

Article

# Estimating the Capital Asset Pricing Model with Many Instruments: A Bayesian Shrinkage Approach <sup>†</sup>

Cássio Roberto de Andrade Alves <sup>‡</sup> and Márcio Laurini <sup>\*‡</sup> 

Department of Economics, School of Economics, Business Administration and Accounting at Ribeirão Preto (FEA-RP/USP), University of São Paulo, Av. dos Bandeirantes 3900, Ribeirão Preto 14040-905, SP, Brazil; cassioalves@usp.br

\* Correspondence: laurini@fearp.usp.br; Tel.: +55-16-3329-0867

<sup>†</sup> This article is an extended version of our work presented in the XII Brazilian Finance Meeting, 2022, Vitória-ES.

<sup>‡</sup> These authors contributed equally to this work.

**Abstract:** This paper introduces an instrumental variable Bayesian shrinkage approach specifically designed for estimating the capital asset pricing model (CAPM) while utilizing a large number of instruments. Our methodology incorporates horseshoe, Laplace, and factor-based shrinkage priors to construct Bayesian estimators for CAPM, accounting for the presence of measurement errors. Through the use of simulated data, we illustrate the potential of our approach in mitigating the bias arising from errors-in-variables. Importantly, the conventional two-stage least squares estimation of the CAPM beta is shown to experience bias escalation as the number of instruments increases. In contrast, our approach effectively counters this bias, particularly in scenarios with a substantial number of instruments. In an empirical application using real-world data, our proposed methodology generates subtly distinct estimated CAPM beta values compared with both the ordinary least squares and the two-stage least squares approaches. This disparity in estimations carries notable economic implications. Furthermore, when applied to average cross-sectional asset returns, our approach significantly enhances the explanatory power of the CAPM framework.

**Keywords:** Bayesian estimation; shrinkage; instrumental variables; CAPM

**MSC:** 62F15



**Citation:** de Andrade Alves, C.R.; Laurini, M. Estimating the Capital Asset Pricing Model with Many Instruments: A Bayesian Shrinkage Approach. *Mathematics* **2023**, *11*, 3776. <https://doi.org/10.3390/math11173776>

Academic Editor: Hsien-Chung Wu

Received: 7 August 2023

Revised: 23 August 2023

Accepted: 31 August 2023

Published: 2 September 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Many asset pricing models incorporate the return of the market portfolio as an independent variable in the estimation process. The capital asset pricing model (CAPM) is perhaps the most renowned example of such models, owing to its theoretical simplicity and ease of interpretation. Estimating the CAPM necessitates a proxy for the return of the market portfolio due to the unavailability of the actual market portfolio's data [1,2]. The introduction of a substitute for the market portfolio return introduces an error-in-variables (EIV) issue, which biases the estimates and complicates results interpretation. This concern is referred to as Roll's critique [3]. The EIV problem poses challenges in estimating the CAPM and assessing its empirical validity, and notably impacts investment decisions [4,5] as well as testing procedures for portfolio efficiency [6–9].

The usual econometric solution to the error-in-variables problem is to use instrumental variables (IVs). However, it can be challenging to find 'strong' instruments for the market return [10,11]. The data-rich environment of financial data sets offers many candidates for instrumental variables, although they are usually only weakly correlated with the returns of the market portfolio. Alternatively, in contrast to low-dimensional model settings, in which we select a small set of instruments—imposing an ad hoc sparsity, all these many candidates for instrumental variables may be incorporated into the model, leading to

high-dimensional model settings. Unfortunately, conventional econometric techniques cannot deal with high-dimensional asset pricing model settings [12].

This paper proposes an instrumental variable Bayesian shrinkage approach to estimate the capital asset pricing model using a large set of instruments. Bayesian shrinkage techniques can deal with high-dimensional models by using regularization priors. This approach has been increasingly adopted in financial econometrics [13–16] for example. For a comprehensive view of recent advancements in this literature, see [17–21]. Regularization priors are particularly helpful when there are several potential instruments, as in the case of CAPM estimation. Without these regularization priors, using many instruments generates biased estimates [22–24]. In terms of the two-stage least squares (2SLS), for example, a large set of instruments implies an overfitting in the first stage, because of the tendency for the ordinary least squares (OLS) method to fit too well. Then, the second stage will be closer to a simple OLS method, which is biased in the presence of EIVs. The regularization priors avoid this overfitting in the first stage and, consequently, also avoid the bias in the many instruments setting. Thus, the high-dimensional model setting combined with prior regularization may offer a new approach to deal with Roll's critique.

Although high-dimensional models have become increasingly popular in financial econometrics literature, the use of high-dimensional models for instrumental variables combined with regularization techniques is still a little-explored field. In this paper, we estimate the capital asset pricing model using a large set of instruments and shrinkage priors over the parameters associated with the instruments. We use the Bayesian approach proposed by [25] to shrink unimportant instruments and compare the size of the estimated bias with that produced by the traditional estimation methods (ordinary least squares and two-stage least squares).

We compare our approach both in simulated data (Monte Carlo experiments) and in observed data. In the simulation exercises, we analyze whether the shrinkage method can help to improve the inference on the CAPM beta. In the empirical application, we use the shrinkage approach to verify if it delivers better estimates for the CAPM beta by comparing the beta estimates between methods. We also verify whether our proposal can help explain the cross-section of returns by running the two-step procedure of [26].

The results indicate that the regularization over the instruments coefficient improves the estimates of CAPM beta. In the Monte Carlo simulation analysis, we find that the regularized Bayesian instrumental variable dramatically reduces the mean bias found with the traditional 2SLS method. Moreover, the bias and root mean squared errors are smaller when using the regularization technique. This evidence shows that high-dimensional settings offer a better way to deal with the CAPM error-in-variables problem. Using many instruments, we can find more precise and unbiased measures of CAPM stock's beta, which helps us to evaluate the systematic risk accurately. In addition, more accurate estimates for betas in the first-pass time-series regression offer an adequate input for the second-pass cross-section regression in the two-step procedure of [26].

In the empirical application, the beta estimates using our approach present a subtle difference to OLS and 2SLS estimates, which varies across the assets. This difference in estimated betas is economically relevant since many financial models are sensitive to beta. To investigate further whether the difference among estimated betas is indeed relevant, we run the second step of the Fama–MacBeth procedure for both individual stocks and portfolios sorted by size and book-to-market. For the portfolio data, the results show that our approach can explain around 15% of the cross-section of portfolio returns, while the standard 2SLS method explains only 5%. For the stocks individually, the betas from our approach can explain 3% of the cross-section of a stock's return, compared with nearly 0% with the standard 2SLS method.

These results shed new light on the estimation of asset pricing models with measurement errors. By using high-dimensional data and proper techniques, we can attenuate the error-in-variables problem in the estimation of systematic risk, which in turn will perform a better job in explaining the cross-section of returns. We note that, even if the power

of explanation of the cross-section of returns found in our empirical application is small, compared with multifactor models, the use of many instruments and regularization priors can improve this power with regards to unregularized approaches.

We list our main contributions below:

- Addressing Roll's critique: the paper tackles Roll's (1977) critique, focusing on the application of instrumental variables approach, particularly within the context of the capital asset pricing model (CAPM).
- Utilizing abundant instruments: leveraging the data-rich environment in finance, the paper capitalizes on a diverse set of instruments to counteract the error-in-variables issue inherent in the CAPM.
- Bayesian shrinkage priors for robust estimation: the introduction of shrinkage priors to the instrument-associated parameters enhances the robustness of estimation, effectively mitigating bias arising from weak instrument correlations.
- Simulation results: through simulation exercises, the proposed methodology is shown to reduce bias significantly in estimating CAPM beta values.
- Empirical validation: empirical application demonstrates the efficacy of the approach, surpassing traditional OLS and 2SLS methods in certain stock cases. This difference in estimated betas proves economically significant in various financial models.
- Enhanced cross-sectional portfolio analysis: the methodology extends its benefits beyond individual stocks, aiding in cross-sectional portfolio analysis and contributing to the understanding of cross-sectional return heterogeneity.
- General applicability: the contribution's scope expands beyond CAPM estimation, encompassing multifactor models and risk factor construction, offering insights into addressing measurement errors and enhancing risk pricing methodologies.

This paper is organized as follows. Section 2 reviews the capital asset pricing model and introduces the notation used in the paper. Section 3 explains the Bayesian estimation and prior regularization. Section 4 presents the Monte Carlo analysis and the empirical results of the paper. We discuss our results in Section 5. Finally, Section 6 concludes the paper.

## 2. The CAPM and Measurement Errors

The seminal paper of [27] prepared the framework for the capital asset pricing model. The author established the investor problem in terms of a trade-off between risk and return, and defined the mean-variance efficiency concept of a portfolio allocation. This definition states that, for a given level of return, the portfolio is mean-variance efficient if it minimizes the variance. Refs. [1,2] worked on the results of [27] to analyze the implication for asset pricing and developed what is called the Sharpe–Lintner CAPM, or just CAPM.

By assuming that investors possess homogeneous expectations, refs. [1,2] showed that, in the absence of market frictions, if all investors choose an efficient portfolio, then the market portfolio is also mean-variance efficient. In this context, the market portfolio includes all assets in the economy, for instance, stocks, real estate, and commodities, which makes it an unobserved variable. In practice, usual surrogates for the market portfolio are market indexes, such as S&P500, but these indexes do not contain all assets and, consequently, the market portfolio is observed only with errors. Despite this practical difficulty, the efficiency of the market portfolio implies a relation between the asset risk premium and the market risk premium:

$$\mathbb{E}[R_i] - R_f = \beta_i (\mathbb{E}[R_m] - R_f), \quad (1)$$

where  $R_i$  denotes the return of asset  $i$ ,  $R_f$  is the return of the risk-free asset,  $R_m$  is the market portfolio, and  $\beta_i \equiv \sigma_{im} / \sigma_m^2$ , with  $\sigma_{im}$  being the covariance between asset  $i$  and the market portfolio and  $\sigma_m^2$  is the variance of the market portfolio. Therefore, the CAPM summarized in Equation (1) is an equilibrium result that holds for a single period.

