Estimating the capital asset pricing model with many instruments: A Bayesian Shrinkage approach

Cássio Roberto de Andrade Alves

Márcio Poletti Laurini *

Abstract

We propose a Bayesian estimation of the capital asset price model (CAPM) with measurement errors using a large set of instruments and shrinkage priors over the parameters associated with the instrumental variables. When using the instrumental variable approach to estimate the CAPM, the challenge is to find "strong" instruments for the market portfolio return, even though the data-rich environment available in finance allows us to use many-instrument (possibly weak) settings. We use regularization priors to deal with the large set of instrumental and improve inference of the estimated beta. Using simulated data, we find that our approach may reduce the size of mean bias caused by error-in-variables up to 90% concerning the traditional two-stage least square. This reduction, however, is attenuated as the number of instruments increases. In an empirical application, our findings show that the estimated CAPM stock's beta using the proposed approach is greater than the ordinary least square estimator, and the traditional two-stage least square converges to the admittedly biased least squares estimator. This difference in estimated betas is economically relevant since many financial models are sensitive to beta.

Keywords: Bayesian Estimation, Shrinkage, Instrumental Variables, CAPM

*emails - cassioalves@usp.br and laurini@fearp.usp.br. The authors acknowledge financial support from CNPq (306023/2018-0) and FAPESP (2018/04654-9). This study was also financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) - Finance Code 001.

1 Introduction

A well-established fact in financial econometrics is that the estimation of the capital asset price model (CAPM) requires a surrogate for the market return since the true market return is unobservable (Roll, 1977; Stambaugh, 1982; Prono, 2015; Simmet and Pohlmeier, 2020). The use of such a surrogate introduces an error-in-variables (EIV) that bias the estimates and turn the interpretation of a rejection of the model ambiguous (Roll, 1977). In this model, the traditional instrumental variable solution to the EIV problem is challenging. Although there are many instruments available for market return, these instruments are weakly correlated with market return, which makes it difficult to elicitate instruments. This paper proposes a Bayesian approach to estimate the capital asset price model using a large set of instruments and shrinkage priors over the parameters associated with the instruments.

Ignoring the measurement error present in the CAPM has drastic consequences for evaluating the model empirically. The biased estimates may mismeasure the exposition to the systematic risk of an asset, and it also has consequences to test the validity of the model. If a test indicates the rejection of CAPM, then there are alternative interpretations: Either the model is false, or the surrogate of the market return is unappropriated. This point was extensively discussed in Roll (1977) and is commonly known as Roll's (1977) critique.

The literature tries to solve the Roll's (1977) critique in several ways (Stambaugh, 1982; Shanken, 1987b,a; Jagannathan and Wang, 1996; Jegadeesh et al., 2019), and a particular one is to use instrumental variable estimation (Coën and Éric Racicot, 2007; Racicot et al., 2019). This approach requires finding instruments, which are variables that are correlated with the market return but uncorrelated with the error term of the single-factor CAPM equation. On the one hand, the martingale property of market return and the difficulty in predicting it limit the IV approach (Simmet and Pohlmeier, 2020). On the other hand, the data-rich environment available in finance allows us to use a large instrument set, even if these instruments are weakly related to the market return. We can use, for instance, the assets return as instruments. Since the market return results from all asset returns, each of these assets (and possibly function of them) can explain the market return to some degree, for instance, which form a large set of candidates to instruments.

Using many instruments may itself be a source of bias to the estimates (Bekker, 1994; Newey and Smith, 2004; Ng and Bai, 2009). In terms of the two-step generalized method of moments, for example, it would imply a large number of moment conditions, which consequently produce biased estimates (Newey and Windmeijer, 2009). However, combining regularization methods with instrumental variable estimation can overcome this drawback. There are several approaches to deal with this regularization, ranging from factor analysis, as in Bai and Ng (2010), to methods that explores the covariance of instruments (Carrasco, 2012). An alternative approach is a Bayesian estimation that uses shrinkage priors over the parameters of the instrument. In particular, Hahn et al. (2018) proposed a factor-motivated prior structure to deal with many-instruments setting. In this context, the Bayesian approach is interesting because the first and second stages are jointly estimated, allowing a mutual influence. Thus, the Hahn et al.'s (2018) proposal to combine a many-instrument setting with a large set of the instrument is attractive.

In asset price models, this combination is an unexplored field. In this paper, we propose to estimate the capital asset price model using a large set of instruments and shrinkage priors over the parameters associated with the instruments. To do so, we use the Bayesian approach proposed by Hahn et al. (2018) to shrink unimportant instruments and compare the size of the estimated bias with that produced by the traditional estimation methods (ordinary least square and two-stage least square).

Our results showed that the Bayesian regularization priors over the instruments coefficient may improve the inference about CAPM beta. In the Monte Carlo simulation analysis, we find that size of mean bias is reduced by 90% in relation to the traditional TSLS, but this improvement is reduced as the set of instruments is enlarged. In an empirical application, we find that the estimated CAPM stock's beta is greater than the ordinary least square (OLS) estimator, while the traditional two-stage least square converges to the admittedly downward biased OLS estimator. The discrepancy between our estimated beta from TSLS and OLS estimates is economically important since many financial models are sensitive to beta.

2 The CAPM and measurement errors

The seminal paper of Markovitz (1959) has prepared the framework for the *Capital Asset Price Model* (CAPM). 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. Sharpe (1964) and Lintner (1965) worked on Markovitz (1959) results to analyze the implication for the asset price and developed what is called the Sharpe-Lintner CAPM, or just CAPM.

By assuming that investors possess homogeneous expectations, meaning that they agree about expectations of future investments, Sharpe (1964) and Lintner (1965) shown 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 state, and commodities, which makes it an unobserved variable. In practice, usual surrogates for the market portfolio are market indexes, such as SP&500, 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 will imply a relation between assets risk-premium and the market risk premium.

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

where $\beta_i \equiv \sigma_{im}/\sigma_m^2$. Therefore, the CAPM summarized in equation (1) is an equilibrium result that holds for a single period.

The relation establish in equation (1) for one period is not enough to empirically assess the CAPM. To proceed with econometric analysis, an additional assumption is required: the returns are independent and identically distributed along time and multivariate normal. Although this hypothesis is a strong one, its posses some benefits. First, it is consistence with the CAPM hold for each period in time. Moreover, it is a good approximation for monthly returns (Campbell et al., 1997). Under this assumption, the CAPM may be represented by the single index model, which is described by

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

In equation (2), if γ_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 tradition of testing the model that became known as the time series approach. To empirically test the CAPM model Jensen et al. (1972) proposed to use time series for return of assets, for return risk-free asset and a proxy to the return of market portfolio to estimate equation (2). Usual choice for the risk-free asset is the US treasure bill and SP&500 for the return of the market portfolio. Then, their approach suggests testing whether the estimated intercept is equal to zero, which may be done by using a Wald test or the test proposed by Gibbons et al. (1989).

