asymptotic distribution of the distribution effect $\hat{\Delta}_{y_o}^{DE} = \hat{F}_Y(y_o) - \hat{F}_Y^c(y_o)$ at a threshold $y_o \in \bar{\mathcal{Y}}$ follows from theorem 4 because the theorem also applies to the counterfactual distribution. In the following corollaries, we show the asymptotic distribution of the distribution effect.

Corollary 2 (Asymptotic distribution of $\hat{\Delta}_{y_o}^{DE}$)**.**
Let assumption 1(b) & assumption 1(c) hold, using results in theorems 2 and 3, the distribution effect at a threshold $y_o \in \mathcal{Y}$, $\hat{\Delta}_{y_o}^{DE} = \hat{F}_Y(y_o) - \hat{F}_Y^c(y_o)$ *is normally distributed,* $\hat{\Delta}_{y_o}^{DE} \sim \mathcal{N}(\Delta_{y_o}^{DE}, \mathcal{V}_{\Delta_o^{DE}})$ *where* $\Delta_{y_o}^{DE} = \Delta(\boldsymbol{\theta}_o) = \int(\Lambda(\mathbf{x}\boldsymbol{\theta}_o) - \Lambda(\mathbf{x}\boldsymbol{\alpha}'\boldsymbol{\theta}_o))dF(\mathbf{x})$, $\mathcal{V}_{\Delta_o^{DE}} = E[\bar{\boldsymbol{\lambda}}_i^{\mathbf{x}}(\boldsymbol{\theta}_o)[I(\boldsymbol{\theta}_o)]^{-1}\bar{\boldsymbol{\lambda}}_i^{\mathbf{x}}(\boldsymbol{\theta}_o)']$, $\bar{\boldsymbol{\lambda}}_i^{\mathbf{x}}(\boldsymbol{\theta}_o) = \Lambda'(\mathbf{x}_i\boldsymbol{\theta}_o) - \Lambda'(\mathbf{x}_i\boldsymbol{\alpha}'\boldsymbol{\theta}_o)\boldsymbol{\alpha}$ *and* $\boldsymbol{\alpha}$ *is a* $k \times k$ *diagonal matrix that multiplicatively creates a counterfactual of* $\mathbf{x}$.

Proof. See appendix B.2. □

The asymptotic result in corollary 2 is extensible to several indices $\{y_{o,g}\}_{g=1}^G \subseteq \bar{\mathcal{Y}}$ using similar arguments in corollary 1 to establish joint normality.

Corollary 3 (Joint asymptotic distribution of $\hat{\boldsymbol{\Delta}}^{DE}$)**.**
Extending results from corollary 2 to the joint distribution at several indices $\{y_{o,g}\}_{g=1}^G \subseteq \bar{\mathcal{Y}}$,

the joint asymptotic distribution of $\hat{\boldsymbol{\Delta}}^{DE} = [\hat{\Delta}^{DE}_{y_{o,1}}, ..., \hat{\Delta}^{DE}_{y_{o,G}}]'$ *is joint normally distributed with*

$$\hat{\boldsymbol{\Delta}}^{DE} \xrightarrow{d} \mathcal{N}(\boldsymbol{\Delta}^{DE}, \boldsymbol{\Omega}_{\Delta}) \tag{4.4}$$

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

Proof of corollary 3 .

See appendix B.2. □

There are three possibilities of carrying out Bayesian distribution regression, the first involving MCMC draws (see algorithm 1), a second that uses asymptotically approximated posterior distributions (which replace MCMC) from Theorem 3 with every other step carried with parameter draws and a third that uses a joint asymptotic distribution. The first one is fairly exact even in small samples whereas the second and third are mostly recommended when the sample size is fairly large because they are based on asymptotic approximations. While the BDR estimator can be deemed purely Bayesian, Bayesian inference on the approximated posterior distributions draws on the Bayesian asymptotic theory and the third is purely asymptotic. Inference results in section 3 on quantile and quantile effect functions hold for all three estimators. In addition, a recent work Montiel Olea and Plagborg-Møller (2018) presents a framework (see algorithm 1 in the paper) usable in the case of the purely asymptotic DR estimator for obtaining simultaneous confidence bands.

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.[19] Inspired by the evidence of Novy-Marx (2013) and Titman, Wei, and Xie (2004), Fama and French (2015) establish 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 eq. (5.1), R_{it} is the excess return on portfolio i for period t, which is equal to the return minus the riskfree return.[20] 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.

[19]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.

