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 approach to estimate the capital asset price model (CAPM) in the presence of 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 to the market portfolio, although the data-rich environment available in finance allows us to use many-instrument (possibly weak) settings. We use regularization priors to deal with a large number of instrumental variables and improve inference of the estimated CAPM beta. Using simulated data, we find that our approach may reduce the average 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, we find that the estimated CAPM stock’s beta is greater than the ordinary least square estimator, while the traditional two-stage least square converges to the admittedly biased least squares estimator.

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. We carry out two simulations exercises: the first one, considers that the market return is a weighted sum of assets returns and the second one we simulate a classical measurement error problem. In the first experiment, we found that our approach can improve the average bias up to 90%, but this gain is attenuated as the number of instruments increase. The second experiment, the average bias is reduced by only 35%, and, again, the 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 biased OLS estimator.

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.

\[ E[R_i] - R_f = \beta_i (E[R_m] - R_f), \]

where \( \beta_i = \frac{\sigma_{im}}{\sigma_m^2} \). 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
\[ R_{it} - R_{ft} = \gamma_i + \beta_i (R_{mt} - R_{ft}) + \varepsilon_{it}, \quad \varepsilon_{it} \sim N(0, \sigma_i^2). \tag{2} \]

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 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

\[ y_{it} = \gamma_i + \beta_i x_t + \varepsilon_{it}, \quad \varepsilon_{it} \sim N(0, \sigma_i^2), \tag{3} \]
\[ x_t = x^*_t + u_t, \quad u_t \sim N(0, \sigma_u^2). \tag{4} \]

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.

Even before Roll’s (1977) critique, the time series approach had to deal with measurement error problems, although of a different nature. The first attempt to test the CAPM was introduced by Jensen et al. (1972), who proposed to regress the excess return on a constant, $\gamma_i$, and the excess of the return on the market portfolio. If the estimation considers securities individually, the test of $\gamma_i = 0$ would introduce inefficiency and bias, since the residual of each regression may be correlated, as argued by King (1966). Considering the estimation with a single security does not consider the information contained in the other assets correlated with this security. The solution to this problem was to group the securities in portfolios, according to their risk, measured by the associated beta. However, if one uses the estimated beta from the sample at hand to group the assets, it would introduce a selection bias due to measurement errors: Assets that are in the group of higher betas may be in this group either because it indeed has a high beta or because it was overestimated (positive measurement error), and vice-versa. To avoid such a problem Jensen et al. (1972) used pre-sample estimated betas to form ten different portfolios and test the CAPM for these ten regressions. The results show that the CAPM is not valid for all portfolios using the CCRSP\(^1\) as a proxy to the market return.

The measurement error discussed by Roll (1977) has a different nature. It is related to the fact that only a proxy of the market return is observable. 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). These authors estimate a model similar to Jensen et al. (1972), but propose an approach to test if $\gamma_i = 0$ jointly. Gibbons et al. (1989) result also points to the failure of CAPM. 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 Classical methods. A first answer to the Roll’s (1977) critique was a branch of research that advocated that the latency of market return is not an empirical problem. Stambaugh (1982) proposed to consider alternative proxies to the market return and to evaluate if inferences about the theory validity change. The chosen proxies ranged from a simple market portfolio, containing only shares, to a more complete portfolio, including real states and consumption of durables in addition to stocks. Testing from the more complete market portfolio and sequentially excluding the components of

\(^{1}\)Chicago Center for Research in Security Prices
this portfolio, the author concluded that there is no change in the inference and, therefore, the Roll’s (1977) critique is not an empirical problem. In the same direction, examples of research that tried to build more precise proxies are Jagannathan and Wang (1996) and Campbell et al. (1997), which include human capital when measuring the aggregate return.

Using multiple proxies, Shanken (1987a) generalizes the problem in a Bayesian framework. This author consider linear combinations of multiples surrogates to the market portfolio and consider a prior information about the correlation between the true market portfolio and the proxy used in estimation. If the prior belief about the correlation between true market portfolio and the market index used as a proxy is enough high, then the following result is true: a rejection of theory using a proxy implies a rejection of CAPM for the true market portfolio (Shanken, 1987b). In this sense, if one believes that the proxy is highly correlated with the true market portfolio, the latency of market portfolio is not an empirical problem.