Testing the CAPM using the Jensen et al. (1972) approach is problematic once the return of market portfolio, R_{mt}, is a variable observed only with errors. The source of measurement errors appears because the market indexes used to estimate the model contain only a sub set of assets. Moreover, even if all universe of assets were observed, the measurement error could appear due to misspecification in weights of assets. This problem is known as Roll's critique, due to Roll (1977) who argued that, once the market portfolio is not observed, the CAPM cannot be tested. According to this author, a rejection of the CAPM could be due to measurement errors in return of 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 the classical measurement errors, let $\tilde{R}_{mt}$ denote the observed return of market portfolio. Also, denote by x_t^* the excess of return on the true market portfolio and by x_t the excess of return on the observed market portfolio. The excess of return on the asset i is denoted by y_{it}, and there is no error in variable in this case. Instead of equation (2), the model to be estimated to test the CAPM should be

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

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

Equation (4) assumes that the error in excess of return on market portfolio is additive. If one ignores this additive measurement error and estimate 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 CAPM holds. Thus, to appropriately deal with error-in-variable problem, the equations (3) and (4) must be considered to estimate and test CAPM model.

3 Literature Review

This section aims to relate previous research with our paper, showing a lack of knowledge still unexplored. Two pieces of literature are associated with our research. First, the literature that estimates and test the CAPM model, considering measurement errors. In general, our paper is related to the time series approach proposed by Jensen et al. (1972), which is also known as one-pass regression. The two-pass regression proposed by Fama and MacBeth (1973) is another form to estimate the CAPM and also is subject to the error-in-variables problem. In this sense, research that use the two-pass regression approach and treat the underlying measurement error is also related to our work to some degree. However, we focus on one-pass regression. The second branch of research that is relevant is the econometric literature on measurement errors, which consider new alternatives to solve the problem, such as the selection or shrinkage of instruments in the context of many instruments.

The measurement error in CAPM model and its consequence for estimating and testing the model has being know since the influential work by Roll (1977). This author argues that, given the latency of the market return, we cannot test the theory of CAPM, since alternative interpretations for rejection may arise: (i) the theory of CAPM is false; (ii) the surrogate of the market return is not adequate, or (iii) both (i) and (ii). Although the Jensen et al. (1972) approach does not deal with the latency of market return, together with the Roll's (1977) criticism, it opened a new agenda of research. In one hand, the one-pass regression approach needed to be generalized to jointly test the hypothesis of no intercept and develop a framework to estimate the model for each asset instead of grouped securities. The generalization to jointly test the hypothesis of $\gamma_i = 0$ for all i was subsequently introduced by Gibbons et al. (1989). In the other hand, the tradition of using a portfolio instead of individual assets has remained unchanged. Moreover, the measurement error, introduced by the proxy of the market return, requires suitable econometric methods.

After Roll's (1977) critique, several researchers tried to deal with measurement errors, using both Bayesian or Frequentist methods. A first branch of research worked on the improvement of the measure of market portfolio, by considering alternatives proxies for the market return, including, for instance, durable goods and human capital (Stambaugh, 1982; Jagannathan and Wang, 1996; Campbell et al., 1997). Another branch of the literature used multiple proxies and generalized the problem in a Bayesian perspective, allowing for the inclusion of prior information about the correlation between true market portfolio and the market index used as a proxy Shanken (1987b,a); Harvey and Zhou (1990); Kandel et al. (1995). Although the results of this line of inquiry has important implications for testing CAPM, the estimates of betas remains unsolved. Apart from use the betas for testing the model, such as in the two-pass approach, the estimated betas is useful for practitioners in finance.

Focused on the estimating the CAPM beta, the econometric literature appealed to instrumental variables (IV) estimation. Frequentist examples of IV approach are Coën and Éric Racicot (2007); Meng et al. (2011); Racicot et al. (2019); Jegadeesh et al. (2019) The challenge of this approach is to find an instrument that is sufficiently correlated to the proxy, once this variable presents martingale properties. The traditional instrument used in the literature, which is the endogenous variable lagged, cannot explain the endogenous variable. The alternative that the literature proposes is to use alternative instruments, such as technical instruments. Coën and Éric Racicot (2007) and Racicot et al. (2019), for instance, use higher moments to correct the measurement error. Their results show that if one does not correct for the measurement error, then there will be differences in inference about the theory validity, showing the importance of considering measurement error when estimate CAPM. It should be noted that using higher moments instruments can lead to serious problems in the presence of outliers.

The difficulty to find instruments for the market portfolio relates the CAPM test with another literature, which is the strand of research that studies the selection or shrinkage of instruments when a data-rich environment is available Ng and Bai (2009). In general, when dealing with financial data we have access to relatively large information data, which can form a large set of instrument candidates in the error-in-variable estimation. In this sense, the literature on the selection of instruments is related to our paper. In particular, there is a branch of research that treats instrumental variable selection in a Bayesian approach (Hahn et al., 2018), which can connect the two pieces of literature exposed above. Since it is difficult to find good instruments in the context of CAPM estimation, one can consider all instruments available (non-technical, such as lagged variables, and technical, such as polynomials) and use the econometric method to deal with shrink the coefficients of less important instruments or selection of the best ones. Therefore, the literature of Bayesian instrumental variable estimation is also related to our proposal.

To our knowledge, there is no research that uses shrinkage methods in the instrumental variable regression of CAPM. This is exactly the gap of knowledge that our paper aims to fill. This sort of method may help to better explain the endogenous variable and therefore produce a more realistic estimation of the CAPM model, by correcting the problems caused by the measurement errors.

4 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 long 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 LASSO, ridge, Elastic Net, or via Bayesian shrinkage priors, which penalize the number of covariates in some fashion.

In the instrumental variable regression, it 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 Hahn et al. (2018) has the advantage of linear combine the information in all possibly weak instruments in such a way that, taken together, makes them stronger. In the next subsection, we present this sort of shrinkage prior to the IV regression context.

4.1 Bayesian regularization methods in IV regression

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

$$x_t = z_t'\delta + \varepsilon_{x_t}, \tag{5}$$

$$y_t = \gamma + x_t\beta + \varepsilon_{y_t}, \quad t \in \{1, \cdots, n\} \tag{6}$$

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

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

Since p may be large (possibly larger than the number of observations, $p >> n$), some regularization on equation (5) is necessary. The Bayesian solution to this problem is to impose shrinkage priors on δ to shrink those parameters which have little power to explain x_t. By imposing such a prior, the usual Gibbs sampler scheme (Lopes and Polson, 2014) used to estimate model (5)-(6) cannot be employed. Hahn et al. (2018) developed an elliptical slice sampler that can be dealt with arbitrary prior on δ, allowing us to use 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 arbitrary prior distribution on δ. To understand the Bayesian estimation of IV regression, consider the reduced form of equations (5) and (6):

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

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

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

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

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

$$f(x, y|z, \Theta) = f(y|x, \delta, \alpha, \beta, \xi^2) \times f(x|z, \sigma_x^2\delta) \tag{9}$$

$$= N\left(\gamma + x_t\beta + \alpha(x_t - z_t'\delta), \xi^2\right) \times \mathcal{N}(z_t'\delta, \sigma_x^2). \tag{10}$$

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

$$\delta \sim \text{arbitrary}, \tag{11}$$

$$\sigma_x^2 \sim \mathcal{IG}(\text{shape} = k_x, \text{scale} = s_x), \tag{12}$$

$$(\xi^2, \gamma, \beta, \alpha)' \sim \mathcal{NIG}\left(0, \xi^2\Sigma_0^{-1}, \text{shape} = \frac{k}{2}, \text{scale} = \frac{s}{2}\right). \tag{13}$$

Combining these priors with likelihood function in equation (10), give us the posterior distribution. To sample from this posteriori distribution, it is possible to break it in three full conditional posteriors to form a Gibbs sampler scheme. To explain these three blocks, it is useful to introduce some definitions. Define[1] $\tilde{x} \equiv (1, x, z'\delta)$, $M \equiv \Sigma_0 + \tilde{x}'\tilde{x}$,

[1] In general, the notation with no subscript t means that the variable contains all observation. For example, $x \equiv (x_1, x_2, ..., x_n)$. A particular case is the vector of instruments, z_t, for which all observations is denoted by Z, since it becomes a matrix of dimension $(p \times n)$.