The relation established in Equation (1) for one period is not enough to assess the CAPM empirically. To proceed with econometric analysis, an additional assumption is required: the returns are independent and identically distributed along time and multivariate Gaussian. Although this hypothesis is a strong one, it possesses some benefits. First, it is consistent with the CAPM holding for each period in time. Moreover, it is a good approximation for monthly returns [28]. Under this assumption, the CAPM may be represented by the single index model, which is described by

$$R_{it} - R_{ft} = \gamma_i + \beta_i(R_{mt} - R_{ft}) + \varepsilon_{it}, \quad \varepsilon_{it} \sim \mathcal{N}(0, \sigma_i^2). \tag{2}$$

where we add a time index,  $t$ , to each variable, and  $\gamma_i$  is an intercept, representing a mean return for the asset  $i$  not explained by the market portfolio. In Equation (2), if  $\gamma_i$  is equal to zero, then the CAPM holds for each period in time.

The representation of the CAPM model given by Equation (2) started a testing tradition that became known as the time series approach. To test the CAPM model empirically, ref. [29] proposed to use time series for return of assets, for return risk-free assets, and as a proxy to the return of the market portfolio to estimate Equation (2). The usual choice for the risk-free asset is the US Treasury Bill and S&P500 for the return of the market portfolio. Then, their approach suggests testing whether the estimated intercept is equal to zero, which may be performed using a Wald test or the test proposed by [30].

Testing the CAPM using the approach of [29] is problematic once the return of the market portfolio,  $R_{mt}$ , is a variable contaminated by measurement errors. The source of measurement errors appears because the market indexes used to estimate the model contain only a subset of assets. Moreover, even if all universe of assets are observed, the measurement error can appear due to misspecification in the weights of assets. This problem is known as Roll’s critique, due to [3], who argued that, once the market portfolio is not observed, the CAPM cannot be tested. According to this author, a rejection of the CAPM can be due to measurement errors in the return of the market portfolio. In an econometric sense, the present problem is a case of classical measurement errors and should be treated as such.

To put the problem in terms of classical measurement errors, let  $\tilde{R}_{mt}$  denote the observed return of the market portfolio. Also, denote by  $x_t^* \equiv R_{mt} - R_{ft}$  the excess of return of the true market portfolio, by  $x_t \equiv \tilde{R}_{mt} - R_{ft}$  the excess of return of the observed market portfolio, and let  $u_t$  be the measurement error, assuming that this process is independent. The excess of the return of asset  $i$  is denoted by  $y_{it}$ , and there is no error-in-variables in this case. Instead of Equation (2), the model to be estimated to test the CAPM should be

$$y_{it} = \gamma_i + \beta_i x_t + \varepsilon_{it}, \quad \varepsilon_{it} \sim \mathcal{N}(0, \sigma_i^2), \tag{3}$$

$$x_t = x_t^* + u_t, \quad u_t \sim \mathcal{N}(0, \sigma_u^2). \tag{4}$$

Equation (4) assumes that the measurement error is additive. If one ignores this additive measurement error and estimates Equation (3) using least squares, then the estimates of betas will suffer from attenuation bias and the intercept will be upward biased, implying positive alphas, even if the CAPM holds. Thus, to deal appropriately with the error-in-variables problem, Equations (3) and (4) must be considered to estimate and test the CAPM model.

### 3. Methods and Data

The data-rich environment available in the financial data set allows us to use many instruments to correct the bias caused by measurement errors, even though these instruments are possibly weak. The many instruments setting needs to be used carefully, as it can itself be a source of bias. To overcome this inconvenience, we need a regularization step, such as variable selection or shrink of less important parameters. Examples of regularization

methods are least absolute shrinkage and selection operator (LASSO), ridge, elastic net, or via Bayesian shrinkage priors, which penalize the number of covariates in some form.

In instrumental variable regression, it is interesting to use a method that jointly estimates “two stages”, and the Bayesian approach has this advantage. The regression of the treatment variable on the instruments and the estimation of the target variable on the treated variable can be estimated in a single step. In this sense, the Bayesian shrinkage priors are preferred rather than other regularization methods. In particular, the factor-based prior proposed by [25] has the advantage of linearly combining the information in all possibly weak instruments in such a way that, taken together, makes them stronger. In the next subsection, we present this structure of shrinkage prior to the IV regression context.

### 3.1. Bayesian Regularization Methods in IV Regression

When dealing with measurement error, instrumental variable regression may be used. Consider the model:

$$\begin{aligned} x_t &= z_t'\delta + \varepsilon_{x_t}, & (5) \\ y_t &= \gamma + x_t\beta + \varepsilon_{y_t}, \quad t \in \{1, \dots, n\} & (6) \end{aligned}$$

where  $x_t$  is the endogenous or treatment variable,  $z_t$  is a  $(p \times 1)$  vector of instruments, and  $y_t$  is the response variable, and it is supposed that

$$\begin{bmatrix} \varepsilon_{x_t} \\ \varepsilon_{y_t} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{yx} & \sigma_y^2 \end{bmatrix}\right)$$

Equation (6) represents the projection of the covariate,  $x$ , onto the instrument vector, and is the functional form used in the first stage of instrumental variable methods. Since  $p$  may be large, some regularization on Equation (5) is necessary. The Bayesian solution to this problem is to impose shrinkage priors on  $\delta$  to shrink those parameters that have little power to explain  $x_t$ . By imposing such a prior, the usual Gibbs sampler scheme [31] used to estimate model (5) and (6) cannot be employed. Ref. [25] developed an elliptical slice sampler that can deal with arbitrary priors on  $\delta$ , allowing us to use a shrinkage prior, such as Laplace distribution, as well as the factor-based-prior, also developed by the same authors. Then, it is instructive to describe the estimation for an arbitrary prior distribution on  $\delta$ . To understand the Bayesian estimation of IV regression, consider the reduced form of Equations (5) and (6):

$$x_t = z_t'\delta + v_{x_t} \tag{7}$$

$$y_t = \gamma + z_t'\delta\beta + v_{y_t} \tag{8}$$

where  $v_{x_t} \equiv \varepsilon_{x_t}$  and  $v_{y_t} \equiv \beta\varepsilon_{x_t} + \varepsilon_{y_t}$ . Defining  $T = \begin{bmatrix} 1 & 0 \\ \beta & 1 \end{bmatrix}$  implies that:

$$\Omega \equiv \text{Cov}\left(\begin{bmatrix} v_{x_t} \\ v_{y_t} \end{bmatrix}\right) = TST' = \begin{bmatrix} \sigma_x^2 & (\alpha + \beta)\sigma_x^2 \\ (\alpha + \beta)\sigma_x^2 & (\alpha + \beta)^2\sigma_x^2 + \zeta^2 \end{bmatrix},$$

with  $\alpha \equiv \frac{\sigma_y}{\sigma_x}\rho$  and  $\zeta^2 \equiv (1 - \rho^2)\sigma_y$ , where  $\rho \equiv \sigma_{xy}/(\sigma_x\sigma_y)$ . Note that the parameters to be estimated are  $\Theta = (\sigma_x^2, \delta, \zeta^2, \gamma, \beta, \alpha)$ . Then, conditional on the set of instruments, the likelihood function may be written as:

$$\begin{aligned} f(x, y|z, \Theta) &= f(y|x, \delta, \alpha, \beta, \zeta^2) \times f(x|z, \sigma_x^2, \delta) \\ &= \mathcal{N}\left(\gamma + x_t\beta + \alpha(x_t - z_t'\delta), \zeta^2\right) \times \mathcal{N}\left(z_t'\delta, \sigma_x^2\right). \end{aligned} \tag{9}$$

This decomposition of the likelihood function allows us to form a Gibbs sampler scheme by choosing the following prior distributions:

$$\delta \sim \text{arbitrary}, \tag{10}$$

$$\sigma_x^2 \sim \mathcal{IG}(\text{shape} = k_x, \text{scale} = s_x), \tag{11}$$

$$(\zeta^2, \gamma, \beta, \alpha)' \sim \mathcal{NIG}\left(0, \zeta^2 \Sigma_0^{-1}, \text{shape} = \frac{k}{2}, \text{scale} = \frac{s}{2}\right). \tag{12}$$

Combining these priors with the likelihood function in Equation (9) gives us the posterior distribution. To sample from this posterior distribution, it is possible to break it into three full conditional posteriors to form a Gibbs sampler scheme. To explain these three blocks, it is useful to introduce some definitions. In general, the notation with no subscript,  $t$ , means that the variable contains all observations. For example,  $x \equiv (x_1, x_2, \dots, x_n)$ . A particular case is the vector of instruments,  $z_t$ , for which all observations are denoted by  $Z$  since it becomes a matrix of dimension  $(p \times n)$ .

Define  $\tilde{x} \equiv (1, x, z'\delta)$ ,  $M \equiv \Sigma_0 + \tilde{x}'\tilde{x}$ ,  $a \equiv k + n$  and  $b \equiv s + y'y - y'\tilde{x}M^{-1}\tilde{x}'y$ . It is possible to show that  $f(y|x, z, \delta) \propto |M|^{-\frac{a}{2}} b^{-\frac{1}{2}}$ . With these definitions, we can describe each of these blocks.

Full conditional posterior for  $\delta|\Theta$ , data.

Given  $\Theta$ , from Equations (9) and (10), the conditional posterior is proportional to  $f(x|\Theta)f(y, |x, \Theta)\pi(\delta)$ . Since we are considering an arbitrary prior for  $\delta$ , this full conditional posterior may not have a closed form, requiring alternative methods to sample it. Although traditional Metropolis–Hastings steps can be used in this case, it scales poorly due to the possibly high dimension and multimodality of the full conditional posterior. Instead, ref. [25] proposed to sample it using an elliptical slice sampler, which only requires the ability to evaluate  $\pi(\delta)$ . This algorithm is described below:

Note that the only requirement of Algorithm 1 is the ability to evaluate the prior density,  $\pi(\delta)$ .