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

5.2 Data

We use Fama and French (2015) US dataset, which can be downloaded from Kenneth R. French's website. We use a 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.[21] We choose value-weighted portfolio returns. Thus, we have 18 portfolios in all. To save space, we show the 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 the 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 portfolio 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

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

5.3 Results

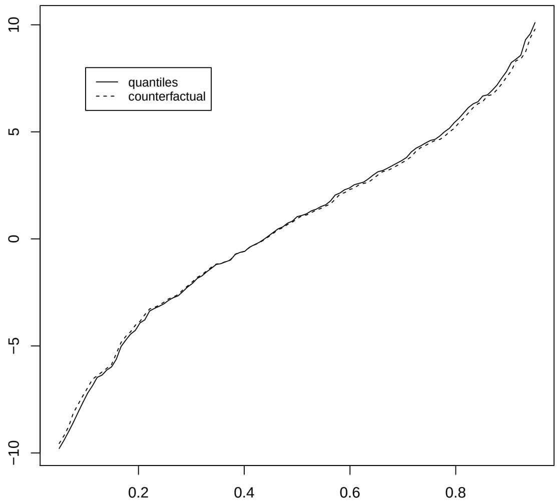
Table 2 shows the results by running the Fama and French's five-factor model. We can see the coefficients on all factors, which are 1.021 for the market factor, 0.972 for SMB, -0.058 for HML, -0.503 for RMW, and 0.033 for CMA. Notice that the coefficients on the market, size, value and profitability 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.[22]

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.[23] 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 of Figure 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.

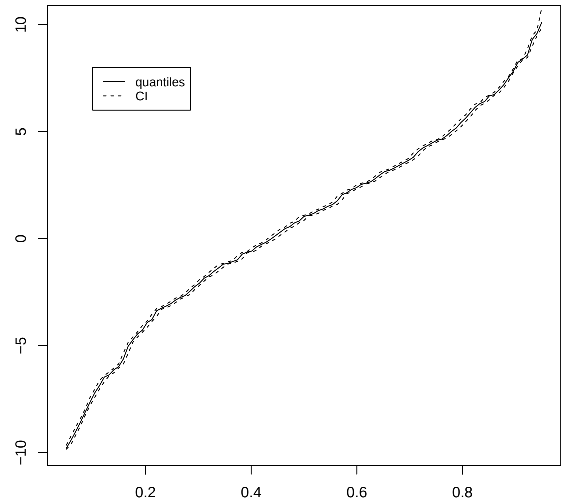
Figure 2 shows the effects (in percentages) of the counterfactual changes in the stock market (from a decline of 20% to an increase of 20%) on the quantiles of the portfolio return. Figure 2a shows that the 5th quantile of portfolio returns increases as the market return

[22]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 appendix A.

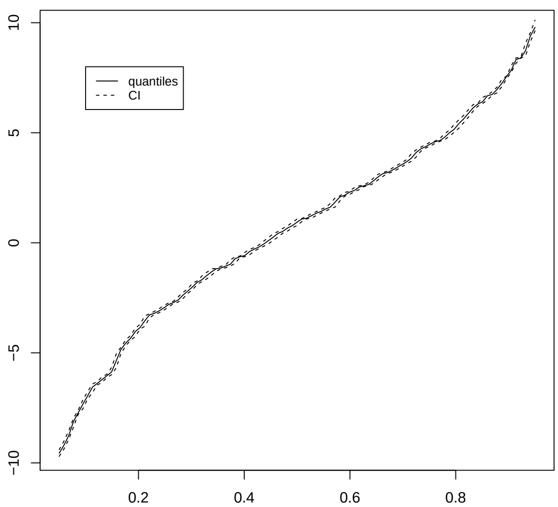
[23]We do not plot the distribution of a distribution which is a three-dimensional figure. Instead, we plot 95% confidence interval, which is simply the 5th and 95th quantile of the distribution of the distribution.