$a \equiv k+n$ 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 Θ, from Equations (10) and (11) the conditional posterior is proportional to $f(x|\Theta)f(y,|x,\Theta)\pi(\delta)$. Since we are considering an arbitrary prior for δ, this full conditional posterior may not have closed form, requiring alternatives methods to sample it. Although traditional Metropolis-Hastings can be used in this case, it scales poorly due to possibly high dimension and multimodality of the full conditional posterior. Instead, Hahn et al. (2018) proposed to sample it using an elliptical slice sampler, which only requires the ability to evaluate $\pi(\delta)$. This algorithm is described below:

Algorithm 1 Elliptical Slice Sampler

```
procedure SliceSampler(δ, σ_x², x, Z, y)
    Define δ̂ = (ZZ')⁻¹Zx and Δ = δ − δ̂
    Draw ζ ~ N(0, σ_x(ZZ')⁻¹) and v ~ U(0,1)
    Compute ℓ ≡ log(f(y|x,Z,δ) + log(π(δ)) + log(v)
    Draw an angle φ ~ U(0,2π), and do lower ← φ − 2π and upper ← φ
    Update Δ and δ: Δ̄ = Δcos(φ) + ζ sin(φ) and δ̄ = δ̂ + Δ
    while log(f(y,x,Z,δ̄)) + log(π(δ̄)) < ℓ do
        If φ < 0, then lower ← φ. Else upper ← φ
        Draw a new angle φ ~ U(lower, upper)
        Update Δ and δ: Δ̄ = Δcos(φ) + ζ sin(φ) and δ̄ = δ̂ + Δ
    δ ← δ̂ + Δ̄.
    return δ
```

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

Full conditional posterior for $\sigma_x^2|\Theta$, data: Fortunately, for an inverse-gamma prior on σ_x^2 the full conditional posterior for σ_x^2 have a closed form. Combining the likelihood $f(x|Z,\delta,\sigma_x^2)$ with the prior given in (12), it is possible to show that the full conditional posterior is an Inverse-Gamma with shape parameter k_x+n and scale $s_x+\sum_{i=1}^{n}(x_t-z_t'\delta)^2$ (see derivation in appendix).

Full conditional posterior for $(\xi^2,\gamma,\beta,\alpha)'|\Theta$, data: This block also have a closed form. By using bivariate normal properties, we can write the likelihood in terms of the transformed variable $\tilde{x}$, which follows $y_t|\tilde{x}_t \sim N(\tilde{x}_t\theta, \xi^2)$. Combining this likelihood with the prior give in (13), it can be show that the full conditional posterior $(\xi^2, \gamma, \beta, \alpha)'|\Theta$, data is a Normal-Inverse-Gamma distribution. Specifically:

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

(See appendix for derivation).

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 δ, and it still works well. In particular, we can elicit several shrinkage prior over δ, since the many instruments setting requires regularization.

4.2 Shrinkage priors for instruments coefficients

There is a large range of shrinkage priors in the literature Van Erp et al. (2019). 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 follow, we present some of these priors that can directly applied to $\delta = (\delta_1, \cdots, \delta_p)$. Then, we proceed with the factor-based prior distribution.

4.2.1 Heavy-tailed priors

Popular choices of shrinkage prior are the Cauchy, Laplace, and Horseshoe densities (Carvalho et al., 2010). The horseshoe density is the stronger one, in the sense that it concentrates 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 much density around zero, and so it also works as a shrinkage prior. The left panel of Figure 1 depicts these three priors for a single coefficient.

An advantage of the Horseshoe prior is that at the same time it concentrates 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 prior 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 δ_j and assume

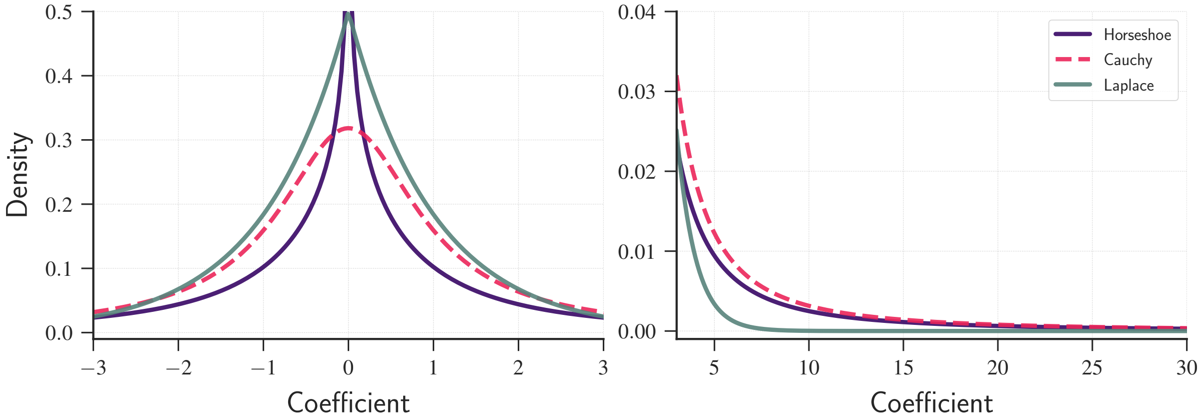

Figure 1: Three sort of shrinkage priors: Horseshoe, Double-exponential and Cauchy, all of them centered at zero.

that they are independent for all $i \neq j$, with $i, j \in \{1, \cdots, p\}$. 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.