---

**Algorithm 1** Elliptical slice sampler.

---

- 1: **procedure** SLICE SAMPLER( $\delta, \sigma_x^2, x, Z, y$ )
  - 2:   Define  $\hat{\delta} = (ZZ')^{-1}Zx$  and  $\Delta = \delta - \hat{\delta}$
  - 3:   Draw  $\zeta \sim \mathcal{N}(0, \sigma_x(ZZ')^{-1})$  and  $v \sim U(0, 1)$
  - 4:   Compute  $\ell \equiv \log(f(y|x, Z, \delta) + \log(\pi(\delta)) + \log(v)$
  - 5:   Draw an angle  $\varphi \sim U(0, 2\pi)$ , and do Lower  $\leftarrow \varphi - 2\pi$  and Upper  $\leftarrow \varphi$
  - 6:   Update  $\Delta$  and  $\delta$ :  $\bar{\Delta} = \Delta \cos(\varphi) + \zeta \sin(\varphi)$  and  $\bar{\delta} = \hat{\delta} + \bar{\Delta}$
  - 7:   **while**  $\log(f(y, x, Z, \bar{\delta})) + \log(\pi(\bar{\delta})) < \ell$  **do**
  - 8:     If  $\varphi < 0$ , then Lower  $\leftarrow \varphi$ . Else Upper  $\leftarrow \varphi$
  - 9:     Draw a new angle  $\varphi \sim U(\text{Lower}, \text{Upper})$
  - 10:    Update  $\Delta$  and  $\delta$ :  $\bar{\Delta} = \Delta \cos(\varphi) + \zeta \sin(\varphi)$  and  $\bar{\delta} = \hat{\delta} + \bar{\Delta}$
  - 11:    $\delta \leftarrow \hat{\delta} + \bar{\Delta}$ .
  - 12:   **return**  $\delta$
- 

Full conditional posterior for  $\sigma_x^2|\Theta$ , data.

Fortunately, for an inverse gamma prior on  $\sigma_x^2$ , the full conditional posterior for  $\sigma_x^2$  has a closed form. Combining the likelihood  $f(x|Z, \delta, \sigma_x^2)$  with the prior given in (11), it is possible to show that the full conditional posterior is an inverse gamma with shape parameter  $k_x + n$  and scale  $s_x + \sum_{i=1}^n (x_i - z_i'\delta)^2$  (see proof in Appendix A).

Full conditional posterior for  $(\zeta^2, \gamma, \beta, \alpha)'|\Theta$ , data.

This block also has a closed form. By using the bivariate normal properties, we can write the likelihood in terms of the transformed variable,  $\tilde{x}$ , which follows  $y_t|\tilde{x}_t \sim N(\tilde{x}_t\theta, \zeta^2)$ .

Combining this likelihood with the prior give in (12), it can be shown that the full conditional posterior  $(\xi^2, \gamma, \beta, \alpha)' | \Theta, \text{data}$  follow a normal inverse gamma distribution. Specifically:

$$(\xi^2, \gamma, \beta, \alpha)' | \Theta, \text{data} \sim \mathcal{NIG} \left( M^{-1} \bar{x}' y, \xi^2 M^{-1}, \text{shape} = \frac{a}{2}, \text{scale} = \frac{b}{2} \right).$$

(See Appendix A for proof).

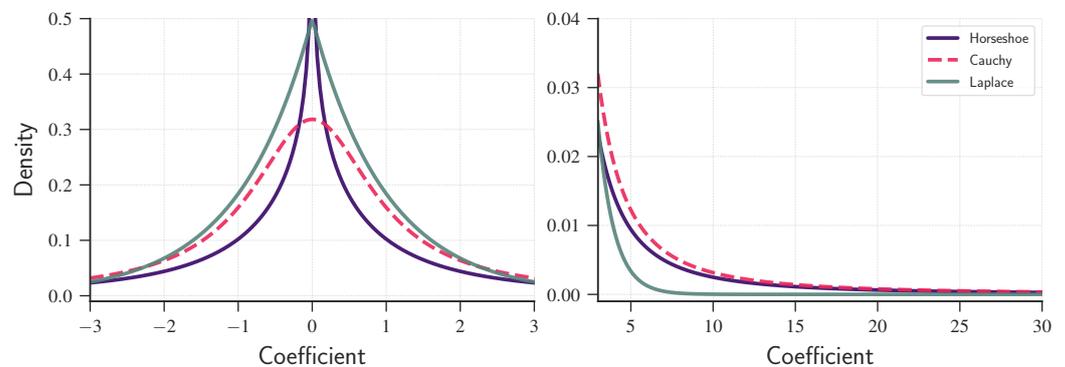
We can use these full conditional posteriors to form a three-block Gibbs sampler, by iteratively sampling over the blocks. This methodology is interesting because we can choose arbitrary priors for  $\delta$ , and it still works well. In particular, we can elicit several shrinkage priors over  $\delta$ , since the many instruments setting requires regularization.

### 3.2. Shrinkage Priors for Instruments Coefficients

There is a large range of shrinkage priors in the literature [32]. The underlying idea of these priors is to give a higher prior probability around zero, such that, if the parameter is not too important, it shrinks to zero. In what follows, we present some of these priors that can be directly applied to  $\delta = (\delta_1, \dots, \delta_p)$ . Then, we proceed with the factor-based prior distribution.

#### 3.2.1. Heavy-Tailed Priors

Popular choices of shrinkage priors are the Cauchy, Laplace, and horseshoe densities [33]. The horseshoe density is the stronger one, in the sense that it concentrates the high-probability density around zero. In the same sense, the Laplace prior is strong, while the Cauchy density is relatively weaker, although it also concentrates the density around zero, and so it also works as a shrinkage prior. The left panel of Figure 1 depicts these three priors.



**Figure 1.** Three examples of shrinkage priors: horseshoe, double-exponential, and Cauchy, all of which are centered at zero.

An advantage of the horseshoe prior is that, at the same time as concentrating the high-probability density around zero, shrinking unimportant coefficients to this point, it also has heavy tails. The right panel of Figure 1 compares the right tails of the three priors considered here. Notice that the tail of the horseshoe prior is above the Laplace and closer to the Cauchy tail. This heavy tail allows the identification of parameters that are different from zero.

In the IV regression case, we can choose one of these priors for each  $\delta_j$  and assume that they are independent for all  $i \neq j$ , with  $i, j \in \{1, \dots, p\}$ . Although it may work, it neglects the covariance between the instruments. To consider the covariance between the instruments, a more sophisticated prior is required, and this subject is discussed in the next subsection.

### 3.2.2. Factor-Based Shrinkage Prior

The idea underlying the factor-based prior, proposed by [25], is to explore the covariance of instruments to extract factors that represent ‘strong’ instruments. To formalize this intuition, consider the following decomposition of the covariance matrix of instruments:

$$\text{Cov}(z_t) = BB' + \Psi^2, \tag{13}$$

where  $B$  is a  $(p \times k)$  matrix and  $\Psi^2$  is a diagonal  $(p \times p)$ . Despite the fact that every covariance matrix admits this decomposition, the interest here is in the case where  $k \ll p$ , where  $k$  represents the number of factors to be extracted, denoted by  $f_t$ . Suppose that the instruments,  $z_t$ , and the factors,  $f_t$ , are jointly normally distributed as follows:

$$\begin{bmatrix} z_t \\ f_t \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} BB' + \Psi^2 & B \\ B' & I_k \end{bmatrix}\right).$$

This assumption implies that  $\mathbb{E}[f_t|z_t] = Az_t =: \hat{f}_t$ , with  $A \equiv B'(BB' + \Psi^2)^{-1}$ . Now, consider the factor regression model:

$$x_t = \theta \hat{f}_t + \varepsilon_t, \tag{14}$$

where  $\theta$  is a  $(1 \times k)$ . From Equation (14) and the definition of  $\hat{f}_t$ , it is possible to show that  $\delta' = \theta A$ . However, this specification is only correct if  $\delta$  lies in the row space of  $A$ ; otherwise, the model is misspecified. Then, it is necessary to extend the model to include the possibility that  $\delta$  lies in the row space of  $A$ . For that end, the specification in Equation (14) needs to be modified to

$$x_t = \theta \hat{f}_t + \eta \hat{r}_t + \varepsilon_t \tag{15}$$

where  $\eta$  is a  $(1 \times p)$  vector of parameter and  $\hat{r}_t \equiv (I_p + A^+A)z_t$  and  $A^+$  denote the Moore–Penrose pseudo-inverse of  $A$ . In this case, it can be shown that  $\delta' = \theta A + \eta(I_p + A^+A)$ .

Defining  $\tilde{\delta}' = (\theta, \eta)$ , we note that  $\delta = H\tilde{\delta}$ , where

$$H' = \begin{bmatrix} A \\ I_p + A^+A \end{bmatrix}.$$

Consequently, we can rewrite (15) as  $x_t = H\tilde{\delta}z_t + \varepsilon_t$ .

Assuming we know  $A$  (and then  $H$ ), this specification allows us to attribute a prior over  $\delta$  by imposing a strong shrinkage prior over  $\tilde{\delta}$ . If we solve the system  $\delta = H\tilde{\delta}$ , using the theory of pseudo-inverses, we have  $\tilde{\delta} = H^+\delta + (I_{p+k} + H^+H)\omega$ , for an arbitrary vector  $\omega$ . With this identity, conditional on  $\omega$ , we can impose a horseshoe prior, for instance, on  $\tilde{\delta}$ , and it induces a prior on  $\delta$ . That is:

$$\begin{aligned} \pi(\delta|\omega) &= \prod_{j=1}^{k+p} \left\{ (2\pi^3)^{-\frac{1}{2}} \log\left(1 + \frac{4}{\tilde{\delta}_j^2}\right)\right\} \\ \pi(\delta|\omega) &= \prod_{j=1}^{k+p} \left\{ (2\pi^3)^{-\frac{1}{2}} \log\left(1 + \frac{4}{\left\{ [H^+\delta + (I_{k+p} - H^+H)\omega]_j \right\}^2}\right)\right\}. \end{aligned}$$

Following [25], we assume that  $\omega \sim \mathcal{N}(0, I_{k+p})$ . Once we know  $\omega$ , we can evaluate the prior  $\pi(\delta|\omega)$ , which is the only requirement of the slice sampler presented in Algorithm 1. Then we can sample  $\delta$  by inducing a prior on  $\delta$  via a horseshoe prior over  $\tilde{\delta}$ . Note that, under this specification, the factor structure derived in this section is taken into account in the prior over  $\delta$ . In practice, however, the matrices  $B$  and  $\Psi$  are unknown and, consequently,  $A$  and  $H$  are also unknown. Instead of estimating it in a Bayesian fashion, we use point

estimates of these matrices, which are found by minimizing the trace of  $\text{Cov}(z_t) - D$ , by choosing  $D$ , subject to  $D$  be diagonal and positive definite.