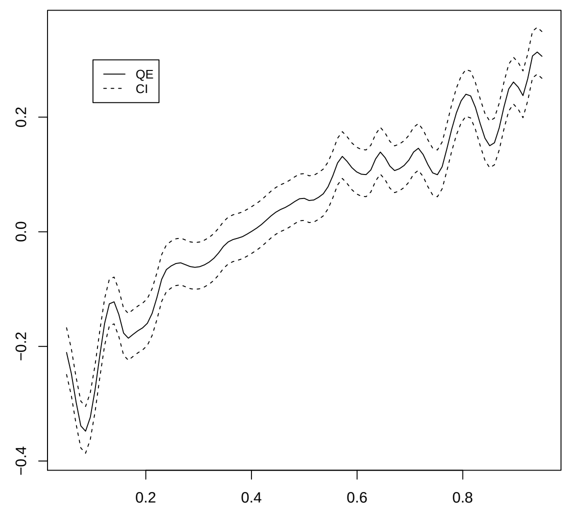
(a) Quantiles and counterfactual



(b) Quantiles



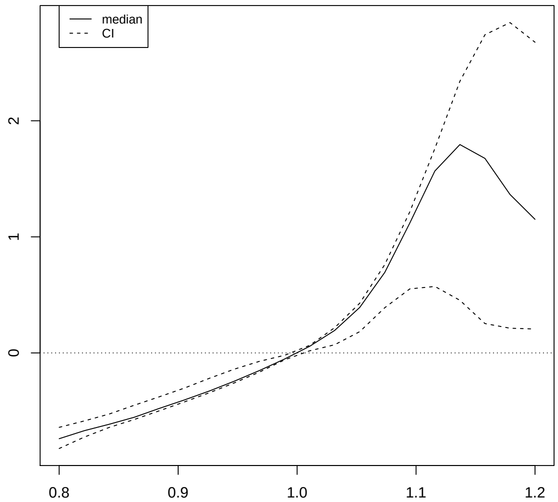
(c) Couterfactual quantiles



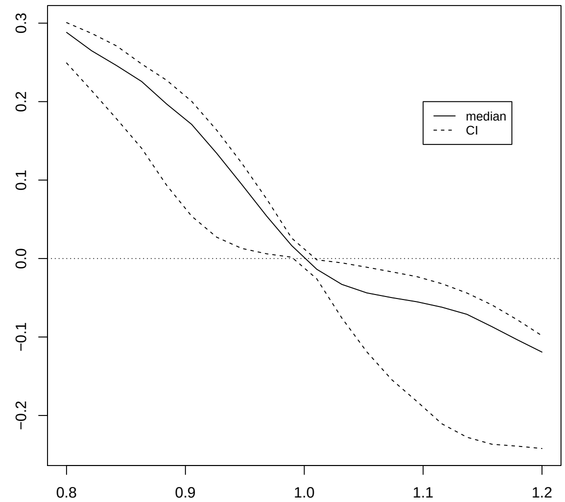
(d) Quantile effects

Figure 1: (Counterfactual) quantiles and quantile effects

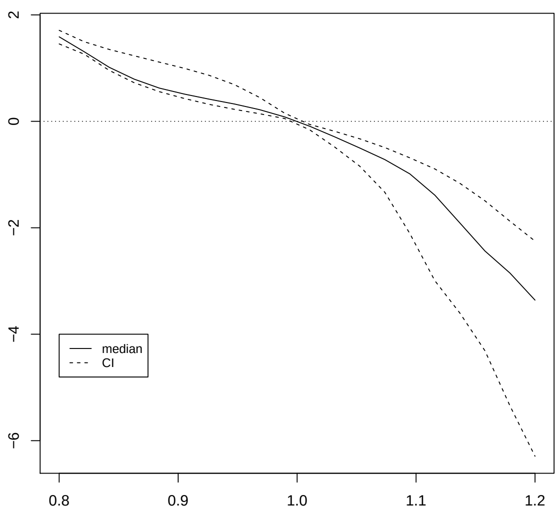
Notes: The top left panel plots the (counterfactual) quantiles of the portfolio return, in which the counterfactual quantiles are the quantiles that would prevail when the stock market declines by 5%. The top left panel plots the quantiles of the portfolio return and its 95% confidence interval. The 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



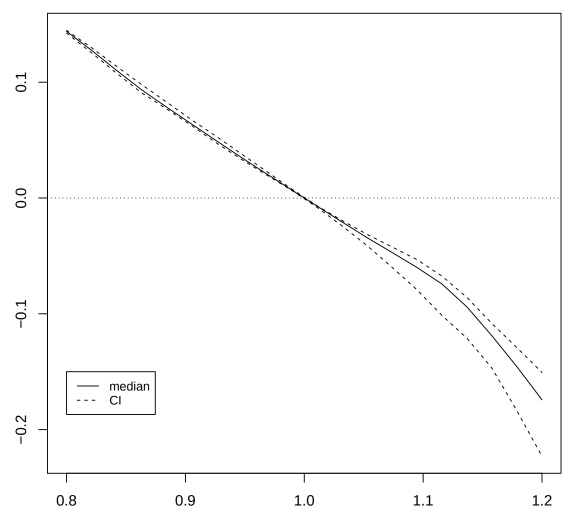
(a) 5%



(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 the market on the 5th quantile of portfolio returns and the 95% confidence bands. The top right panel plots the quantile effect on the median of portfolio returns and the 95% confidence bands. The bottom left panel plots the quantile effect on the 95th quantile of portfolio returns and the 95% confidence bands and the bottom right plots the effects on the mean.
Sources: Kenneth R. French's website

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.[24] However, Figure 2b and Figure 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.

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.

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

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] Albert, Jim. *Bayesian computation with R*. Springer Science & Business Media, 2009.

[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] Berger, James O. *Statistical decision theory and Bayesian analysis*. Springer Science & Business Media, 2013.

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

[8] 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.

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

[10] 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.

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

[12] Chernozhukov, Victor, Iván Fernández-Val, Sukjin Han, and Amanda Kowalski. *Censored quantile instrumental variable estimation with Stata*. Tech. rep. National Bureau of Economic Research, 2018.