Different from the traditional frequentist approach, the Bayesian inference arises as a new form to incorporate prior information and test the theory (Shanken, 1987a; Harvey and Zhou, 1990; Kandel et al., 1995). In this setup, the CAPM can be tested using odds-ratio. Moreover, prior information for all parameters can be used, instead of only for the correlation parameter introduced by Shanken (1987a). Regarding this parameter, Kandel et al. (1995) estimated the posterior distribution of the correlation between proxy and market portfolio. Their result, however, shows that this posterior distribution is sensitive to the choice of a prior distribution. More recent research also used the Bayesian framework to test and dealt with the imperfection of proxies for market return (Cederburg and O’Doherty, 2015). The focus, however, is to econometrically control for the effects of the measurement error instead of using the correlation between proxies and market portfolios. In this sense, Cederburg and O’Doherty (2015) built a hierarchical Bayesian approach to estimate the CAPM, considering the error-in-variables problem. The authors find evidence in favor of CAPM.

On the frequentist side, the literature appealed to the instrumental variables estimation to deal with the measurement error problem. Examples of this approach are Coën and Éric Racicot (2007); Meng et al. (2011); Racicot 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.

Another recent example of instrumental variables is the work of Jegadeesh et al. (2019). Besides, these authors solve the problem of estimation for single securities (instead of the grouped securities)
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:

\[
x_t = z_t^\prime \delta + \varepsilon_{xt}, \quad (5)
\]

\[
y_t = \gamma + x_t^\prime \beta + \varepsilon_{yt}, \quad t \in \{1, \cdots, n\} \quad (6)
\]

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

\[
\begin{bmatrix}
\varepsilon_{xt} \\
\varepsilon_{yt}
\end{bmatrix}
\sim N
\begin{bmatrix}
0 \\
0
\end{bmatrix},
\begin{bmatrix}
\sigma^2_x & \sigma_{xy} \\
\sigma_{yx} & \sigma^2_x
\end{bmatrix}
\]
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 $\delta$ 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 $\delta$, 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 $\delta$. To understand the Bayesian estimation of IV regression, consider the reduced form of equations (5) and (6):

$$x_t = z_t' \delta + \nu_{x_t}, \quad (7)$$
$$y_t = \gamma + z_t' \delta \beta + \nu_{y_t}, \quad (8)$$

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:

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

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

$$f(x, y|z, \Theta) = f(y|x, \delta, \alpha, \beta, \xi^2) \times f(x|z, \sigma^2_x \delta)$$
$$= \mathcal{N} \left( \gamma + x_t \beta + \alpha (x_t - z_t' \delta), \xi^2 \right) \times \mathcal{N}(z_t' \delta, \sigma^2_x). \quad (10)$$

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

$$\delta \sim \text{arbitrary}, \quad (11)$$
$$\sigma^2_x \sim \mathcal{IG}(\text{shape} = k_x, \text{scale} = s_x), \quad (12)$$
$$(\xi^2, \gamma, \beta, \alpha)' \sim \mathcal{NIG} \left(0, \xi^2 \Sigma_0^{-1}, \text{shape} = \frac{k}{2}, \text{scale} = \frac{s}{2} \right). \quad (13)$$

Combining these priors with likelihood function in equation (10), give us the posterior distribution. To sample from this posterior 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 $^3 \hat{x} \equiv (1, x, z'\delta)$, $M \equiv \Sigma_0 + \hat{x}'\hat{x}$, $a \equiv k+n$ and $b \equiv s+y'y - y'\hat{x}M^{-1}\hat{x}'y$. 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:

\footnote{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 $(p \times n)$.}
Full conditional posterior for $\delta|\Theta, \text{data}$: given $\Theta$, 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 $\delta$, 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

1: procedure SliceSampler($\delta, \sigma^2_x, x, Z, y$)
2: Define $\hat{\delta} = (ZZ')^{-1}Zx$ and $\Delta = \delta - \hat{\delta}$
3: Draw $\zeta \sim \mathcal{N}(0, \sigma^2_x(ZZ')^{-1})$ 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 $\delta$: $\Delta = \Delta \cos(\varphi) + \zeta \sin(\varphi)$ and $\bar{\delta} = \hat{\delta} + \Delta$
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 $\varphi \sim U(lower, upper)$
10: Update $\Delta$ and $\delta$: $\Delta = \Delta \cos(\varphi) + \zeta \sin(\varphi)$ and $\bar{\delta} = \hat{\delta} + \Delta$
11: $\delta \leftarrow \hat{\delta} + \Delta$
12: return $\delta$

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

Full conditional posterior for $\sigma^2_x|\Theta, \text{data}$: Fortunately, for an inverse-gamma prior on $\sigma^2_x$ the full conditional posterior for $\sigma^2_x$ have a closed form. Combining the likelihood $f(x|Z, \delta, \sigma^2_x)$ 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 $s_x + \sum^n_{i=1}(x_t - z_i\delta)^2$ (see derivation in appendix).