Finally, for all shrinkage priors, we incorporate a global shrinkage parameter  $\lambda$ . Introducing this as a parameter in the model is a key feature in Bayesian regularization because it avoids procedures like cross-validation or setting it as a fixed parameter. We sampled the global shrinkage parameter via a Metropolis–Hastings step.

### 3.3. Data

To analyze whether our empirical method performs well, we start by using it in simulated data by means of a Monte Carlo analysis. When we know the true generating process, we can calculate error measures of the estimate (mean bias, mean absolute bias, and root mean squared error) and compare it with alternative methodologies (for instance, OLS, 2SLS, etc.). Besides, in the simulation exercise, we also apply the empirical method to real financial data. To estimate the CAPM, we need asset return data, a surrogate for the market return, and a risk-free asset. As the risk-free asset, we consider the one-month Treasury Bill rate and take the surrogate of the market return data from Kenneth French’s website. The returns on market return and the risk-free asset are taken from <https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/> (accessed on 31 January 2022). The returns of the stocks are based on the close price of the stocks and are taken from Yahoo Finance. Finally, we consider the return of 277 stocks listed in the S&P500 with data availability in the last five years. All data are daily and range from 1 January 2017 to 31 December 2021, resulting in 1260 observations.

## 4. Results

In this section, we describe and discuss the results of our paper. We begin by describing the outcome of a simulation exercise, in which we compare the Bayesian regularization discussed in the previous section with traditional ordinary least squares and two-stage least squares estimation. Then, we present the result of the CAPM estimation for observed data using the proposed Bayesian shrinkage approach.

### 4.1. Monte Carlo Analysis: Simulation Procedures

To simulate the CAPM, we consider a classical additive measurement error model, as follows:

$$x_t = x_t^* + u_t, \quad u_t \sim \mathcal{N}(0, \sigma_u^2) \text{ and } x_t^* \sim \mathcal{N}(0, \sigma_x^2), \tag{16}$$

$$y_{it} = \beta_i x_t^* + \varepsilon_{it}, \tag{17}$$

for  $i \in \{1, \dots, p + 1\}$  and  $t \in \{1, \dots, n\}$ . In Equations (16) and (17),  $x_t^*$  is the true market return, and it is assumed to be Gaussian with mean zero and variance  $\sigma_x^2$ ,  $u_t$  is the Gaussian measurement error, with mean zero and variance  $\sigma_u^2$ , and  $x_t$  is the observed market return. The sensitivity to the market return, which is measured by  $\beta_i$ , is assumed to be known, and, based on these values and given the error term,  $\varepsilon_t$ , we construct the assets return, as described in Equation (17). The error term,  $\varepsilon_t$ , is also assumed to be Gaussian, with mean zero and variance  $\sigma_\varepsilon^2$ .

To simulate the model, we need to calibrate the parameters  $(\sigma_u^2, \sigma_x^2, \sigma_\varepsilon^2, \beta_i)$ . For  $\beta_i$ , we consider a linear grid between 0.3 and 1.5. Based on data of a proxy of market return, we calibrated  $\sigma_x = 0.01$ . We calibrated one of the assets with  $\sigma_\varepsilon = 0.001$  and the other  $p - 1$  with  $\sigma_\varepsilon = 0.9$ . We use these different values for  $\sigma_\varepsilon$  to create one strong instrument and  $p - 1$  weak instruments. A lower standard deviation creates assets that will be stronger instruments than those with a higher standard deviation. Besides these  $p$  assets, we consider an additional one with  $\beta = 1$  and  $\sigma_\varepsilon = 0.04$ , which is used as the target variable in the CAPM estimation. Finally, we calibrate  $\sigma_u = 2\sigma_x$  to create a situation with a high measurement error. These parameter calibrations allow us to simulate all variables of

interest in the CAPM. We set the hyper-parameters of the remaining prior distribution parameters to reflect a diffuse prior.

To evaluate the accuracy of each method, we simulate the model  $N_{\text{sim}} = 1000$  times. At each iteration, we estimate the parameters using six methods. The first one is the traditional OLS estimator, regressing  $y_{it}$  on  $x_{it}$ , which is known to be inconsistent in the presence of measurement errors. Second, we consider the 2SLS estimator, using all asset returns but the regressand in the CAPM equation as instruments (see [10], for a similar approach). We believe that these variables satisfy the requirement of an instrument: they are correlated with the market return but uncorrelated with the error term in the CAPM equation. Third, we use the limited information maximum likelihood (LIML) [34] estimator with the same set of instruments. Although our main interest is to verify whether the Bayesian regularization of the two-stage method can improve the inference about beta, we include the results of the LIML estimator since it is known to be unbiased in the presence of many instruments [35]. The LIML estimator, however, is also known to have no moments. The modified version of LIML, due to [34], solves this drawback. Still, the modification introduces an additional parameter that must be chosen by the econometrician. In the last three methods, we consider the same set of instruments to estimate the model using the Bayesian method described in Section 3. We use the horseshoe, Laplace, and factor-based shrinkage prior distributions over  $\delta$ . These estimations are referred to as the BHS, BLA, and BFB.

We simulate and estimate the model for different numbers of assets and hence different numbers of instruments. Specifically, we start with  $p = 2$ , and then increase it to 10, 20, 40, 80, and 160. In the estimation process, we consider the asset with  $\beta = 1$ . Because of measurement error, the OLS estimator is downward biased. To assess the ability of each estimation method to correct this bias, we use three criteria: mean bias, mean absolute bias, and root mean of squared error (RMSE). According to these criteria, the lower its absolute value is, the better the estimator is. Since we are in a context of measurement error, the OLS beta estimates are downward biased. Even for the standard 2SLS method, or its regularized versions, we cannot ensure it will be unbiased in finite samples. Thus, we consider these three criteria to evaluate both the bias (mean bias and mean absolute bias) and the variance (RMSE) since the RMSE is a decomposition of the squared bias and the variance. Thus, by using these criteria, we can evaluate the bias and the efficiency of the estimators.

Table 1 summarizes the Monte Carlo results. In this table, we highlight in bold the best estimator for the criterion and the number of instruments. The Bayesian “two-stage” procedure with the horseshoe regularization prior (BHS) is better than the traditional 2SLS method for all criteria and choices of  $p$ , except for the mean bias criteria and  $p = 2$ . These results show that the regularization over the instrumental variables indeed avoids bias in the presence of many weak instruments. The mean bias of the BHS is closer to the mean bias of the LIML. For  $p = 80$ , it is even smaller than that of the LIML, which corrects the many instruments bias [35]. We note that, for the other values of  $p \neq 80$ , the LIML has a smaller mean bias. However, considering the mean bias and the root mean squared error, which also penalizes for variance, the Bayesian approach always dominates the LIML, except for  $p = 2$ . Thus, we conclude that the Bayesian regularization in the estimation of the CAPM model improves the inference about the betas in two ways: decreasing the bias and reducing the estimates’ variance.

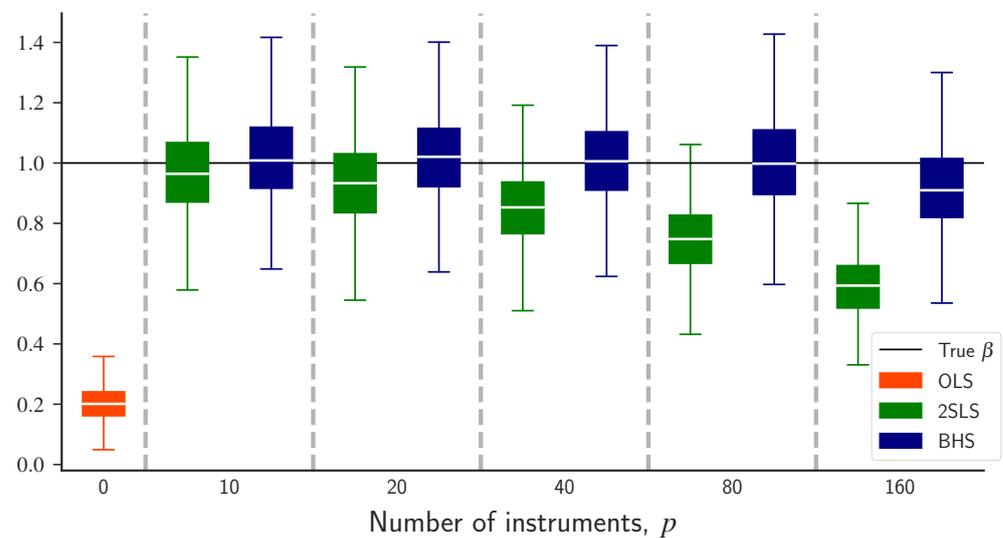
Among the three types of regularization priors, the horseshoe outperforms the others at least in terms of mean bias, except for  $p = 10$ . For this reason, we now focus on the horseshoe prior in comparison with the traditional 2SLS method. In Figure 2 we present the distribution of the BHS and 2SLS for several numbers of instruments, as well as the OLS estimates. For small sets of instruments, say up to 20, the 2SLS bias is small and can be entirely corrected by Bayesian regularization. When the number of instruments increases to 40 and 80, the bias becomes greater and the Bayesian regularization still entirely corrects the bias, maintaining the distribution of the estimates around the true CAPM beta, which is one. The bias of the traditional 2SLS method can be explained by the tendency of the OLS, in the first stage, to

fit too well as the number of instruments increases, see [36] p. 222. The BHS penalizes the “first-stage” estimation, avoiding this tendency to overfitting, and hence reducing the bias (see Figure 2). The size of the correction diminishes when we increase the number of instruments to 160, but it is still better than 2SLS, as shown both by Figure 2 and Table 1.