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

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

[15] Chernozhukov, Victor, Ivan Fernandez-Val, Blaise Melly, and Kaspar Wüthrich. “Generic inference on quantile and quantile effect functions for discrete outcomes”. *arXiv preprint arXiv:1608.05142* (2016).

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

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

[18] 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.

[19] 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.

[20] Doorley, Karina and Eva Sierminska. "Myth or fact? The beauty premium across the wage distribution" (2012).

[21] Dube, Arindrajit. "Minimum wages and the distribution of family incomes" (2017).

[22] 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.

[23] 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.

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

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

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

[27] 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.

[28] 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.

[29] 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.

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

[31] Han, Lu, Chandler Lutz, and B Sand. *The effects of macroprudential mortgage insurance regulation during a housing boom: Evidence from Canada*. Tech. rep. Citeseer, 2016.

[32] 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.

[33] 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.

[34] 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.

[35] Hosseini, Reza. "Quantiles equivariance". *arXiv preprint arXiv:1004.0533* (2010).

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

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

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

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

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

[41] 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.

[42] Montiel Olea, José Luis and Mikkel Plagborg-Møller. "Simultaneous Confidence Bands: Theory, Implementation, and an Application to SVARs" (2018).

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

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

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

[46] 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.

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

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

[49] Scheffe, Henry, John W Tukey, et al. "Non-parametric estimation. I. Validation of order statistics". *The Annals of Mathematical Statistics* 16.2 (1945), pp. 187–192.

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

[51] 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.

[52] Shepherd, Debra et al. "A question of efficiency: decomposing South African reading test scores using PIRLS 2006". *Stellenbosch University, Department of Economics Working Papers (20* (2013).

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

[54] 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.

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

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

[57] Wüthrich, Kaspar. *Semiparametric estimation of quantile treatment effects with endogeneity*. Tech. rep. Discussion Papers, Universität Bern, Department of Economics, 2015.

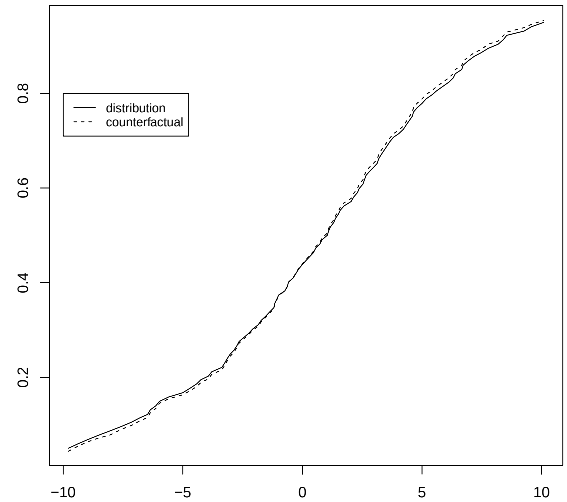
[58] Yang, Yunwen and Huixia Judy Wang. "Bayesian Quantile Regression". *Handbook of Quantile Regression*. Chapman and Hall/CRC, 2017, pp. 61–74.

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

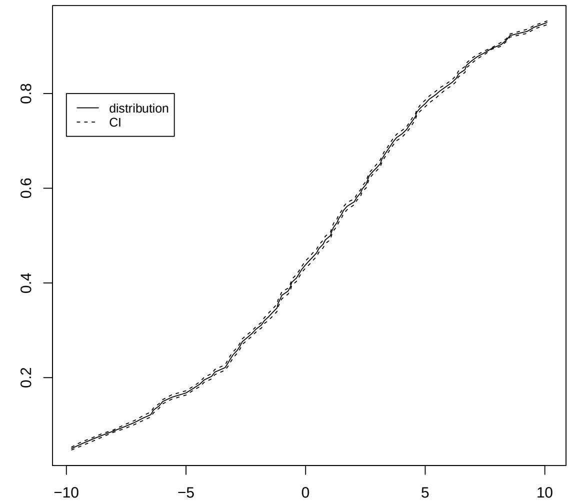
A Distribution effects

Figure 3 shows the (counterfactual) distribution of the portfolio return and the distribution effects of a market decline by 5%. Figure 3a plots the distribution and its counterfactual. Figure 3b plots the distribution and its 95% confidence band. Figure 3c plots the counterfactual distribution and its 95% pointwise confidence band. Figure 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%.