Full conditional posterior for $(\xi^2, \gamma, \beta, \alpha)'|\Theta, \text{data}$: 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 $y_t|x_t \sim N(\tilde{x}_t\theta, \xi^2)$. Combining this likelihood with the prior give in (13), it can be show that the full conditional posterior $(\xi^2, \gamma, \beta, \alpha)'|\Theta, \text{data}$ is a Normal-Inverse-Gamma distribution. Specifically:

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

(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 $\delta$, and it still works well. In particular, we can elicit several shrinks prior over $\delta$, since the many instruments setting requires regularization.
4.2 Shrinkage priors for $\delta$

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, Double-exponential (also known as Laplace), and Horseshoe densities (Carvalho et al., 2010). The horseshoe density is the stronger one, followed by Double-exponential. Although the Cauchy density is relatively weaker, it also concentrates much density around zero, and so it also works as a shrinkage prior. Figure 1 depicts these three priors for a single parameter.

![Density plot](image)

Figure 1: Three sort of shrinkage priors: Horeseshoe, Double-exponential and Cauchy, all of them centered at 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 $i, j \in \{1, \cdots, p\}$. Although it may work, it neglects the covariance between the instruments. To consider the covariance between the instruments, a more sophisticated prior needed, 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:

\[
\text{Cov}(z_t) = BB' + \Psi^2, \tag{14}
\]

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 on the case where $k << p$. 

10
$k$ 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 $E[f_t|z_t] = A z_t =: \hat{f}_t$, with $A \equiv B'(BB' + \Psi^2)^{-1}$.

Now, consider the factor regression model:

$$
\begin{aligned}
x_t &= \theta \hat{f}_t + \varepsilon_t, \\
\end{aligned}
$$

where $\theta$ 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 $\delta$ 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 (15) needs to be modified

$$
\begin{aligned}
x_t &= \theta \hat{f}_t + \eta \hat{r}_t + \varepsilon_t, \\
\end{aligned}
$$

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 pseudo-inverse of $A$. In this case, it can be show that $\delta' = \theta A + \eta(I_p + A^+ A)$.

Defining $\delta' = (\theta, \eta)$, 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 $\delta$ 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 $\omega$. With this identity, conditional on $\omega$, we can impose an horseshoe prior, for instance, on $\delta$ and it induce a prior on $\tilde{\delta}$. That is:

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

Following Hahn et al. (2018), 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 induce a prior on $\delta$ 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 $\delta$. In practice, however, the matrices $B, \Psi$ 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, LASSO, 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, which we took from Kenneth website. We consider the sp500 returns as the surrogate for the market return. Finally, we consider 200 assets listed in the sp500 as the return of the assets. The data ranges from 2020-03-23 to 2021-12-31, totaling 451 observations. This choice of a relatively short sample period and high dimensionality in the set of instruments represents the context in which the measurement error problem in the measurement of the market portfolio is more relevant, and also where the bias is introduced by the dimensionality of the instruments in the estimation can be stronger.

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. We consider two scenarios of simulated data. Finally, we use the observed data presented before to estimate the CAPM using our proposal procedure.

5.1 Monte Carlo analysis: Simulation procedures

We consider two scenarios to simulate CAPM. First, we begin by simulating the assets return and build the true market return based on these assets returns. In the second scenario, we use a traditional classical measurement error model. In both cases, we use descriptive statistics of observed data to calibrate the parameters, so that the simulated data is a good approximation of the data.

5.1.1 First strategy: Market return as weighted sum of assets return

In the first simulation exercise, we consider that the true market return is a weighted average of all $M$ assets of the economy, similar to the simulation used in Simmet and Pohlmeier (2020). In this case, we need to specify the weights $b_i$ of each asset $i$. With these weights and assuming that each asset return is Gaussian, we can build the true market return $x^*_t$, which, in turn, is used to build the observed market return by adding an error $u_t$. Specifically, we simulated the CAPM model following the steps below.