**Table 1.** Measures for the beta estimation error,  $n = 1000$ .

		OLS	2SLS	LIML	BHS	BLA	BFB
$p = 2$	Mean bias	-0.798	<b>-0.001</b>	0.003	0.007	0.012	0.013
	Mean abs. bias	0.798	0.633	<b>0.117</b>	0.117	0.119	0.119
	RMSE	0.800	4.970	<b>0.143</b>	0.144	0.146	0.146
$p = 10$	Mean bias	-0.799	-0.029	0.007	0.017	0.018	<b>-0.005</b>
	Mean abs. bias	0.799	0.123	0.121	0.120	0.123	<b>0.119</b>
	RMSE	0.801	0.154	0.151	0.151	0.155	<b>0.149</b>
$p = 20$	Mean bias	-0.799	-0.066	<b>0.012</b>	0.021	0.023	-0.055
	Mean abs. bias	0.799	0.126	0.122	0.121	0.125	<b>0.120</b>
	RMSE	0.801	0.158	0.155	0.155	0.160	<b>0.152</b>
$p = 40$	Mean bias	-0.798	-0.146	<b>0.003</b>	0.010	0.006	-0.166
	Mean abs. bias	0.798	0.167	0.126	<b>0.122</b>	0.159	0.181
	RMSE	0.800	0.198	0.161	<b>0.158</b>	0.308	0.211
$p = 80$	Mean bias	-0.803	-0.252	0.017	<b>0.008</b>	-1.133	-0.320
	Mean abs. bias	0.803	0.253	0.135	<b>0.129</b>	1.234	0.320
	RMSE	0.805	0.279	0.172	<b>0.163</b>	1.444	0.339
$p = 160$	Mean bias	-0.799	-0.407	<b>0.010</b>	-0.079	-1.149	-0.508
	Mean abs. bias	0.799	0.407	0.158	<b>0.138</b>	1.153	0.508
	RMSE	0.801	0.421	0.199	<b>0.168</b>	1.249	0.517

Note: We report mean bias, mean absolute bias and the root mean squared error (RMSE) of each estimator. We highlight in bold type the best estimator for each criterion and number of instruments.



**Figure 2.** Boxplots of the estimated CAPM betas for different numbers of instruments: comparison between Bayesian horseshoe (BHS) and two-stage least squares (2SLS). The horizontal black line represents the true CAPM beta value, which is one.

This is an important limitation that affects estimators based on regularization/selection structures in the high-dimensional setting. Although the shrinkage framework is effective in reducing model complexity, the Bayesian shrinkage estimators, and their frequentist competitors, strike a balance between bias and variance. As the number of variables increases, the penalty imposed on the coefficients may lead to bias in the estimates. This bias arises from the tendency of shrinkage estimators to force some coefficients towards zero, even if they are truly non-zero. This trade-off becomes more pronounced when dealing with a large number of variables, potentially leading to underestimation of the true relationships.

As discussed in [37,38], the assumption for the validity of shrinkage estimators, whether in a frequentist or Bayesian context, is that the true number of statistically relevant variables in the final model must be less than the sample size. Thus, variable selection methods in a context of sparsity allow starting from a number of possible covariates greater or much greater than the number of observations, but the final number of variables included in the final specification must be smaller than the sample size. Theoretical conditions for the consistency of the selection procedure (the Oracle properties) for Bayesian estimators based on shrinkage priors are discussed in [37] (see Theorems 1–5), but are difficult to be evaluated in empirical models due to dependence on quantities not observed. Another important point is that, with more variables, the computational burden of performing Bayesian estimation increases substantially, and this point is especially important due to the fact that we are using MCMC estimators to estimate the hyperparameters of the models, and, in situations of high dimensionality, problems of convergence of chains can become worse, also affecting the properties of the estimators in finite samples. In Appendix A, we present a robustness check, in which we change some parameters of the simulated CAPM.

#### 4.2. Empirical Application

This section uses the Bayesian regularization procedure to estimate the CAPM using observed data and compare it with the traditional 2SLS, LIML, and OLS estimators of the CAPM beta. We estimate the model for 277 stocks listed in the S&P500 index. The instruments for each model consist of all other stock returns listed in the S&P500, except the one used in the CAPM target equation. The original set contains 365 stocks for the period between January 2017 and December 2021, but we exclude 88 stocks with a correlation above 0.7. This exclusion is needed to avoid numerical approximation errors in the inversion of  $ZZ'$ , a requirement of Algorithm 1. All data are daily and range from January 2017 to December 2021, totaling 1260 observations.

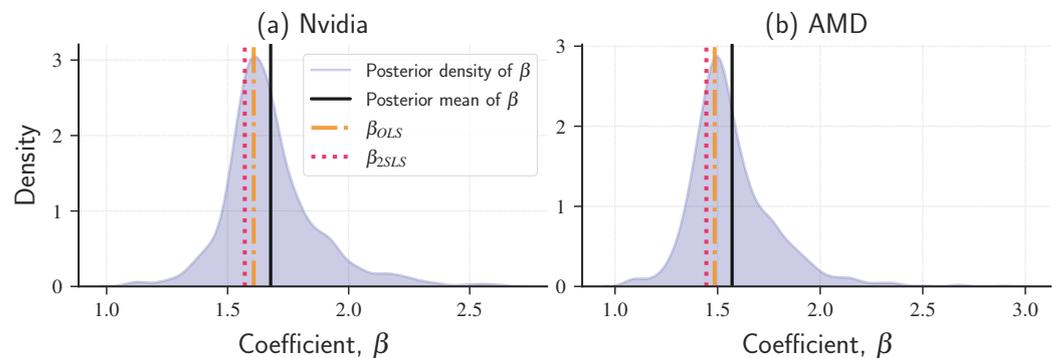
The posterior distribution for the CAPM beta was similar for the factor-based prior and Laplace prior (BFB and BLA, respectively, see Figure A1 in Appendix A). This result indicates that, for these stocks, these two priors perform equally. The reason why the factor-based shrinkage prior cannot do a better job than the straight horseshoe prior may be related to the covariance structure of the instruments. Indeed, the eigenvalues of the covariance matrix decay drastically from the first to the second eigenvalues and then decline slowly. The covariance of the minimized trace (the  $\text{Cov}(z_t) - D$  discussed in Section 3.2.2) presents a similar behavior (see Figure A2 in Appendix A). Thus, we cannot isolate “commonalities” and, hence, the prior information is unable to help in the shrinkage of parameters. The horseshoe prior presents a subtly different result, compared with BLA and BFB (see Figure A1 in Appendix A). We focus on the horseshoe prior because of its better performance in the simulation exercise.

When comparing the Bayesian IV (BIV) estimates with the OLS and 2SLS estimates, we find different results across the stocks. While the Bayesian IV delivers greater betas than the other methods for some stocks, there is also a group of stocks whose BIV estimates are less than (or very similar to) the unregularized estimates. On one hand, a beta greater than the OLS is a better estimate, since the OLS is downward biased in the presence of the EIV. On the other hand, the appearance of betas smaller than the OLS may reflect another hypothesis that may be failing in the model, such as omitted variables.

Figure 3 shows two examples of stocks where BIV delivers a greater beta than the other estimators. The stocks are Nvidia (NVDAO) and AMD, which are enterprises from the technology sector and present the highest return in our data set. Since they possess the high returns we expected, according to the CAPM theory, they hold a higher systematic risk and, consequently, a higher beta.

For both the Nvidia and AMD stock, the 2SLS and OLS estimates are similar, though the 2SLS is slightly smaller. It is an expected result since, when the number of the instruments is large (as in this case,  $p = 277$ ) relative to the number of observations ( $n = 1260$ ), the 2SLS estimates tend to the OLS estimates. In the extreme case, where  $p = n$ , the two

estimators are equivalent. In addition, since the market return contains a measurement error, we know that the OLS estimates are downward biased and, consequently, so are the 2SLS estimates.



**Figure 3.** Posterior distribution of the CAPM beta estimated by the Bayesian instrumental variable with the horseshoe prior for three assets: (a) Nvidia and (b) AMD. The figure also presents the mean of the posterior distribution and the OLS and 2SLS estimates.

The Bayesian approach, in turn, delivers a greater CAPM beta for the two stocks, considering the posterior mean as a point estimation (see black line in Figure 3). For Nvidia stock, while  $\hat{\beta}_{OLS} = 1.61$  and  $\hat{\beta}_{2SLS} = 1.57$  the posterior mean of beta is 1.68, using the horseshoe prior. For AMD stock, while  $\hat{\beta}_{OLS} = 1.49$  and  $\hat{\beta}_{2SLS} = 1.44$ , the posterior mean of beta is 1.68. Although we do not know the true beta in this case, we know that OLS is downward biased, which puts our approach in a better position. Even though the difference between the BIV and 2SLS beta estimates is not too stark, these discrepancies in the estimated beta may have drastic implications for finance practitioners. As noted by [6], some analysis in finance, such as valuation, is very sensitive to the estimated beta. In valuation procedures, small differences in beta estimation lead to relevant changes in the final value, as cash flows are discounted using the estimated beta for an infinite sequence of future values. Thus, the many instruments approach with shrinkage priors can offer a better way to conduct such financial analysis.

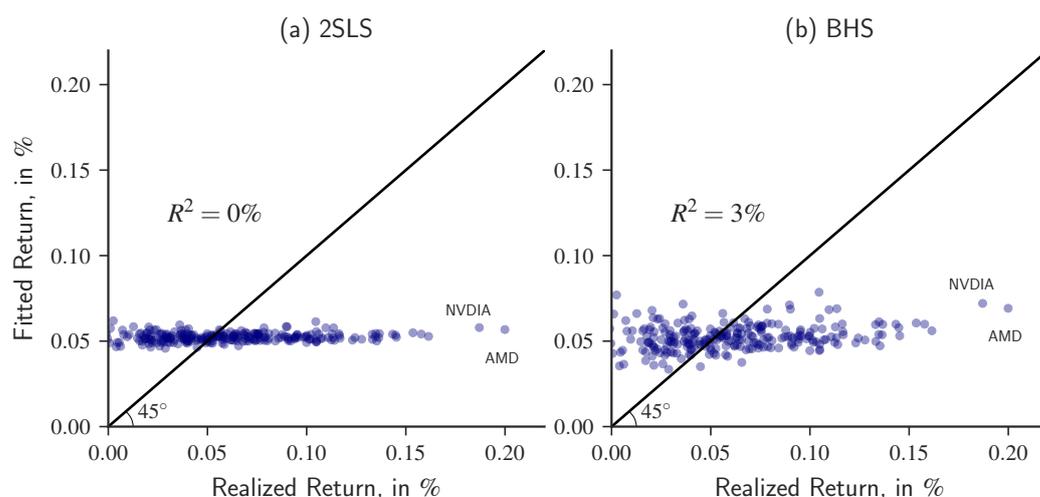
The above discussion only considers two stocks with the highest return of the analyzed data set. We extend this analysis to include all 277 stocks and, hence, consider all the differences among the different estimators. To do so, we analyze how the estimated betas of each estimator can explain the heterogeneity in the cross-section return by running the second-pass regression of the two-pass procedure proposed by [26]. The idea of the second-pass regression is to regress the cross-section of the returns against the betas. Thus, we can verify how much of the cross-sectional variation in the returns is explained by the betas. Specifically, we follow an approach similar to [39], regressing:

$$\bar{R}_i = \lambda_0 + \lambda_1 \hat{\beta}_i + u_i \tag{18}$$

where  $\bar{R}_i \equiv \sum_{t=0}^n R_{it}$ .