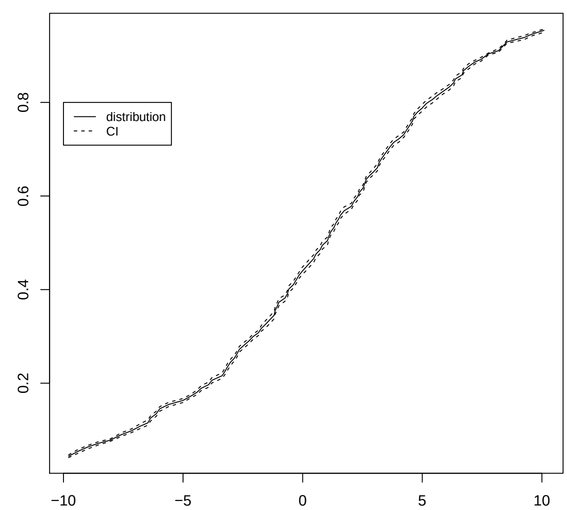
Figure 4 shows the counterfactual distribution effects (in percentages) of changes in the market returns (from a decline of 20% to an increase of 20%). Figure 4a shows that the probability of the portfolio returns being less than the 5th percentile of the portfolio return, which is -9.79% per month, is 1.2% when the market return declines by 10%. However, Figure 4c shows that the probability of the portfolio returns being less than the 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, Figure 4b shows that the probability of the portfolio returns being less than the 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.



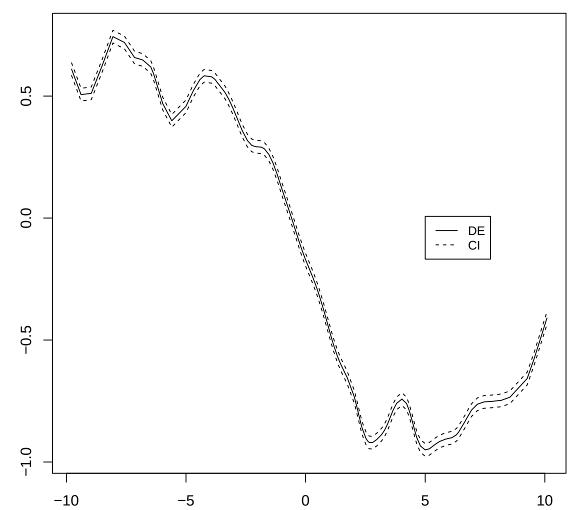
(a) Distribution and counterfactual



(b) Distribution



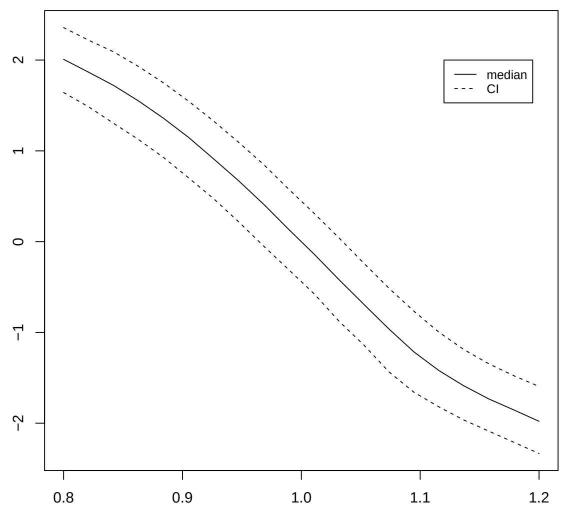
(c) Couterfactual distribution



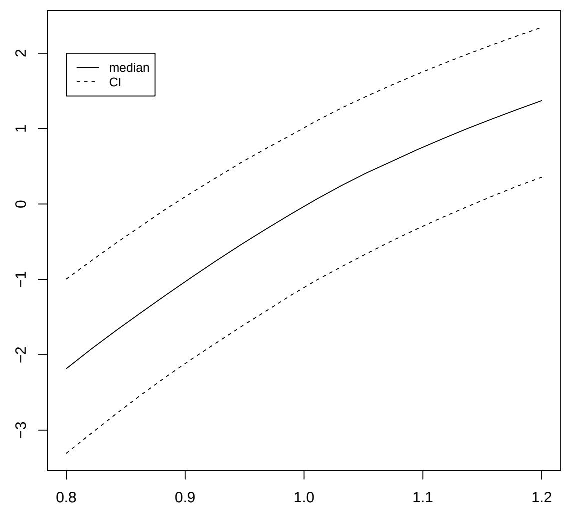
(d) Distribution effects

Figure 3: (Counterfactual) distribution and distribution effects

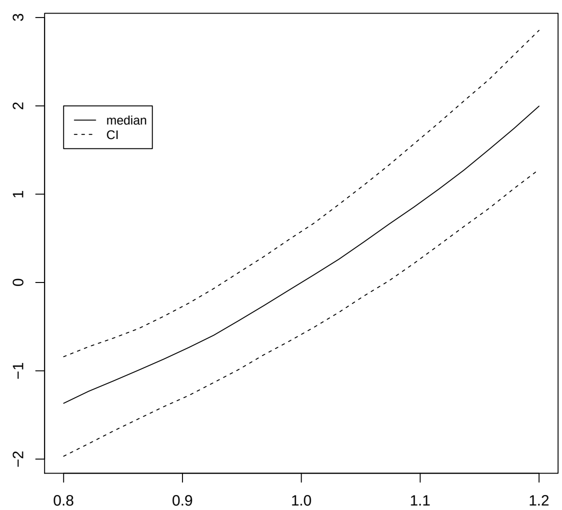
Notes: The top left panel plots the (counterfactual) distribution of the portfolio return, in which the counterfactual distribution is the distribution that would prevail when the stock market declines by 5%. The top right panel plots the distribution of the portfolio return and its 95% confidence bands.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