We begin by drawing the weights $b_i$ from a uniform distribution: half of $b_i$ are drawn from a $U(0, 1)$ and the other half from a $U(1000, 1001)$ so that we can differentiate between assets with more weights. We normalize the weights so that $\sum_{i=1}^{M} b_i = 1$. Then

4https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/
we simulate the assets return from a Gaussian distribution with mean zero and variance \( \sigma^2 \). We build the true market return \( x_t^* = \sum_{i=1}^{M} b_i y_{it} \) and assume that the total number of assets in the economy is \( M \in \{20, 80, 140\} \). The observed market return is \( x_t = x_t^* + u_t \), where \( u_t \sim \mathcal{N}(0, .001) \). In this case, although we do not know the values of the true \( \beta \)'s we can estimate it by regressing \( y_{it} \) on the true market return \( x_t^* \). We can use the OLS estimate of this regression as the true \( \beta \)'s, since it is a consistent estimator for \( \beta \), even though it possesses sample error. We generate \( 10^4 \) observations to obtain this consistent estimation, although we use only 500 observations for the comparison of the estimators.

We repeat this simulated procedure 100 times, and at each iteration, we estimate \( \beta \) using OLS, TSLS, and Bayesian IV with shrinkage prior over \( \delta \). We consider three types of priors for \( \delta \): i) the horseshoe, ii) the Laplace, which is known as Bayesian LASSO, iii) and the factor based shrinkage prior explained in section 4.2.2. For all instrumental variable estimation we consider the set of instruments is formed by all assets except the regressand in the CAPM equation. 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 CAPM equation. Table 1 presents the descriptive statistic for true betas and the estimated betas using five different methods:

<table>
<thead>
<tr>
<th>Number of assets, ( M )</th>
<th>True</th>
<th>OLS</th>
<th>TSLS</th>
<th>BIV-HS</th>
<th>BIV-LASSO</th>
<th>BIV-FSP</th>
</tr>
</thead>
<tbody>
<tr>
<td>min</td>
<td>0.895</td>
<td>0.482</td>
<td>-0.447</td>
<td>0.465</td>
<td>0.481</td>
<td>0.468</td>
</tr>
<tr>
<td>q25</td>
<td>0.977</td>
<td>0.797</td>
<td>-0.061</td>
<td>0.799</td>
<td>0.813</td>
<td>0.805</td>
</tr>
<tr>
<td>median</td>
<td>1.003</td>
<td>0.898</td>
<td>0.051</td>
<td>0.896</td>
<td>0.911</td>
<td>0.911</td>
</tr>
<tr>
<td>mean</td>
<td>1.004</td>
<td>0.901</td>
<td>0.036</td>
<td>0.897</td>
<td>0.911</td>
<td>0.907</td>
</tr>
<tr>
<td>q75</td>
<td>1.033</td>
<td>0.997</td>
<td>0.143</td>
<td>0.996</td>
<td>1.009</td>
<td>1.006</td>
</tr>
<tr>
<td>max</td>
<td>1.123</td>
<td>1.290</td>
<td>0.540</td>
<td>1.307</td>
<td>1.297</td>
<td>1.302</td>
</tr>
<tr>
<td>Number of assets, ( M )</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>min</td>
<td>0.829</td>
<td>-0.007</td>
<td>-0.587</td>
<td>-0.095</td>
<td>0.018</td>
<td>0.013</td>
</tr>
<tr>
<td>q25</td>
<td>0.951</td>
<td>0.597</td>
<td>0.020</td>
<td>0.550</td>
<td>0.570</td>
<td>0.555</td>
</tr>
<tr>
<td>median</td>
<td>1.001</td>
<td>0.776</td>
<td>0.185</td>
<td>0.730</td>
<td>0.732</td>
<td>0.725</td>
</tr>
<tr>
<td>mean</td>
<td>0.996</td>
<td>0.743</td>
<td>0.166</td>
<td>0.704</td>
<td>0.705</td>
<td>0.702</td>
</tr>
<tr>
<td>q75</td>
<td>1.052</td>
<td>0.903</td>
<td>0.346</td>
<td>0.858</td>
<td>0.859</td>
<td>0.870</td>
</tr>
<tr>
<td>max</td>
<td>1.150</td>
<td>1.274</td>
<td>0.851</td>
<td>1.229</td>
<td>1.232</td>
<td>1.206</td>
</tr>
<tr>
<td>Number of assets, ( M )</td>
<td>80</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>min</td>
<td>0.698</td>
<td>-0.179</td>
<td>-0.503</td>
<td>-0.188</td>
<td>-0.198</td>
<td>-0.115</td>
</tr>
<tr>
<td>q25</td>
<td>0.944</td>
<td>0.359</td>
<td>-0.005</td>
<td>0.352</td>
<td>0.334</td>
<td>0.352</td>
</tr>
<tr>
<td>median</td>
<td>0.991</td>
<td>0.587</td>
<td>0.212</td>
<td>0.545</td>
<td>0.551</td>
<td>0.544</td>
</tr>
<tr>
<td>mean</td>
<td>0.999</td>
<td>0.574</td>
<td>0.227</td>
<td>0.534</td>
<td>0.528</td>
<td>0.535</td>
</tr>
<tr>
<td>q75</td>
<td>1.054</td>
<td>0.781</td>
<td>0.478</td>
<td>0.725</td>
<td>0.724</td>
<td>0.717</td>
</tr>
<tr>
<td>max</td>
<td>1.294</td>
<td>1.431</td>
<td>1.080</td>
<td>1.257</td>
<td>1.356</td>
<td>1.308</td>
</tr>
<tr>
<td>Number of assets, ( M )</td>
<td>140</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The main result in Table 1 is that while the distribution of the TSLS estimator dramatically shifted to the left, concerning the distribution of the true Beta (column 1), the Bayesian Instrumental Variable estimator does not accompany this shift (columns 4-6). All the three Bayesian estimations presented similar results, meaning that the effect of the three sorts of priors over instrumental variable coefficients is equivalent in this simulation. This result remains valid for all number of assets, \( M \), considered.
The true beta in the first column of Table 1 is distributed about one. Because of
the measurement error introduced, the OLS estimator is less than the true beta. Although
the Bayesian instrumental variable estimator has improved concerning the two-stage least
squares, the mean of estimated beta estimated by BIV approach is slightly lesser than
the OLS estimator, especially for the large set of instruments ($M = 80$ and $M = 140$). In
this sense, the estimation of a model remains problematic.