The second-pass regression also suffers from the EIV problem because  $\hat{\beta}_i$  carries the error estimation from the first pass (the time-series regression). As argued in the simulation exercise, our approach can improve the estimation of the betas both in terms of bias and variance. Thus, a more efficient estimate and possibly bias correction can improve the estimation of the second pass, alleviating its EIV problem. Traditionally, the two-pass procedure is performed by grouping the stock in portfolios to avoid the EIV in the second pass. Grouping stocks in portfolios has some shortcomings. See [40] for a discussion. Since our approach improves the estimation in the first step, we consider the two-pass procedure for stocks instead of for portfolios. We also consider, however, the analysis for portfolios to allow some comparison with previous literature.

Figure 4 plots the results of the second-pass regression using the betas from 2SLS and from BHS. In this figure, the horizontal axis measures the realized return and the vertical axis measures the fitted return obtained from the regression (18). If the CAPM fitted the data perfectly, all points should lie exactly in the 45-degree line. The CAPM, however, has been challenging in explaining the cross-section return. In part (a) of Figure 4, we observe that the return of the market portfolio can explain little of the cross-section variation of stock returns when we use the betas from 2SLS ( $R^2 = 0\%$ ). Using the betas from the BHS, in part (b) of Figure 4, we can observe an increase in the explanation of cross-section return, with an  $R^2 = 3\%$ . This power of explanation is still low compared with alternative models in the literature, such as multifactor models, but it represents some improvement in the explanation of heterogeneity of cross-section return by the market factor.



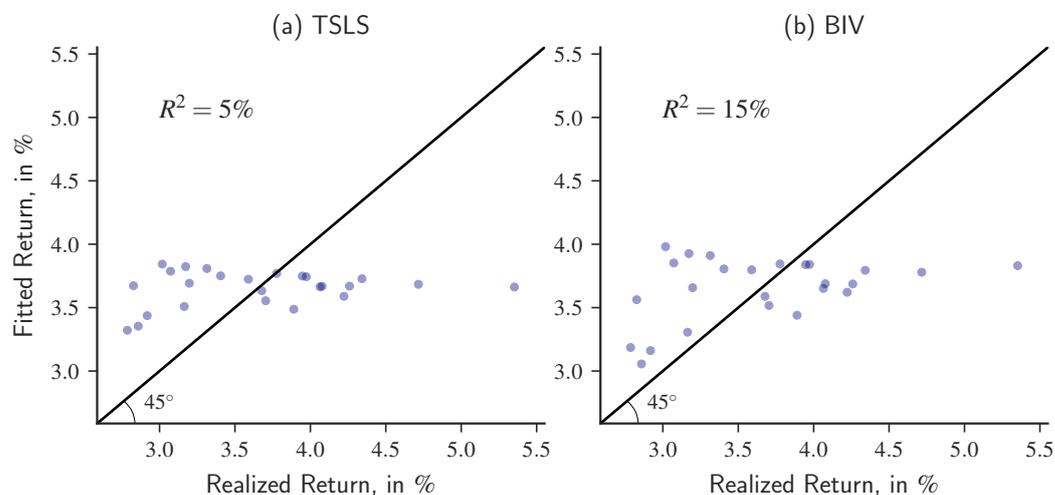
**Figure 4.** Realized versus fitted stocks return: (a) fitted return obtained from 2SLS betas estimates; and (b) fitted return obtained from BHS.

We can interpret this result as a “possible improvement in the measure”. Using many instruments and regularization techniques, we can deal directly with the EIV of the return on the market portfolio, alleviating the problems caused by the EIV. In a certain sense, our approach deals with the critique of [3]: even though we cannot measure the return on the market portfolio exactly, we can use a data-rich environment to treat the mismeasured variable. The shrinkage approach makes it possible to downplay the significance of unimportant instruments and use all the relevant information in the data. This better use of the data allows for a better estimate of the betas in the first step; hence, the second step can explain a higher variation of the cross-section return.

To illustrate the argument above, we highlight in Figure 4 the points representing the AMD and Nvidia stocks. Notice that these two stocks present the highest realized return among all stocks. They are, however, distant from the 45-degree line. For a given estimate of  $\lambda_0$  and  $\lambda_1 > 0$ , a higher Nvidia or AMD beta would increase its fitted return, bringing the points of these stocks closer to the 45-degree line. We note, however, that different values of  $\beta_i$  would affect the estimates of  $\lambda_j$ ,  $j = 0, 1$ . We also note that, in our case, the estimated  $\lambda_0$  and  $\lambda_1$  are positive, both for betas from 2SLS and betas from BHS, with the latter being greater than the former. As shown in Figure 3, the point BHS estimate of the CAPM beta is greater than the 2SLS, which brings the points closer to the 45-degree line and increases the  $R^2$ .

We also analyze the two-pass procedure using portfolios instead of stocks. Specifically, we consider the 25 Fama–French portfolios, sorted by size and book-to-market, to run the two-step procedure of the [26], as in the case of the stocks presented above. The portfolio data range from 1963-Q3 to 2019-Q4. Following [10,39], we use the data in quarterly frequency. Figure 5 presents the fitted return versus the realized return of the 25 portfolios.

Comparing the results using the betas from 2SLS in part (a) with the results using BHS in part (b) of Figure 5, we again observe an increase in the  $R^2$ . That is, using the betas from the BHS increases the percentage of variation in the portfolio returns explained by the market beta.



**Figure 5.** Realized versus fitted portfolio return: (a) fitted return obtained from 2SLS betas estimates; and (b) fitted return obtained from BHS.

A large body of literature has documented the CAPM's incapability to explain the heterogeneity of cross-section average return [41,42]. By using methods that ignore the EIV problems, such as OLS, or instrumental variables without regularization techniques and high-dimensional settings, previous literature has reported that the CAPM can explain around only 1% of the variation in the cross-section return of 25 Fama–French portfolios [10,39]. In contrast, by exploring the high-dimensional settings in instrumental variables, our approach shows that the CAPM betas can explain 15% of the variation in the return of 25 Fama–French portfolios.

These findings provide fresh insight into the estimate of asset pricing models with measurement errors. We can lessen the error-in-variables problem in the assessment of systematic risk, which in turn will better explain the cross-section of returns, by employing high-dimensional data and appropriate approaches. Although the power of explanation of the cross-section of returns reported in our empirical application is modest compared with multifactor models, we emphasize that the inclusion of several instruments and regularization priors may increase this power relative to unregularized techniques. Therefore, using a variety of instruments with regularization priors can be beneficial in other contexts as well, particularly in asset pricing models that incorporate variables with significant measurement errors, such as macroeconomic variables.

## 5. Discussion

This study enhances CAPM estimation by introducing Bayesian regularization for a large number of instrumental variables in the presence of measurement errors. Results highlight significant improvements in estimating CAPM beta values using shrinkage priors for instrument coefficients. Simulation and empirical tests provide supportive evidence for our approach. In a Monte Carlo simulation analysis, the results show a remarkable reduction in mean bias when utilizing the regularized Bayesian instrumental variable approach, particularly in contrast to traditional 2SLS estimation. Furthermore, the introduced regularization technique results in smaller bias and root mean squared errors, indicative of its effectiveness in high-dimensional settings to address the error-in-variables challenge within the CAPM. The empirical application reinforces the significance of the approach. Notably, the estimated CAPM beta values exhibit subtle deviations compared with OLS and 2SLS estimates, with variations discernible across different assets. This distinction in

estimated betas holds substantial economic relevance, especially considering the sensitivity of numerous financial models to beta values.

Expanding on the empirical context, the study delves into the second step of the Fama–Macbeth procedure, focusing on both individual stocks and portfolios sorted by size and book-to-market ratios. The results highlight the pronounced explanatory power of the proposed approach. In the case of portfolio data, the introduced methodology effectively accounts for approximately 15% of the cross-sectional variation in portfolio returns, a significant improvement over the standard 2SLS method, which explains only 5%. Similarly, for individual stocks, the CAPM beta estimates derived from the proposed approach contribute to explaining 3% of the cross-section of stock returns, a notable advance from the nearly 0% explanatory power offered by the standard 2SLS method.

The key element in our findings is to enhance the information set available for estimation. Asset pricing models are complex due to measurement errors in their components. However, the abundant data environment offers a solution by introducing more information into the model. Bayesian regularization instrumental variables serve as the mechanism for unveiling this relevant information, mitigating the impact of mismeasurement.

An important comment about the contribution of our work is its general application in more general contexts than CAPM estimation. Note that the same estimation methodology can be applied in the context of multifactor risk models, especially in the construction of risk factors based on asset sorting using characteristics such as the so-called Fama–French factors [42]. These factors are built by ordering the assets based on observable characteristics, and then building the risk factors based on quantile separation. Note that in this procedure there are several possible sources of measurement risk, such as the frequency of reordering based on the observed characteristic, the choice of quantile level for separating the portfolios, etc. These ad hoc choices can introduce a component of measurement error in the construction of these risk factors, as discussed, for example, in [40,43].

## 6. Conclusions

Our research offers a significant advancement in the estimation of asset pricing models that suffer from the EIV problem. We present a novel approach that addresses the EIV issue within the capital asset pricing model (CAPM) by utilizing a large set of instruments and incorporating shrinkage priors. This method overcomes the challenge of weak instrument correlation with the endogenous market return, providing more robust estimates and mitigating potential bias. Thus, by using our approach, one can find a better estimate for the systematic risk of an asset, measured by the betas, and also better explain the cross-section variation in expected return. Our research not only addresses CAPM estimation but also opens up avenues for improving risk pricing methodologies in more complex multifactor contexts by considering and mitigating the impact of measurement errors.