4.2.2 Factor-based shrinkage prior

The idea underlying the factor-based prior, proposed by Hahn et al. (2018), 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:

$$\mathrm{Cov}(z_t) = BB' + \Psi^2, \tag{14}$$

where, B is a $(p \times k)$ matrix and Ψ^2 is a diagonal $(p \times p)$. Despite the fact that every covariance matrix admits this decomposition, the interest here is on the case where $k << p$, 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 normal jointly distributed as follow:

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

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:

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

where θ is a $(1 \times k)$. From equation (15) and the definition of $\hat{f}_t$ it is possible to show that $\delta' = \theta A$. However, this specification is only correct if δ 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 δ lies in the row space of A. For that end, the specification in equation (15) needs to be modified

$$x_t = \theta \hat{f}_t + \eta \hat{r}_t, + \varepsilon_t \tag{16}$$

where η is a $(1 \times p)$ vector of parameter and $\hat{r}_t \equiv (I_p + A^{+}A)z_t$ and A^{+} denote the pseudo-inverse of A. In this case, it can be show 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 (16) 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 δ by imposing 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 ω. With this identity, conditional on ω, we can impose an horeseshoe prior, for instance, on $\tilde{\delta}$ and it induce a prior on δ. That is:

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

Following Hahn et al. (2018), we assume that $\omega \sim \mathcal{N}(0, I_{k+p})$. Once we know ω, 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 δ by induce a prior on δ via horseshoe prior over $\tilde{\delta}$. Notice that, under this specification, the factor structure derived in this section is taken into account in the prior over δ. In practice, however, the matrices B, Ψ are unknown and, consequently, A and H is also unknown. Instead of estimating it in a Bayesian fashion, we use point estimate of these matrices, which is found by minimizing the trace of $\text{Cov}(z_t) - D$, by choosing D, subject to D be diagonal and positive-definite.

4.3 Data description

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 generate process, we can calculate the error of estimate and compare it with alternatives methodology (for instance, OLS, 2SLS, etc.). Besides, the simulation exercise, we also apply the empirical method in 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 web site[2]. Finally, we consider all assets listed in the SP&500 with data availability in the last fifteen years. All data is daily and ranges from 2007-01-01 to 2021-12-31, resulting in $3,774$ observations.

5 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 square and two-stage least square estimation. Then, we use the observed data presented before to estimate the CAPM using our proposal procedure.

[2]https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/

5.1 Monte Carlo analysis: Simulation procedures

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

$$x_t = x_t^* + u_t, \quad u_t \sim \mathcal{N}(0, \sigma_u^2) \text{ and } x_t^* \sim \mathcal{N}(0, \sigma_x^2), \tag{17}$$

$$y_{it} = \beta_i x_t^* + \varepsilon_t, \tag{18}$$

for $i \in \{1, ..., M\}$ and $t \in \{1, ..., n\}$. In equations (17) to (18) x_t^* is the true market return and it is assumed to be Gaussian with mean zero and variance σ_x^2, u_t is Gaussian measurement error, with mean zero and variance σ_u^2, x_t is the observed market return. The sensitivity to the market return, which is measured by β_i, is assumed to be known, and based on these values and given the error term ε_t, we can construct the assets return. The first and second moment of the error term ε_t are, respectively, $\mathbb{E}[\varepsilon_t] = 0$ and $\mathbb{E}[\varepsilon_t \varepsilon_t] = \sigma_\varepsilon^2$, and we also assume that it is Gaussian.

To simulate the model, we need to calibrate the parameters $(\sigma_u^2, \sigma_x^2, \sigma_\varepsilon^2, \beta_i)$. For the β_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 half of the assets with $\sigma_\varepsilon = 0.03$ and the other half with $\sigma_\varepsilon = 0.04$. These two different values are chosen based on the distribution of the standard deviation of the assets listed in the SP&500, and the two values are intended to create strong and weak instruments. A lower standard deviation creates assets that will be stronger instruments than those with a higher standard deviation. Finally, we calibrate $\sigma_u = 0.9 \times \sigma_x$ to create a situation with a high measurement error but assure the model is identifiable. These parameters calibration allow us to simulate all variables of interest in CAPM.

To evaluate the accuracy of each method, we simulated the model $N_{\text{sim}} = 1,000$ times. At each iteration, we estimated the parameters using six methods. The first one was the traditional Ordinary Least Squares (OLS), regressing y_{it} on x_t, which is known to be inconsistent in the presence of measurement errors. Second, we consider the Two-Stage Least Squares (TSLS), with the set of instruments formed by all asset returns except the regressand in the CAPM equation (see Meng et al. (2011), 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) estimator with the same set of instruments. Although our main interest is to verify whether the Bayesian regularization of the Two-stage 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 (Hansen et al., 2008) [3]. 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 used the Horseshoe, Laplace, and Factor-Based Shrinkage prior distributions over δ. This estimation is referred to as BHS, BLASSO, and BFBS, once the Laplace distribution is known as Bayesian LASSO (Park and Casella, 2008).

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 different criteria: mean bias, mean absolute bias and the root mean of squared error (RMSE). According to these criteria, the lower its value, the better the estimator.

The Monte Carlo results are summarized in Table 1. The Bayesian "Two-stage" procedure with Horseshoe regularization prior (BHS) is better than traditional TSLS for all criteria and choices of p. When the number of instrument is $p = 20$, the BHS approach reduce the size of the mean bias in 90% in relation to the traditional TSLS[4]. In this case, the mean bias of BHS is even smaller than those of LIML, which is known for correct the bias in the presence of many instruments (Hansen et al., 2008). We note that, for the other values of p, the LIML indeed has the smallest mean bias. However, considering the other criteria, which also penalize for variance, the Bayesian approach dominates the LIML for all cases, except for $p = 160$. In all case, the improvement of the Bayesian estimator is attenuated as the number of instruments increases. Thus, we can conclude that the Bayesian regularization in the estimation of CAPM model ameliorate the inference about the betas, both in terms of bias and variance.

[3]The LIML estimator, however, is also known to have no moments. The modification version of LIML, due to Fuller (1977), solves this drawback. Still, the modification introduces an additional parameter that has to be chosen by the econometrician.

[4]By the "size of mean bias" we mean the absoluve value of the mean bias. Thus, in the case with $p = 20$, the size of mean bias reduction is $(|0.001| - |0.013|)/| - 0.013|$, which is approximately -0.92.

Table 1: Different Measures for the beta estimation error, $n = 1000$

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

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

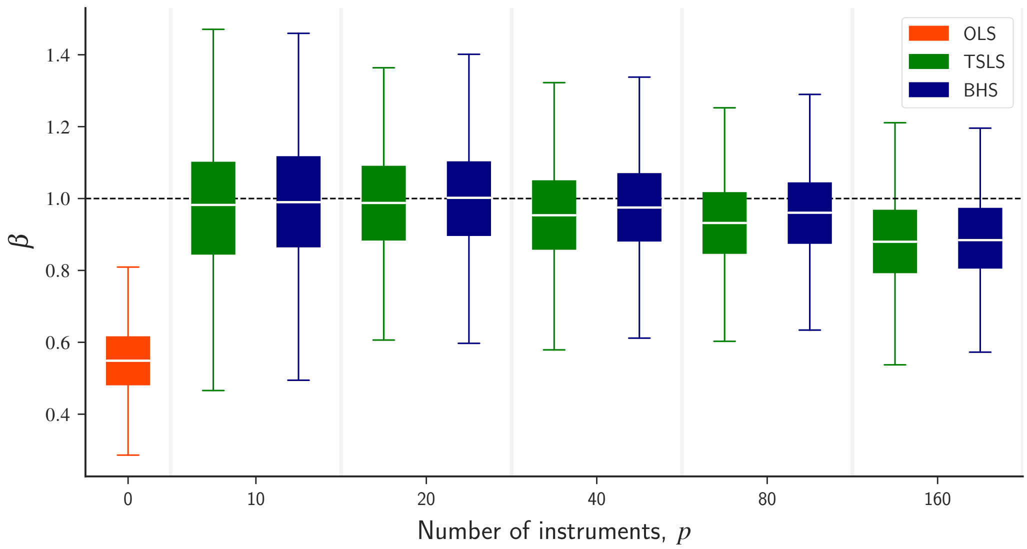

Figure 2: Box plots of the estimated CAPM betas for different numbers of instruments: Comparison between Bayesian Horseshoe (BHS) and Two-Stage Least Square (TSLS). The horizontal black-dashed line represents the true CAPM beta value, which is one