To further investigate the improvement of the Bayesian estimator concerning the
traditional two-stage least square, we analyze the mean error (ME) and the root of mean
squared error (RMSE), and the results are presented in Table 2. Both in terms of ME
and RMSE, the Bayesian approach has ameliorated concerning the TSLS. This gain of
accuracy, however, is attenuated as the number of instruments become large. For $M = 20$,
the mean error has improved around 90% and the root of mean squared error around 83%.
When we consider $M = 140$ assets, these gains diminish to 40% and 35% for the mean
error and the root of squared error.

Table 2: Different Measures of accuracy for the beta estimation relative to the TSLS.

<table>
<thead>
<tr>
<th>Number of assets</th>
<th>TSLS</th>
<th>BHS</th>
<th>BLASSO</th>
<th>BFBS</th>
</tr>
</thead>
<tbody>
<tr>
<td>ME</td>
<td>1.000</td>
<td>0.111</td>
<td>0.096</td>
<td>0.100</td>
</tr>
<tr>
<td>RMSE</td>
<td>1.000</td>
<td>0.175</td>
<td>0.164</td>
<td>0.170</td>
</tr>
<tr>
<td>Number of assets</td>
<td>80</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ME</td>
<td>1.000</td>
<td>0.352</td>
<td>0.351</td>
<td>0.354</td>
</tr>
<tr>
<td>RMSE</td>
<td>1.000</td>
<td>0.427</td>
<td>0.423</td>
<td>0.423</td>
</tr>
<tr>
<td>Number of assets</td>
<td>140</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ME</td>
<td>1.000</td>
<td>0.603</td>
<td>0.610</td>
<td>0.601</td>
</tr>
<tr>
<td>RMSE</td>
<td>1.000</td>
<td>0.646</td>
<td>0.651</td>
<td>0.643</td>
</tr>
</tbody>
</table>

Note: Values less than one indicate that the estimator outperforms the TSLS.

5.1.2 Second strategy: Classical measurement error

Consider a classical additive measurement error:

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

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

for $i \in \{1, ..., M\}$ and $t \in \{1, ..., 1000\}$. In equations (17) to (18) $x_t^*$ is the true market
return and it is assumed to be Gaussian with mean equal to zero and variance $\sigma_x^2$, $u_t$ is
Gaussian measurement error, with variance $\sigma_u^2$, $x_t$ 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 can construct the assets return. The
first and second moment of the error term $\varepsilon_t$ are, respectively, $E[\varepsilon_t] = 0$ and $E[\varepsilon_t \varepsilon_{t'}] = \sigma_{\varepsilon}^2$,
and we assume that it is Gaussian.