These conclusions contribute to the literature that works on the solution of the critique of [3], in particular the strand of literature that uses the instrumental variable. The data-rich environment available in finance offers a wide range of instruments to solve the EIV present in the capital asset pricing model. These instruments, however, are usually weakly correlated with the endogenous market return, and using too many instruments may induce bias. Our approach mitigates this induced bias by shrinking unimportant instruments toward zero, which allows using all the available information efficiently.

Future research can extend the results of this work in at least three ways. First, the natural extension of our findings lies in the realm of risk pricing for multifactor models, where the presence of contamination by measurement errors can lead to biased estimates of risk premia. The second way is to generalize the model presented here to consider the joint estimation of all beta assets. This generalization can be a way to increase the efficiency of the estimates since it will consider the covariance of all stock returns in the estimation. Finally, a limited information Bayesian approach may be designed to also include prior regularization, in an analogous manner, as presented here.

**Author Contributions:** Conceptualization, C.R.d.A.A. and M.L.; Methodology, C.R.d.A.A. and M.L.; Investigation, C.R.d.A.A. and M.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by CNPq (310646/2021-9), FAPESP (2018/04654-9) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) Finance Code 001.

**Data Availability Statement:** Data from Kenneth French website <https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/> and Yahoo Finance.

**Acknowledgments:** The authors acknowledge the comments of four anonymous referees and the editors Chiara Di Maria and Antonino Abbruzzo. We appreciate the financial support from Capes, CNPq (310646/2021-9) and FAPESP (2018/04654-9).

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A

### Appendix A.1. Full Conditional Posterior

First full conditional posterior  $\sigma_x^2 | \delta, \beta, \alpha, \xi^2$

From Equation (5) we know that

$$f(x_t | z_t, \delta, \sigma_x^2) = \mathcal{N}(z_t' \delta, \sigma_x^2),$$

which represents the likelihood. It implies that the density of  $x \equiv (x_1, \dots, x_n)'$ , given  $z \equiv (z_1, \dots, z_n)'$ , is:

$$f(x | z, \delta, \sigma_x^2) \propto (\sigma_x^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma_x^2} (x - z' \delta)' (x - z' \delta) \right\}.$$

Consider the inverse gamma prior with an  $s_x/2$  scale parameter and  $k_x/2$ , it follows that the conditional posterior is:

$$\pi(\sigma_x | x, y, Z, \delta) \propto (\sigma_x^2)^{-\frac{k_x + n - 2}{2}} \exp \left\{ -\frac{1}{2\sigma_x^2} \left[ s_x + \sum_{i=1}^n (x_i - z_i' \delta)^2 \right] \right\},$$

which is the kernel of an inverse gamma.

Second full conditional posterior  $\delta | \sigma_x^2, \alpha, \beta, \xi^2$

Use the elliptical slice sampler for this parameter, described in Algorithm 1.

Third full conditional posterior  $(\gamma, \beta, \alpha, \xi^2) | \delta, \sigma_x^2, \alpha, \beta, \xi^2$

To simplify the notation, define  $\theta \equiv (\gamma, \beta, \alpha)'$ , which is a  $(3 \times 1)$  vector. From Equations (5) and (6) and using the bivariate normal properties, we can find the conditional distribution  $\varepsilon_{y_t} | \varepsilon_{x_t} \sim \mathcal{N}(\alpha(x_t - z_t' \delta), \xi^2)$ , where  $\alpha \equiv \rho \frac{\sigma_y}{\sigma_x}$ ,  $\xi^2 \equiv (1 - \rho^2) \sigma_y^2$  and note that  $\varepsilon_{x_t} = x_t - z_t' \delta$ . Then, from Equation (6), we can conclude that  $y_t | x_t \sim \mathcal{N}(\gamma + x_t \beta + \alpha(x_t - z_t' \delta), \xi^2)$ .

Define  $\tilde{x}_t \equiv (1, x_t, x_t - z_t' \delta)$ , which is a  $(1 \times 3)$ . It will be useful to consider  $n \times 3$  matrix of observation:  $\tilde{x} = (\tilde{x}_1', \dots, \tilde{x}_n')'$ . Thus, we can write  $y_t | x_t \sim \mathcal{N}(\tilde{x}_t \theta, \xi^2)$ . The conditional likelihood will be:

$$f(y | x, Z, \dots) \propto (\xi^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\xi^2} (y' y - y' \tilde{x} \theta - \theta' \tilde{x}' y - \theta' \tilde{x}' \tilde{x} \theta) \right\}$$

Combining this likelihood with the normal inverse gamma prior:

$$\pi(\theta, \xi^2 | \tilde{x}^2) \sim \mathcal{NIG} \left( 0, \xi^2 \Sigma_0^{-1}, \text{shape} = \frac{k}{2}, \text{scale} = \frac{s}{2} \right)$$

and defining  $a \equiv k + n$  and  $M = \Sigma_0 + \tilde{x}'\tilde{x}$  allows us to find the third full conditional posterior:

$$\pi(\theta, \zeta^2 | y, x, Z, \sigma_x^2, \delta) = \mathcal{NIG}\left(M^{-1}\tilde{x}'y, \zeta^2 M^{-1}, \text{shape} = \frac{a}{2}, \text{scale} = \frac{b}{2}\right)$$

where  $b \equiv s + y'y - y'\tilde{x}M^{-1}\tilde{x}'y$ .

Appendix A.2. Robustness Check

We also consider alternative calibration in the simulation exercise. Besides the results present in the main text, we also consider a context in which the numbers of weak instruments and strong instruments are always equal, that is, half of the instruments are strong, and half are weak. Specifically, we set the standard deviation of half of the assets to  $\sigma_i = 0.01$  and the other half to  $\sigma_i = 0.04$ . We also consider a less intense measurement error, setting the standard deviation of measurement error to  $\sigma_u = 0.9\sigma_x$ . The calibration of the other parameters is the same as described in the main text. Table A1 presents the results for this setting.

**Table A1.** Measures for the beta estimation error,  $n = 1000$ , with an equal number of weak and strong instruments.

		OLS	2SLS	LIML	BHS	BLASSO	BFBS
$p = 2$	Mean bias	-0.449	1.009	<b>0.008</b>	0.022	0.018	0.019
	Mean abs. bias	0.449	2.000	0.230	0.236	<b>0.235</b>	0.236
	RMSE	0.458	28.312	0.288	0.297	<b>0.295</b>	0.296
$p = 10$	Mean bias	-0.452	-0.024	-0.008	-0.011	-0.017	-0.019
	Mean abs. bias	0.452	0.154	0.150	0.149	<b>0.148</b>	0.148
	RMSE	0.462	0.196	0.188	0.187	<b>0.185</b>	0.185
$p = 20$	Mean bias	-0.445	-0.013	0.009	<b>0.001</b>	-0.011	-0.014
	Mean abs. bias	0.445	0.124	0.125	0.123	<b>0.120</b>	0.120
	RMSE	0.455	0.158	0.158	0.156	0.154	<b>0.154</b>
$p = 40$	Mean bias	-0.449	-0.044	-0.006	-0.022	-0.044	-0.053
	Mean abs. bias	0.449	0.121	0.118	<b>0.117</b>	0.118	0.120
	RMSE	0.458	0.152	0.151	<b>0.147</b>	0.149	0.150
$p = 80$	Mean bias	-0.445	-0.070	<b>-0.004</b>	-0.042	-0.084	-0.105
	Mean abs. bias	0.445	0.115	0.109	<b>0.106</b>	0.117	0.128
	RMSE	0.454	0.145	0.137	<b>0.134</b>	0.148	0.159
$p = 160$	Mean bias	-0.450	-0.120	<b>0.000</b>	-0.114	-0.179	-0.221
	Mean abs. bias	0.450	0.143	<b>0.113</b>	0.136	0.184	0.222
	RMSE	0.460	0.171	<b>0.142</b>	0.163	0.209	0.244

Note: We report mean bias, mean absolute bias, and root mean squared error (RMSE) of each estimator. We highlight in bold type the best estimator for each criterion and number of instruments.

From Table A1, we can notice that the Bayesian horseshoe approach dominates the 2SLS for all criteria and choices of  $p$ , in line with the results in the main text. Also, except for  $p = 2$ , the horseshoe prior performs better than the other regularization priors in terms of mean bias. In this setting, however, the improvement of the Bayesian horseshoe in relation to the unregularized 2SLS is attenuated as the number of instruments increases.

The results of this configuration show that, in situations where regularity in the construction of the instruments is met, in the sense that the instruments present a greater correlation with the covariate measured with error, the gains from the use of estimators based on shrinkage priors is reduced. These results are expected since in this configuration it is desirable that the shrinkage power be less effective due to the need to maintain a greater number of instruments in the first stage of the estimation. In the limit situation where all instruments are strong, it is expected that all instruments are selected, and, in this situation, the differences between the 2SLS estimators and the Bayesian estimators are reduced.

## Appendix A.3. Additional Bayesian Estimation Results

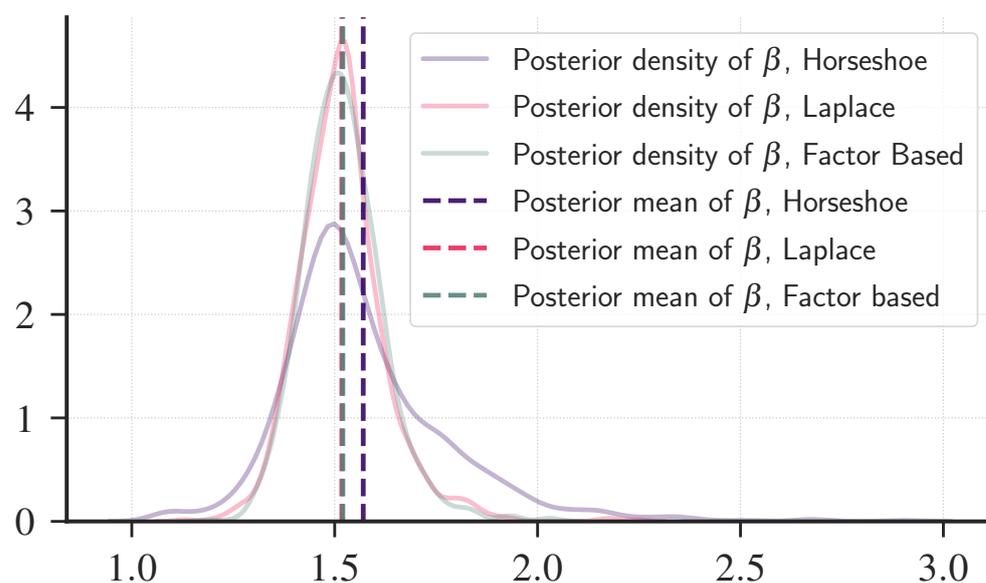


Figure A1. Posterior distributions of beta for the three different types of priors for NVIDIA stock.