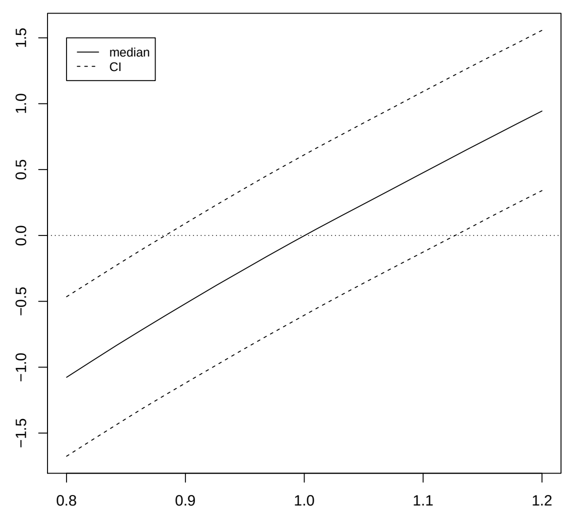
(a) 5%



(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 5th quantile of portfolio returns and the 95% confidence bands. The top right panel plots the distribution effect on the median of portfolio returns and the 95% confidence bands. The bottom left panel plots the distribution effect on the 95th quantile of portfolio returns and the bottom right plots the distribution effect on the mean.
Sources: Kenneth R. French's website

B Proofs

B.1 Lemmata

B.2 Theorems

Proof of Theorem 1.

We need to show that $P(L^*(y) \leq F(y) \leq U^*(y),\ \forall y \in \bar{\mathcal{Y}}) \geq 1-\alpha$.

$$
\begin{aligned}
&P(\hat{F}(y) - c^*_{1-\alpha} \leq F(y) \leq \hat{F}(y) + c^*_{1-\alpha}, \forall y \in \bar{\mathcal{Y}}) \\
=&P(-c^*_{1-\alpha} \leq F(y) - \hat{F}(y) \leq c^*_{1-\alpha}, \forall y \in \bar{\mathcal{Y}}) \\
=&P(-\max_{y\in\bar{\mathcal{Y}}}(q_{1-\alpha}(|F(y)-\hat{F}(y)|)) \leq F(y) - \hat{F}(y) \leq \max_{y\in\bar{\mathcal{Y}}}(q_{1-\alpha}(|F(y)-\hat{F}(y)|))) \\
=&P(|F(y)-\hat{F}(y)| \leq \max_{y\in\bar{\mathcal{Y}}}(q_{1-\alpha}(|F(y)-\hat{F}(y)|))) \\
=&P(|F(y)-\hat{F}(y)| \leq q_{1-\alpha}(\max_{y\in\bar{\mathcal{Y}}}(|F(y)-\hat{F}(y)|))) \\
\geq& 1-\alpha \quad \forall\, y \in \bar{\mathcal{Y}}
\end{aligned}
\tag{B.1}
$$

The first line uses the definition in theorem 1 and the second line rearranges the terms. The second equality holds for all $y \in \bar{\mathcal{Y}}$ because of the $\max(\cdot)$ operator and the third equality is a simplification of the second. The fourth equality uses the fact that the quantile passes through monotonic functions.[25][26] The last line follows from the definition of the quantile function $q_{1-\alpha}(\cdot)$. Notice that the argument to the quantile function is the Kolmogorov maximal statistic. □

Proof of Theorem 3.

This proof follows Gelman, Carlin, Stern, and Rubin (1995, Appendix B). The log of the

[25] See Wooldridge (2010, p. 453) for a similar illustration with the median.