To simulate the model, we need to calibrate the parameters ($\sigma_u^2, \sigma_x^2, \sigma_{\varepsilon}^2, \beta$). For the
$\beta$, we consider a linear grid between 0.5 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 SP500, 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 $\sigma_u = \sigma_x \times 0.9$ to create a situation with a high measurement error but assure the model is identifiable. These parameters allow us to simulate all variables of interest in CAPM.

To evaluate the accuracy of each method, we simulated the model $N_{sim} = 100$ times. At each iteration, we estimated the parameters using four methods. The first one was the traditional Ordinary Least Squares (OLS), regressing $y_t$ on $x_t$, which is known to be inconsistent in the presence of measurement errors. Second, we consider an instrumental variable approach, by estimating a Two-Stage Least Squares (TSLS) and we used the set of instruments formed by all assets except the regressand in the CAPM equation. 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. Finally, we consider the same set of instruments to estimate the model using the Bayesian method described in section 3, and, in particular, we used the Horseshoe, Laplace, and Factor-Based Shrinkage prior distributions over $\delta$. This estimation is referred to as BIV-HS, BIV-LASSO, and BIV-FBS, once the Laplace distribution is known as Bayesian LASSO.

We simulate and estimate the model for three different numbers of assets, 20, 80, and 140, as shown by Table 3. In the estimation process, we consider the asset with $\beta = 1$. Because of measurement error, the OLS estimator is downward biased.

<table>
<thead>
<tr>
<th>Number of assets, $M$</th>
<th>OLS</th>
<th>TSLS</th>
<th>BIV-HS</th>
<th>BIV-LASSO</th>
<th>BIV-FBS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>20</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>min</td>
<td>0.357</td>
<td>0.667</td>
<td>0.661</td>
<td>0.657</td>
<td>0.656</td>
</tr>
<tr>
<td>q25</td>
<td>0.519</td>
<td>0.894</td>
<td>0.903</td>
<td>0.896</td>
<td>0.893</td>
</tr>
<tr>
<td>median</td>
<td>0.562</td>
<td>0.976</td>
<td>0.984</td>
<td>0.973</td>
<td>0.973</td>
</tr>
<tr>
<td>mean</td>
<td>0.555</td>
<td>0.973</td>
<td>0.982</td>
<td>0.972</td>
<td>0.969</td>
</tr>
<tr>
<td>q75</td>
<td>0.604</td>
<td>1.055</td>
<td>1.062</td>
<td>1.050</td>
<td>1.044</td>
</tr>
<tr>
<td>max</td>
<td>0.788</td>
<td>1.247</td>
<td>1.216</td>
<td>1.198</td>
<td>1.196</td>
</tr>
<tr>
<td><strong>80</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>min</td>
<td>0.400</td>
<td>0.680</td>
<td>0.706</td>
<td>0.671</td>
<td>0.663</td>
</tr>
<tr>
<td>q25</td>
<td>0.503</td>
<td>0.862</td>
<td>0.881</td>
<td>0.843</td>
<td>0.832</td>
</tr>
<tr>
<td>median</td>
<td>0.550</td>
<td>0.933</td>
<td>0.940</td>
<td>0.906</td>
<td>0.884</td>
</tr>
<tr>
<td>mean</td>
<td>0.552</td>
<td>0.936</td>
<td>0.949</td>
<td>0.908</td>
<td>0.893</td>
</tr>
<tr>
<td>q75</td>
<td>0.607</td>
<td>0.998</td>
<td>1.006</td>
<td>0.967</td>
<td>0.951</td>
</tr>
<tr>
<td>max</td>
<td>0.712</td>
<td>1.224</td>
<td>1.240</td>
<td>1.181</td>
<td>1.151</td>
</tr>
<tr>
<td><strong>140</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>min</td>
<td>0.373</td>
<td>0.620</td>
<td>0.639</td>
<td>0.610</td>
<td>0.590</td>
</tr>
<tr>
<td>q25</td>
<td>0.494</td>
<td>0.819</td>
<td>0.813</td>
<td>0.752</td>
<td>0.736</td>
</tr>
<tr>
<td>median</td>
<td>0.543</td>
<td>0.885</td>
<td>0.853</td>
<td>0.803</td>
<td>0.778</td>
</tr>
<tr>
<td>mean</td>
<td>0.545</td>
<td>0.882</td>
<td>0.869</td>
<td>0.808</td>
<td>0.785</td>
</tr>
<tr>
<td>q75</td>
<td>0.596</td>
<td>0.939</td>
<td>0.924</td>
<td>0.857</td>
<td>0.840</td>
</tr>
<tr>
<td>max</td>
<td>0.761</td>
<td>1.191</td>
<td>1.102</td>
<td>1.044</td>
<td>1.030</td>
</tr>
</tbody>
</table>