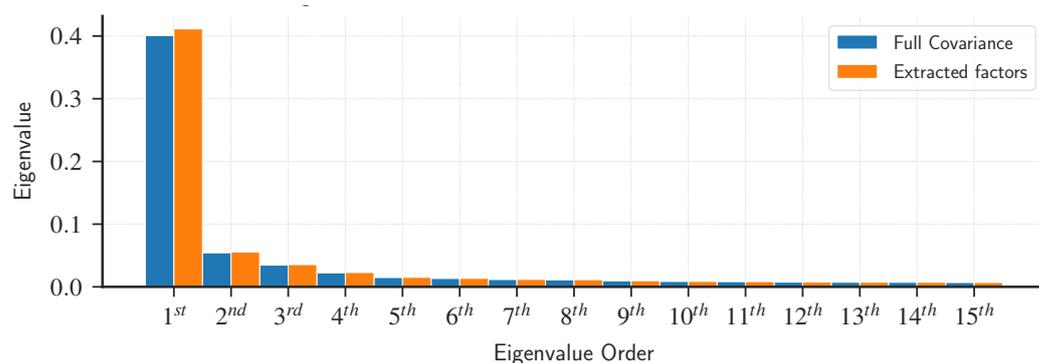


Figure A2. The eigenvalues of a full covariance matrix and the eigenvalues of the trace-minimized covariance matrix.

## References

1. Sharpe, W.F. Capital asset prices: A theory of market equilibrium under conditions of risk. *J. Financ.* **1964**, *19*, 425–442.
2. Lintner, J. The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets. *Rev. Econ. Stat.* **1965**, *47*, 13–37. [[CrossRef](#)]
3. Roll, R. A critique of the asset pricing theory's tests Part I: On past and potential testability of the theory. *J. Financ. Econ.* **1977**, *4*, 129–176. [[CrossRef](#)]
4. Yu, Z.J. Cross-Section of Returns, Predictors Credibility, and Method Issues. *J. Risk Financ. Manag.* **2023**, *16*, 34. [[CrossRef](#)]
5. Korkie, R.; Turtle, H.J. *Limiting Investment Opportunity Sets, Asset Pricing, and the Roll Critique*; Technical Report; Elsevier: Amsterdam, The Netherlands, 2023. [[CrossRef](#)]
6. Malloch, H.; Philip, R.; Satchell, S. Estimation with Errors in Variables via the Characteristic Function. *J. Financ. Econom.* **2021**, *21*, 616–650. [[CrossRef](#)]
7. Cai, Z.; Fang, Y.; Xu, Q. Testing capital asset pricing models using functional-coefficient panel data models with cross-sectional dependence. *J. Econom.* **2022**, *227*, 114–133. [[CrossRef](#)]
8. Vigo-Pereira, C.; Laurini, M. Portfolio Efficiency Tests with Conditioning Information: Comparing GMM and GEL Estimators. *Entropy* **2022**, *24*, 1705. [[CrossRef](#)]
9. Agrawal, P. The Gibbons, Ross, and Shanken Test for Portfolio Efficiency: A Note Based on Its Trigonometric Properties. *Mathematics* **2023**, *11*, 2198. [[CrossRef](#)]
10. Meng, J.G.; Hu, G.; Bai, J. OLIVE: A simple method for estimating betas when factors are measured with error. *J. Financ. Res.* **2011**, *34*, 27–60. [[CrossRef](#)]

11. Simmet, A.; Pohlmeier, W. The CAPM with Measurement Error: ‘There’s life in the old dog yet!’. *Jahrbücher Für Natl. Stat.* **2020**, *240*, 417–453. [[CrossRef](#)]
12. Nagel, S. Machine learning in asset pricing. In *Machine Learning in Asset Pricing*; Princeton University Press: Princeton, NJ, USA, 2021.
13. Hotz-Behofsits, C.; Huber, F.; Zörner, T.O. Predicting crypto-currencies using sparse non-Gaussian state space models. *J. Forecast.* **2018**, *37*, 627–640. [[CrossRef](#)]
14. Kowal, D.R.; Matteson, D.S.; Ruppert, D. Dynamic shrinkage processes. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **2019**, *81*, 781–804. [[CrossRef](#)]
15. Kozak, S.; Nagel, S.; Santosh, S. Shrinking the cross-section. *J. Financ. Econ.* **2020**, *135*, 271–292. [[CrossRef](#)]
16. De Nard, G. Oops! I shrunk the sample covariance matrix again: Blockbuster meets shrinkage. *J. Financ. Econom.* **2022**, *20*, 569–611. [[CrossRef](#)]
17. Yu, J. Bayesian Methods in Economics and Finance: Editors Introduction. *J. Econom.* **2022**, *230*, 1–2. [[CrossRef](#)]
18. Feng, G.; He, J. Factor investing: A Bayesian hierarchical approach. *J. Econom.* **2022**, *230*, 183–200. [[CrossRef](#)]
19. Lopes, H.F.; McCulloch, R.E.; Tsay, R.S. Parsimony inducing priors for large scale state-space models. *J. Econom.* **2022**, *230*, 39–61. [[CrossRef](#)]
20. Fisher, M.; Jensen, M.J. Bayesian nonparametric learning of how skill is distributed across the mutual fund industry. *J. Econom.* **2022**, *230*, 131–153. [[CrossRef](#)]
21. Brignone, R.; Gonzato, L.; Lütkebohmert, E. Efficient Quasi-Bayesian Estimation of Affine Option Pricing Models Using Risk-Neutral Cumulants. *J. Bank. Financ.* **2023**, *148*, 106745. [[CrossRef](#)]
22. Bekker, P.A. Alternative approximations to the distributions of instrumental variable estimators. *Econom. J. Econom. Soc.* **1994**, *62*, 657–681. [[CrossRef](#)]
23. Newey, W.K.; Smith, R.J. Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators. *Econometrica* **2004**, *72*, 219–255. [[CrossRef](#)]
24. Ng, S.; Bai, J. Selecting Instrumental Variables in a Data Rich Environment. *J. Time Ser. Econom.* **2009**, *1*, 1–32. [[CrossRef](#)]
25. Hahn, P.R.; He, J.; Lopes, H. Bayesian factor model shrinkage for linear IV regression with many instruments. *J. Bus. Econ. Stat.* **2018**, *36*, 278–287. [[CrossRef](#)]
26. Fama, E.F.; MacBeth, J.D. Risk, Return, and Equilibrium: Empirical Tests. *J. Political Econ.* **1973**, *81*, 607–636. [[CrossRef](#)]
27. Markovitz, H.M. *Portfolio Selection: Efficient Diversification of Investments*; John Wiley and Sons: Hoboken, NJ, USA, 1959.
28. Campbell, J.Y.; Lo, A.W.; MacKinlay, A.C. *The Econometrics of Financial Markets*; Princeton University Press: Princeton, NJ, USA, 1997.
29. Jensen, M.C.; Black, F.; Scholes, M.S. The capital asset pricing model: Some empirical tests. In *Studies in the Theory of Capital Markets*; Praeger Publishers: Westport, CT, USA, 1972.
30. Gibbons, M.R.; Ross, S.A.; Shanken, J. A Test of the Efficiency of a Given Portfolio. *Econometrica* **1989**, *57*, 1121–1152. [[CrossRef](#)]
31. Lopes, H.F.; Polson, N.G. Bayesian instrumental variables: Priors and likelihoods. *Econom. Rev.* **2014**, *33*, 100–121. [[CrossRef](#)]
32. Van Erp, S.; Oberski, D.L.; Mulder, J. Shrinkage priors for Bayesian penalized regression. *J. Math. Psychol.* **2019**, *89*, 31–50. [[CrossRef](#)]
33. Carvalho, C.M.; Polson, N.G.; Scott, J.G. The horseshoe estimator for sparse signals. *Biometrika* **2010**, *97*, 465–480. [[CrossRef](#)]
34. Fuller, W.A. Some properties of a modification of the limited information estimator. *Econometrica* **1977**, *45*, 939–953. [[CrossRef](#)]
35. Hansen, C.; Hausman, J.; Newey, W. Estimation with Many Instrumental Variables. *J. Bus. Econ. Stat.* **2008**, *26*, 398–422. [[CrossRef](#)]
36. Davidson, R.; MacKinnon, J.G. *Estimation and Inference in Econometrics*; Oxford Press: New York, NY, USA, 1993; Volume 63.
37. Castillo, I.; Schmidt-Hieber, J.; Van der Vaart, A. Bayesian linear regression with sparse priors. *Ann. Stat.* **2015**, *43*, 1986–2018. [[CrossRef](#)]
38. Wu, T.; Narisetty, N.N.; Yang, Y. Statistical inference via conditional Bayesian posteriors in high-dimensional linear regression. *Electron. J. Stat.* **2023**, *17*, 769–797. [[CrossRef](#)]
39. Lettau, M.; Ludvigson, S. Resurrecting the (C) CAPM: A cross-sectional test when risk premia are time-varying. *J. Political Econ.* **2001**, *109*, 1238–1287. [[CrossRef](#)]
40. Jegadeesh, N.; Noh, J.; Pukthuanthong, K.; Roll, R.; Wang, J. Empirical tests of asset pricing models with individual assets: Resolving the errors-in-variables bias in risk premium estimation. *J. Financ. Econ.* **2019**, *133*, 273–298. . [[CrossRef](#)]
41. Fama, E.F.; French, K.R. The cross-section of expected stock returns. *J. Financ.* **1992**, *47*, 427–465. [[CrossRef](#)]
42. Fama, E.F.; French, K.R. Common risk factors in the returns on stocks and bonds. *J. Financ. Econ.* **1993**, *33*, 3–56. [[CrossRef](#)]
43. Giglio, S.; Xiu, D. Asset Pricing with Omitted Factors. *J. Political Econ.* **2021**, *129*, 1947–1990. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.