[26] Note that the exchangeability of the $\max(\cdot)$ and quantile $q_{1-\alpha}(\cdot)$ operators needs continuity in y though $\max(\cdot)$ is continuous and monotonic. See Hosseini (2010) for an interesting discussion on the equivariance of the quantile operator. Our simulations (not presented in the paper) show that the equality fails to hold when y is discrete, for example on the set of natural numbers $y = [1, ..., 100]$.

posterior obtains as

$$L(\boldsymbol{\theta}_o|\breve{y}^o) = \log p(\boldsymbol{\theta}_o|\breve{y}^o) = \sum_{i=1}^{N} \mathbf{x}_i \boldsymbol{\theta}_o \mathbb{1}\{y_i \le y_o\} - \sum_{i=1}^{N} \log(1 + \exp(\mathbf{x}_i \boldsymbol{\theta}_o)) \tag{B.2}$$

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

$$\mathbf{s}(\boldsymbol{\theta}_o) = \nabla_{\boldsymbol{\theta}_o} L(\boldsymbol{\theta}_o|\breve{y}^o) = \sum_{i=1}^{N} \Big(\mathbb{1}\{y_i \le y_o\} - \frac{\exp(\mathbf{x}_i \boldsymbol{\theta}_o)}{1 + \exp(\mathbf{x}_i \boldsymbol{\theta}_o)}\Big) \mathbf{x}_i' \tag{B.3}$$

Taking the second derivative with respect to $\boldsymbol{\theta}$ and evaluating at $\boldsymbol{\theta}_o$,

$$\frac{d^2}{d\boldsymbol{\theta}_o} L(\boldsymbol{\theta}_o|\breve{y}^o) = -\sum_{i=1}^{N} \Big(\frac{\exp(\mathbf{x}_i \boldsymbol{\theta}_o)}{(1 + \exp(\mathbf{x}_i \boldsymbol{\theta}_o))^2}\Big) \mathbf{x}_i \mathbf{x}_i' \tag{B.4}$$

obtains as the hessian matrix. Notice that the expression in eq. (B.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

$$L(\boldsymbol{\theta}_o|\breve{y}^o) = L(\hat{\boldsymbol{\theta}}_o|\breve{y}^o) + \frac{1}{2}(\boldsymbol{\theta}_o - \hat{\boldsymbol{\theta}}_o)' \Big[\frac{d^2}{d\boldsymbol{\theta}_o} L(\boldsymbol{\theta}_o|\breve{y}^o)\Big]\Big|_{\boldsymbol{\theta}_o = \hat{\boldsymbol{\theta}}_o} (\boldsymbol{\theta}_o - \hat{\boldsymbol{\theta}}_o) + (s.o.) \tag{B.5}$$

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

$$p(\boldsymbol{\theta}_o|\breve{y}^o) \approx \mathcal{N}(\hat{\boldsymbol{\theta}}_o, [\mathcal{I}(\hat{\boldsymbol{\theta}}_o)]^{-1}) \tag{B.6}$$

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 **??**, $\boldsymbol{\theta}_o$ is continuous in y_o and by extension, the (asymptotic) posterior distribution eq. (B.6) is a continuous function in y_o. □

Proof of Theorem 4.

The above results follow from the delta method[27] and noting the exchangeability of the

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

derivative and the integral which holds under general regular conditions.

$$\begin{aligned}\sqrt{N}(\hat{F}_Y(y_o|\hat{\boldsymbol{\theta}}_o) - F_Y(y_o|\boldsymbol{\theta}_o)) &= \sqrt{N}(\hat{F}_Y(y_o|\hat{\boldsymbol{\theta}}_o) - \hat{F}_Y(y_o|\boldsymbol{\theta}_o) + \hat{F}_Y(y_o|\boldsymbol{\theta}_o) - F_Y(y_o|\boldsymbol{\theta}_o)) \\ = \sqrt{N}(N^{-1}\sum_{i=1}^{N}\Lambda(\mathbf{x}_i\hat{\boldsymbol{\theta}}_o) - N^{-1}\sum_{i=1}^{N}\Lambda(\mathbf{x}_i,\boldsymbol{\theta}_o)) &+ \sqrt{N}(N^{-1}\sum_{i=1}^{N}\Lambda(\mathbf{x}_i,\boldsymbol{\theta}_o) - F_Y(y_o|\boldsymbol{\theta}_o))\end{aligned} \tag{B.7}$$

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

$$\sqrt{N}(\hat{F}_Y(y_o|\hat{\boldsymbol{\theta}}_o) - F_Y(y_o|\boldsymbol{\theta}_o)) = N^{-1}\sum_{i=1}^{N}\Lambda'(\mathbf{x}_i\boldsymbol{\theta}_o)\mathbf{x}_i'\sqrt{N}(\hat{\boldsymbol{\theta}}_o - \boldsymbol{\theta}_o) + o_p(1) \tag{B.8}$$

Applying the central limit theorem,

$$\sqrt{N}(\hat{F}_Y(y_o|\hat{\boldsymbol{\theta}}_o) - F_Y(y_o|\boldsymbol{\theta}_o)) \xrightarrow{d} \mathcal{N}(\mathbf{0}, N\mathcal{V}_{F_{y_o}}) \tag{B.9}$$

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

Proof of corollary 1 .