When we consider 20 instruments, all instrumental variable estimators achieve a mean value close to the true beta, and the BIV-HS is the one with the best performance.
Increasing the number of instruments to 80, all the instruments-based estimators diminish their performance, but the BIV-HS is still closest to the true value of one. The other Bayesian estimators based on Laplace and factor-based shrinkage prior, however, are worse than the two-stage least square, on average. Finally, when the number of instruments is 140, all estimators get closer to OLS estimator, and in this case, any Bayesian estimator can beat the two-stage least square.

We also compare some measures of accuracy that take into account both the mean error (ME) and the root mean of squared error (RMSE), which are presented in Table 4. These metrics are presented relative to the TSLS, so that values less than one mean that the estimator outperforms the TSLS, while values greater than one means that the estimator underperforms the TSLS. The results in Table 4 are in line with those presented in Table 3: the only Bayesian prior able to beat the TSLS is the Horseshoe and it just happens up to 80 instruments. Specifically, the BIV-HS produced a gain of 34% in the mean error for 20 instruments and 20% for 80 instruments. These gains are attenuated when we consider the RMSE measure.

Table 4: Different Measures of accuracy for the beta estimation relative to the TSLS.

<table>
<thead>
<tr>
<th>Number of assets, M</th>
<th>OLS</th>
<th>TSLS</th>
<th>BIV-HS</th>
<th>BIV-LASSO</th>
<th>BIV-FBS</th>
</tr>
</thead>
<tbody>
<tr>
<td>ME</td>
<td>16.566</td>
<td>1.000</td>
<td>0.658</td>
<td>1.045</td>
<td>1.146</td>
</tr>
<tr>
<td>RMSE</td>
<td>3.559</td>
<td>1.000</td>
<td>0.976</td>
<td>0.978</td>
<td>0.983</td>
</tr>
<tr>
<td>Number of assets, M</td>
<td>80</td>
<td>ME</td>
<td>7.027</td>
<td>1.000</td>
<td>0.804</td>
</tr>
<tr>
<td>RMSE</td>
<td>3.831</td>
<td>1.000</td>
<td>0.965</td>
<td>1.136</td>
<td>1.210</td>
</tr>
<tr>
<td>Number of assets, M</td>
<td>140</td>
<td>ME</td>
<td>3.857</td>
<td>1.000</td>
<td>1.112</td>
</tr>
<tr>
<td>RMSE</td>
<td>3.025</td>
<td>1.000</td>
<td>1.031</td>
<td>1.369</td>
<td>1.503</td>
</tr>
</tbody>
</table>

Note: Values less than one indicate that the estimator outperforms the TSLS.

5.2 Empirical Application

In this section, we use the Bayesian regularization procedure to estimate the CAPM using observed data. We estimate the model for three arbitrary assets: Agilent Technology, eBay, and Amazon. The instrument consists of all the other 200 stock returns listed in the SP500, and for each, the data are available for the period between 2020-03-23 and 2021-12-31. We estimate the model using three types of prior over the instrumental variable coefficients and the traditional OLS and TSLS.

The posterior distribution for CAPM beta was similar for the three Bayesian estimations (BIV-HS, BIV-LASSO, and BIV-FBS). Thus, it indicates that, for these stocks, the three priors perform equally. In the estimation process, the Gibbs schemes required many iterations to converge to the stationary distribution. Specifically, it required $10^4$ iterations to warm up the Gibbs scheme, and $10^7$ iterations saved every 100th to diminish autocorrelation of the Markov chain.

When comparing the CAPM beta estimated by the Bayesian IV approaches with those found by OLS and TSLS estimators, there is a difference, especially for the Amazon stock. Figure 2 presents the posterior distribution for the beta estimated by BIV-HS, as
well as its mean, which is used as a punctual estimator compared with OLS and TSLS estimates.