Among the three type of regularization priors, the horseshoe outperforms the others at least in terms of mean bias. For this reason, we now focuses on the horseshoe prior in the comparison with the traditional TSLS. For small sets of instruments, say up to 20, the TSLS bias is small and can be entirely corrected by the Bayesian regularization. When the number of instruments increases to 40 and 80, the bias becomes greater and the regularization alleviate it, but it also shifted down from the true value, which is one. The bias of the traditional TSLS can be explained by the tendency of the OLS, in the first-stage, to fit too well, as the number of instruments increases, see (Davidson et al., 1993, pp. 222). The BHS penalizes the "first-stage" estimation, avoiding this tendency to over-fitting, 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 TSLS and also shows a smaller dispersion, as shown both by Figure 2 and Table 1).

5.2 Empirical Application

This section uses the Bayesian regularization procedure to estimate the CAPM using observed data. We select two stocks based on a preliminary OLS estimation of CAPM beta. We chose the stocks with the lowest and highest beta, and these stocks are General Mills and Lincoln Financial, with betas equal to 0.37 and 1.81. The instruments consist of all stock returns listed in the SP&500 with data available in the last fifteen years. The original set contains 366 instruments, 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 is daily and ranges from January 2007 to December 2021, totaling $3,774$ observations.

The posterior distribution for CAPM beta was similar for the three types of prior (BHS, BLASSO, and BFBS, see Figure 5 in Appendix). This result indicates that, for these stocks, the three 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 drastic from the first to the second eigenvalue and then decline slowly. The covariance of the minimized-trace (the $\text{Cov}(z_t) - D$ discussed in section 4.2.2) presents a similar behavior, see Figure 8 in Appendix. Thus, we cannot isolate "commonalities" and, hence, the prior information is unable to help in the shrinkage of parameters more than a straight horseshoe prior can do.

We also that the Gibbs schemes required many iterations to converge to the stationary distribution. Specifically, it required 10^5 iterations to warm up the Gibbs scheme, and 10^5 iterations saved every 100th to diminish the autocorrelation of the Markov Chain. When comparing the Bayesian IV estimates with OLS and TSLS estimates, there is a difference, especially for the Lincoln Financial stock. Figure 3 presents the posterior distribution for the CAPM estimated beta by BIV-HS, as well as its mean, to compare with OLS and TSLS estimates.

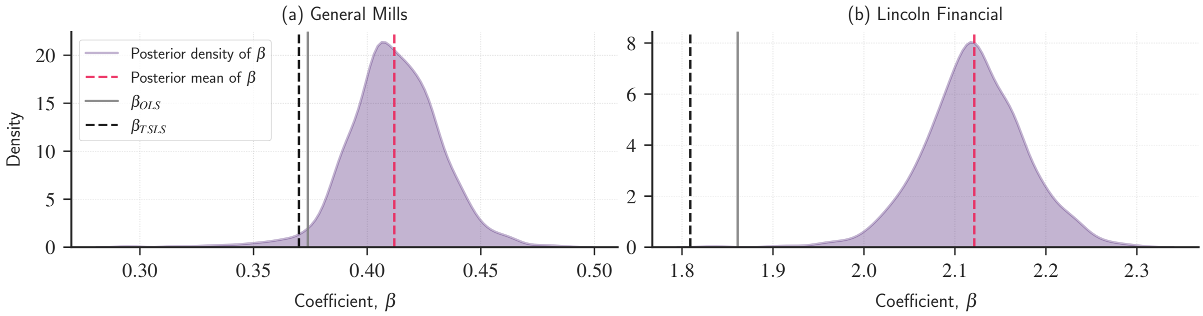

Figure 3: Posterior distribution of CAPM beta estimated by Bayesian Instrumental Variable with Horseshoe prior for three assets: (a) General Mills and (b) Lincoln Financial. The figure also present mean of the posterior distribution and the OLS and TSLS estimates.

For the General Mills stock, the TSLS and OLS estimates are similar. It is an expected result since when the number of the instrument is large (as in this case, $p = 278$) relative to the number of observations ($n = 3,774$), the TSLS estimates tend to the OLS estimates[5]. In addition, since the market return contains measurement error, we know that the OLS estimates are downward biased and, consequently, so are the TSLS estimates. In the case of the highest beta, the Lincoln Financial stocks, the TSLS is slightly above the OLS estimates.

The Bayesian approach, in turn, delivers a greater CAPM beta for the two stocks, considering the posterior mean as a punctual estimation (see dashed red line in Figure 3). For General Mills stock, while $\hat{\beta}_{OLS} = 0.37$ and $\hat{\beta}_{TSLS} = 0.37$ the posterior mean of beta is 0.89, using the Horseshoe prior. For Lincoln Financial stock, while $\hat{\beta}_{OLS} = 1.81$ and $\hat{\beta}_{TSLS} = 1.86$ the posterior mean of beta is 2.12. Although we do not know the true beta in this case, we know that OLS is downward biased, which puts our approach in a

[5] In the extreme case where $p = n$ the two estimators are equivalent.

better position.

These discrepancies in estimated beta may have drastic implications for finance practitioners. As noted by Malloch et al. (2021), some analysis in finance, such as valuation, is very sensitive to the estimated beta. Thus, the many instruments approach with shrinkage priors can offer a better way to make such financial analysis.

The many instruments bias of TSLS arises because the first stage tends to fit too well, as discussed in section 5.1. The regularized Bayesian IV estimation, in turn, may avoid this tendency by shrinking unimportant instrument coefficients toward zeros. We now compare the regularized Bayesian IV and TSLS estimates of the coefficients of the instrumental variables δ_i. As before, we consider the Horseshoe prior in this analysis. Figure 4 presents the posterior distribution of δ's (in boxplot form) and the punctual TSLS estimator (red dot points). This Figure shows that for several coefficients the Bayesian regularization does shrink the coefficients toward zero. Also, in these cases, the punctual TSLS lies out of the interquartile range of the boxplot. It is the case, for example, of the coefficients $(\delta_{15}, \delta_{71}, \delta_{72}, \delta_{106}, \delta_{116}, \delta_{162}, \delta_{183}, \delta_{187}, \delta_{230}, \delta_{257}, \delta_{27})$. Thus, the discrepancy of CAPM beta observed in Figure 3 may be related to this regularization, which avoids overfitting in the "first stage" and delivers a better estimate of beta.

6 Conclusion

This paper contributes to the literature that works on the solution of Roll's (1977) critique, 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 error-in-variable present in the capital asset price model. These instruments, however, are usually week correlated with the endogenous market return, and use too many instruments may induce bias. This paper proposes to estimate the capital asset price model using a large set of instruments and shrinkage priors over the parameters associated with the instruments.