From the score function of the posterior by $\mathbf{s}(\boldsymbol{\theta}_o)$ eq. (B.3), the influence function representation of $\hat{\boldsymbol{\theta}}_o$ (see Wooldridge (2010, equations 12.15 - 12.17)) obtains as

$$\sqrt{N}(\hat{\boldsymbol{\theta}}_o - \boldsymbol{\theta}_o) = [I(\boldsymbol{\theta}_o)]^{-1}N^{-1/2}\sum_{i=1}^{N}\mathbf{s}_i(\boldsymbol{\theta}_o) + o_p(1) \tag{B.10}$$

Expanding eq. (B.8) using eq. (B.10) obtains

$$\sqrt{N}(\hat{F}_Y(y_{o,g}|\hat{\boldsymbol{\theta}}_{o,g}) - F_Y(y_{o,g}|\boldsymbol{\theta}_{o,g})) = N^{-1}\sum_{i=1}^{N}\Lambda'(\mathbf{x}_i\boldsymbol{\theta}_{o,g})\mathbf{x}_i'[I(\boldsymbol{\theta}_{o,g})]^{-1}N^{-1/2}\mathbf{s}_i(\boldsymbol{\theta}_{o,g}) + o_p(1) \tag{B.11}$$

Applying the multivariate central limit theorem (see Van der Vaart (1998, Section 2.18)) to $\sqrt{N}[\mathbf{F}_Y - \mathbf{F}_Y]' = \sqrt{N}[(\hat{F}_Y(y_{o,1}|\hat{\boldsymbol{\theta}}_{o,1}) - F_Y(y_{o,1}|\boldsymbol{\theta}_{o,1})), ..., (\hat{F}_Y(y_{o,G}|\hat{\boldsymbol{\theta}}_{o,G}) - F_Y(y_{o,G}|\boldsymbol{\theta}_{o,G}))]'$ using the representation in eq. (B.11),

$$\sqrt{N}[\hat{\mathbf{F}}_Y - \mathbf{F}_Y]' \xrightarrow{d} \mathcal{N}(\mathbf{0}, N\boldsymbol{\Omega}_{F_y}) \tag{B.12}$$

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

Proof of corollary 2.

$$\begin{aligned}
\sqrt{N}(\hat{\Delta}_{y_o}^{DE} - \Delta_{y_o}^{DE}) &= \sqrt{N}((\hat{F}_Y(y_o) - \hat{F}_Y^c(y_o)) - (F_Y(y_o) - F_Y^c(y_o))) \\
&= \sqrt{N}(\hat{F}_Y(y_o) - F_Y(y_o)) - \sqrt{N}(\hat{F}_Y^c(y_o) - F_Y^c(y_o)) \\
&= N^{-1}\sum_{i=1}^{N}(\Lambda'(\mathbf{x}_i\boldsymbol{\theta}_o) - \Lambda'(\mathbf{x}_i\boldsymbol{\alpha}'\boldsymbol{\theta}_o)\boldsymbol{\alpha})\mathbf{x}_i'[I(\boldsymbol{\theta}_o)]^{-1}N^{-1/2}\mathbf{s}_i(\boldsymbol{\theta}_o) + o_p(1) \\
&= N^{-1}\sum_{i=1}^{N}\bar{\boldsymbol{\lambda}}_i^{\mathbf{x}}(\boldsymbol{\theta}_o)[I(\boldsymbol{\theta}_o)]^{-1}N^{-1/2}\mathbf{s}_i(\boldsymbol{\theta}_o) + o_p(1)
\end{aligned} \tag{B.13}$$

where $\bar{\boldsymbol{\lambda}}_i^{\mathbf{x}}(\boldsymbol{\theta}_o) = \Lambda'(\mathbf{x}_i\boldsymbol{\theta}_o) - \Lambda'(\mathbf{x}_i\boldsymbol{\alpha}'\boldsymbol{\theta}_o)\boldsymbol{\alpha}$. From the above influence function representation, it follows from CLT that

$$\sqrt{N}(\hat{\Delta}_{y_o}^{DE} - \Delta_{y_o}^{DE}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, N\mathcal{V}_{\Delta_o^{DE}}) \tag{B.14}$$

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

Proof of corollary 3. This result obtains by applying the multivariate central limit theorem (see Van der Vaart (1998, Section 2.18)) to $\sqrt{N}(\hat{\boldsymbol{\Delta}}^{DE} - \boldsymbol{\Delta}^{DE}) = \sqrt{N}[(\hat{\Delta}_{y_{o,1}}^{DE} - \Delta_{y_{o,1}}^{DE}), ..., (\hat{\Delta}_{y_{o,G}}^{DE} - \Delta_{y_{o,G}}^{DE})]'$. Using the representation in eq. (B.13),

$$\sqrt{N}(\hat{\boldsymbol{\Delta}}^{DE} - \boldsymbol{\Delta}^{DE}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, N\boldsymbol{\Omega}_{\Delta}) \tag{B.15}$$

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