In a simulation exercise, the results show that the proposed approach reduces the bias of the CAPM beta estimates. We also verify empirically that our proposal produces greater betas than the OLS and TSLS methods. Since OLS is downward biased, the regularized Bayesian approach delivers better results than the traditional TSLS for this

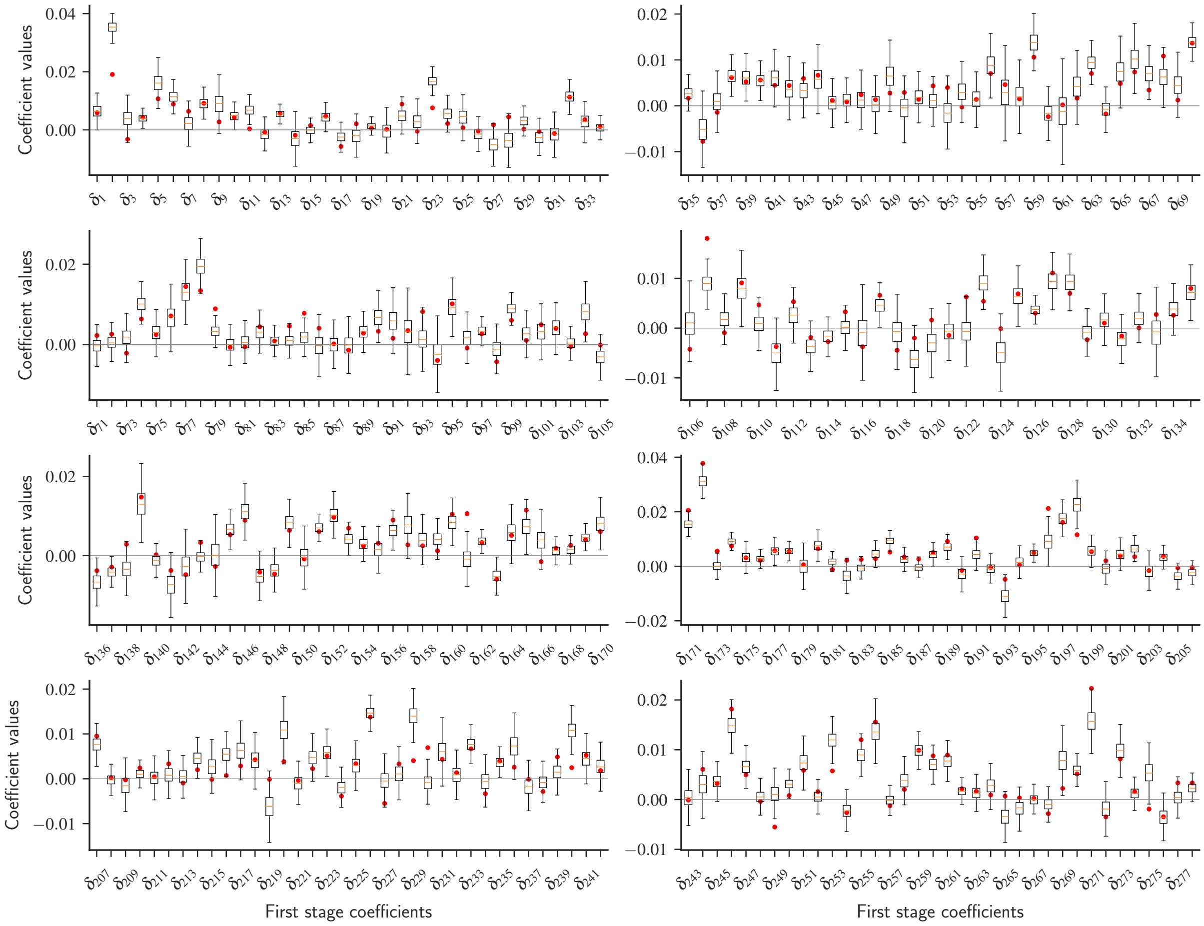

Figure 4: Posterior distribution of "first stage" δ_i coefficients. Red dots represents the Classical TSLS estimates. We consider the estimation of Lincoln Financial stocks

application. This find is supported by a large literature that shows beta's incapacity to appropriately measure the magnitude of risk. This difference in estimated betas is economically important since many financial models are sensitive to beta. A word of caution, however, is needed since even the Bayesian approach loses performance as the dimensionality increases.

Future research can extend the results of this work in at least two ways. The first is to generalize the model presented here to consider the joint estimation of all betas assets. This generalization can be a way to increase the efficiency of the estimates since it will take into account the covariance of all stock returns in the estimation. Second, a limited information Bayesian approach may be designed to also include prior regularization, in a similar manner as presented here.

References

Bai, J. and S. Ng (2010). Instrumental variable estimation in a data rich environment. *Econometric Theory 26*(6), 1577–1606.

Bekker, P. A. (1994). Alternative approximations to the distributions of instrumental variable estimators. *Econometrica: Journal of the Econometric Society 62*(30), 657–681.

Campbell, J. Y., A. W. Lo, A. W. Lo, and A. C. MacKinlay (1997). *The econometrics of financial markets*. princeton University press.

Carrasco, M. (2012). A regularization approach to the many instruments problem. *Journal of Econometrics 170*(2), 383–398.

Carvalho, C. M., N. G. Polson, and J. G. Scott (2010). The horseshoe estimator for sparse signals. *Biometrika 97*(2), 465–480.

Coën, A. and F. Éric Racicot (2007). Capital asset pricing models revisited: Evidence from errors in variables. *Economics Letters 95*(3), 443 – 450.

Davidson, R., J. G. MacKinnon, et al. (1993). *Estimation and inference in econometrics*, Volume 63. Oxford New York.

Fama, E. F. and J. D. MacBeth (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy 81*(3), 607–636.

Fuller, W. A. (1977). Some properties of a modification of the limited information estimator. *Econometrica: Journal of the Econometric Society*, 939–953.

Gibbons, M. R., S. A. Ross, and J. Shanken (1989). A test of the efficiency of a given portfolio. *Econometrica 57*(5), 1121–1152.

Hahn, P. R., J. He, and H. Lopes (2018). Bayesian factor model shrinkage for linear iv regression with many instruments. *Journal of Business & Economic Statistics 36*(2), 278–287.

Hansen, C., J. Hausman, and W. Newey (2008). Estimation with many instrumental variables. *Journal of Business & Economic Statistics 26*(4), 398–422.

Harvey, C. R. and G. Zhou (1990). Bayesian inference in asset pricing tests. *Journal of Financial Economics 26*(2), 221 – 254.

Jagannathan, R. and Z. Wang (1996). The conditional capm and the cross-section of expected returns. *The Journal of finance 51*(1), 3–53.

Jegadeesh, N., J. Noh, K. Pukthuanthong, R. Roll, and J. Wang (2019). Empirical tests of asset pricing models with individual assets: Resolving the errors-in-variables bias in risk premium estimation. *Journal of Financial Economics 133*(2), 273 – 298.

Jensen, M. C., F. Black, and M. S. Scholes (1972). The capital asset pricing model: Some empirical tests. *Studies in the Theory of Capital Markets*.

Kandel, S., R. McCulloch, and R. F. Stambaugh (1995). Bayesian inference and portfolio efficiency. *The Review of Financial Studies 8*(1), 1–53.

Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics 47*(1), 13–37.

Lopes, H. F. and N. G. Polson (2014). Bayesian instrumental variables: priors and likelihoods. *Econometric Reviews 33*(1-4), 100–121.

Malloch, H., R. Philip, and S. Satchell (2021). Estimation with Errors in Variables via the Characteristic Function. *Journal of Financial Econometrics*. nbab011.

Markovitz, H. M. (1959). *Portfolio selection: Efficient diversification of investments*. John Wiley.

Meng, J. G., G. Hu, and J. Bai (2011). Olive: a simple method for estimating betas when factors are measured with error. *Journal of Financial Research 34*(1), 27–60.

Newey, W. K. and R. J. Smith (2004). Higher order properties of gmm and generalized empirical likelihood estimators. *Econometrica 72*(1), 219–255.

Newey, W. K. and F. Windmeijer (2009). Generalized method of moments with many weak moment conditions. *Econometrica 77*(3), 687–719.

Ng, S. and J. Bai (2009). Selecting instrumental variables in a data rich environment. *Journal of Time Series Econometrics 1*(1), 1–32.

Park, T. and G. Casella (2008). The bayesian lasso. *Journal of the American Statistical Association 103*(482), 681–686.

Prono, T. (2015). Market proxies as factors in linear asset pricing models: Still living with the roll critique. *Journal of Empirical Finance 31*, 36 – 53.

Racicot, F.-É., W. F. Rentz, D. Tessier, and R. Théoret (2019). The conditional fama-french model and endogenous illiquidity: A robust instrumental variables test. *PloS one 14*(9), e0221599.

Roll, R. (1977). A critique of the asset pricing theory's tests part i: On past and potential testability of the theory. *Journal of Financial Economics 4*(2), 129 – 176.

Shanken, J. (1987a). A bayesian approach to testing portfolio efficiency. *Journal of financial economics 19*(2), 195–215.

Shanken, J. (1987b). Multivariate proxies and asset pricing relations: Living with the roll critique. *Journal of Financial Economics 18*(1), 91 – 110.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The journal of finance 19*(3), 425–442.

Simmet, A. and W. Pohlmeier (01 Aug. 2020). The capm with measurement error: 'there's life in the old dog yet!'. *Jahrbücher für Nationalökonomie und Statistik 240*(4), 417 – 453.

Stambaugh, R. F. (1982). On the exclusion of assets from tests of the two-parameter model: A sensitivity analysis. *Journal of Financial Economics 10*(3), 237 – 268.

Van Erp, S., D. L. Oberski, and J. Mulder (2019). Shrinkage priors for bayesian penalized regression. *Journal of Mathematical Psychology 89*, 31–50.

Appendix 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 $x \equiv (x_1, ..., x_n)'$ given $z \equiv (z_1, ..., z_n)'$ 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 $s_x/2$ scale parameter and $k_x/2$, it follows that the conditional posterior is:

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

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 (3×1) vector. From eq. (5) and (6) and using the bivariate normal properties, we can find the conditional distribution $\varepsilon_{y_t}|\varepsilon_{x_t} \sim N\left(\alpha(x_t - z_t'\delta), \xi^2\right)$, 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 eq. (6), we can conclude that: $y_t|x_t \sim \mathcal{N}\left(\gamma + x_t\beta + \alpha(x_t - z_t'\delta), \xi^2\right)$.

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

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

Combining this likelihood with the Normal-inverse-gamma prior:

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

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

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

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

Appendix 2: Results for small samples

Table 2: Different Measures for the beta estimation error, $n = 500$

		OLS	TSLS	LIML	BHS	BLASSO	BFBS
$p = 2$	Mean bias	-0.449	-0.340	**0.011**	0.062	0.049	0.054
	Mean abs. bias	0.449	1.963	**0.343**	0.380	0.371	0.376
	RMSE	0.469	9.946	**0.430**	0.483	0.469	0.475
$p = 10$	Mean bias	-0.448	-0.028	**-0.004**	-0.011	-0.027	-0.030
	Mean abs. bias	0.448	0.212	0.209	**0.206**	0.202	0.202
	RMSE	0.468	0.265	0.263	**0.259**	0.253	0.253
$p = 20$	Mean bias	-0.445	-0.044	**0.001**	-0.022	-0.047	-0.058
	MAE	0.445	0.178	0.184	0.176	**0.173**	0.174
	RMSE	0.465	0.225	0.230	0.220	**0.217**	0.218
$p = 40$	Mean bias	-0.448	-0.071	**-0.001**	-0.053	-0.094	-0.119
	Mean abs. bias	0.448	0.166	0.169	**0.161**	0.165	0.174
	RMSE	0.468	0.207	0.211	**0.199**	0.205	0.215
$p = 80$	Mean bias	-0.449	-0.124	**0.002**	-0.126	-0.184	-0.237
	Mean abs. bias	0.449	0.177	0.175	**0.173**	0.204	0.245
	RMSE	0.469	0.221	0.219	**0.217**	0.247	0.285
$p = 160$	Mean bias	-0.441	-0.203	**0.012**	-0.282	-0.357	-0.387
	Mean abs. bias	0.441	0.219	**0.183**	0.285	0.357	0.387
	RMSE	0.462	0.260	**0.228**	0.320	0.384	0.410

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

Appendix 3: Additional Bayesian estimation results

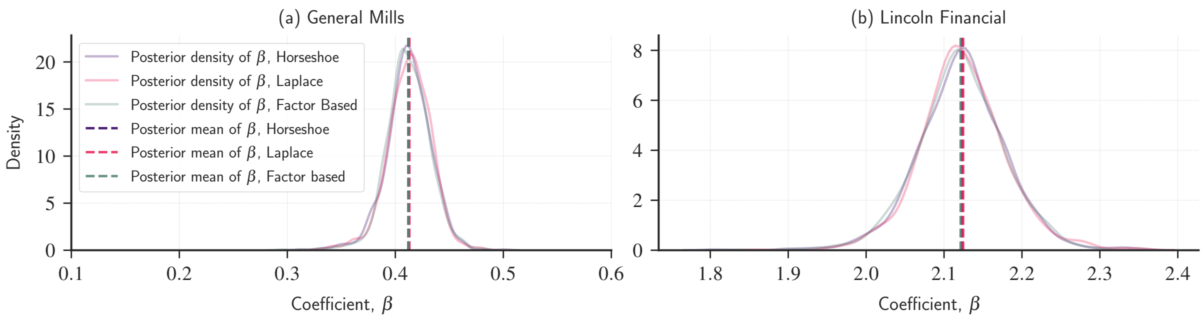

Figure 5

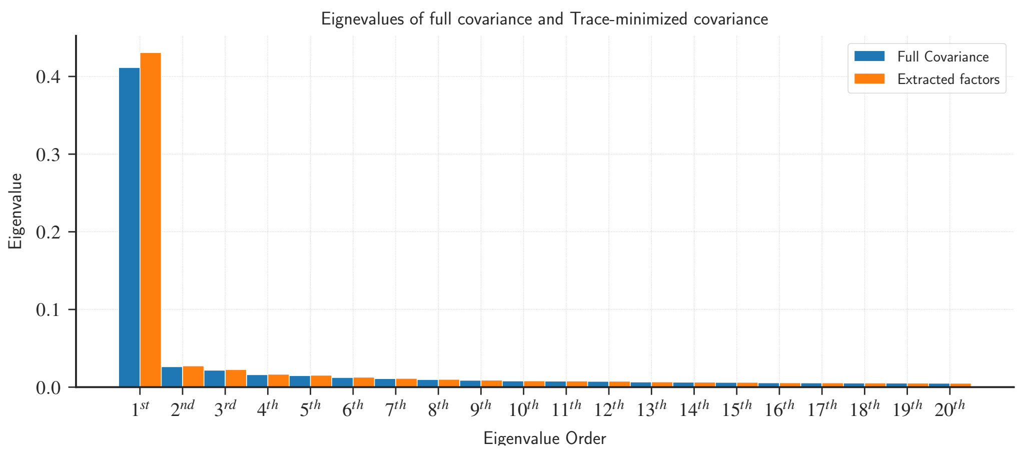

Figure 6

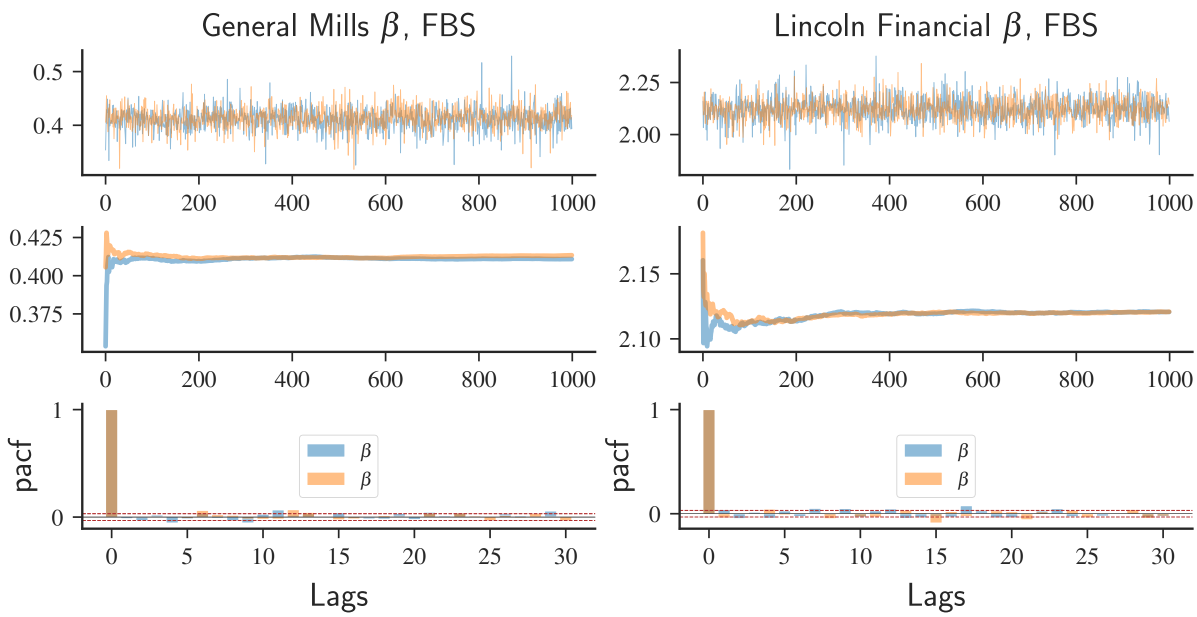

Figure 7

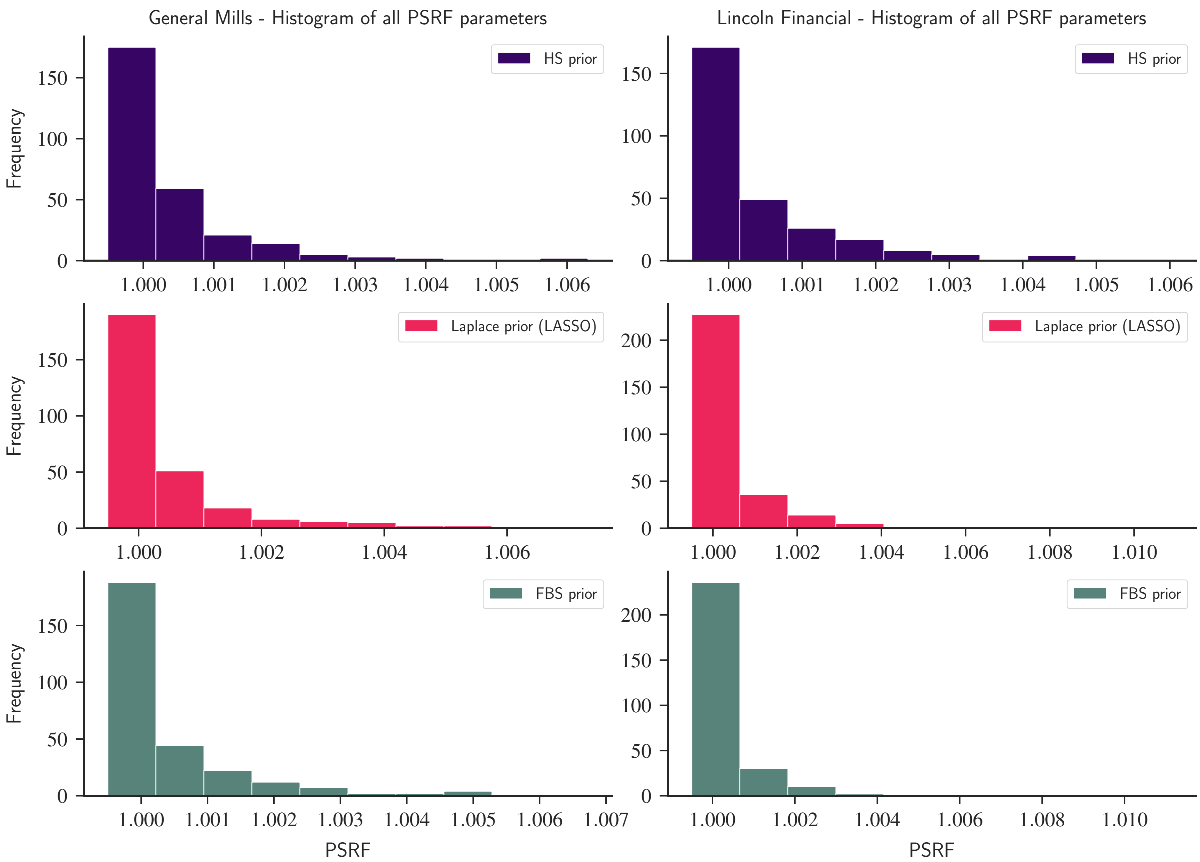

Figure 8