![Figure 2: Posterior distribution of CAPM beta estimated by Bayesian Instrumental Variable with Horseshoe prior for three assets: (a) Agilent Technology, (b) eBay and (c) Amazon. The figure also present mean of the posterior distribution and the OLS and TSLS estimates.](image)

For all three stocks, 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, \( p = 200 \)) relative to the number of observations (\( n = 451 \)), the TSLS estimates tend to the OLS estimates. 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. The Bayesian approach, however, delivers a greater CAPM beta in all three cases (see dashed red line in figure 2). For Agilent Technology stock, while \( \hat{\beta}_{OLS} = 0.86 \) and \( \hat{\beta}_{TSLS} = 0.86 \) the mean of posterior of beta from BIV is 0.89. For eBay stock, while \( \hat{\beta}_{OLS} = 0.72 \) and \( \hat{\beta}_{TSLS} = 0.73 \) the mean of posterior of beta from BIV is 0.80. Finally, for Amazon stock, while \( \hat{\beta}_{OLS} = 0.77 \) and \( \hat{\beta}_{TSLS} = 0.73 \) the mean of posterior of beta from BIV is 1.21. 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.

These discrepancies of estimated beta may have drastic implications for finance practitioners. As noted by Malloch et al. (2020), 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.

6 Final Remarks

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. However, these instruments 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 in the estimation of the CAPM. In an empirical application, we verify that our

---

5In the extreme case where \( p = n \) the two estimators are equivalent.
approach produces greater betas than the OLS and TSLS methods. These results seem to put the Bayesian regularization approach in a better position when estimating CAPM model with many instruments. A word of caution, however, is needed, since even the Bayesian approach lose performance as the dimensionality increases.

References


**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 $x \equiv (x_1, \ldots, x_n)'$ given $z \equiv (z_1, \ldots, z_n)'$ 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:

$$\pi(\sigma_x|x, y, Z, \delta) \propto (\sigma_x^2)^{\frac{-k_x-n-2}{2}} \exp\left\{ -\frac{1}{2\sigma_x^2} \left[ s_x + \sum_{t=1}^{n} (x_t - z_t'\delta)^2 \right] \right\},$$

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 $(3 \times 1)$ vector. From eq. (5) and (6) and using the bivariate normal properties, we can find the conditional distribution $\epsilon_{yi|x_t} \sim \mathcal{N}(\alpha(x_t - z_t'\delta), \xi^2)$, where $\alpha \equiv \rho_{\sigma_x^2}^{-1}$, $\xi^2 \equiv (1 - \rho^2)\sigma_y^2$ and note that $\epsilon_{xi} = x_t - z_t'\delta$. Then, from eq. (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: $\tilde{x} = (\tilde{x}_1', \ldots, \tilde{x}_n')'$. Thus, we can write $y_t|x_t \sim \mathcal{N}(\tilde{x}_t\theta, \xi^2)$. The conditional likelihood will be:

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

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

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

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

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

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

20
Appendix 2: Results for small samples

This appendix presents an analog result to that presented in section 5.1.2, but for small samples. Tables 5 and 6 show the correspondent results of tables 3 and 4, but for a smaller number of observations, namely $n = 500$.

Table 5: Descriptive Statistics for the estimated betas using different methods (simulation 1).

<table>
<thead>
<tr>
<th>Number of assets</th>
<th>OLS</th>
<th>TLS</th>
<th>BIV-HS</th>
<th>BIV-LASSO</th>
<th>BIV-FSP</th>
</tr>
</thead>
<tbody>
<tr>
<td>min</td>
<td>0.293</td>
<td>0.614</td>
<td>0.688</td>
<td>0.692</td>
<td>0.684</td>
</tr>
<tr>
<td>q25</td>
<td>0.486</td>
<td>0.885</td>
<td>0.897</td>
<td>0.880</td>
<td>0.882</td>
</tr>
<tr>
<td>median</td>
<td>0.577</td>
<td>0.977</td>
<td>0.981</td>
<td>0.964</td>
<td>0.957</td>
</tr>
<tr>
<td>mean</td>
<td>0.559</td>
<td>0.984</td>
<td>0.973</td>
<td>0.958</td>
<td>0.952</td>
</tr>
<tr>
<td>q75</td>
<td>0.628</td>
<td>1.071</td>
<td>1.037</td>
<td>1.015</td>
<td>1.007</td>
</tr>
<tr>
<td>max</td>
<td>0.852</td>
<td>1.379</td>
<td>1.230</td>
<td>1.207</td>
<td>1.197</td>
</tr>
<tr>
<td>min</td>
<td>0.365</td>
<td>0.590</td>
<td>0.637</td>
<td>0.613</td>
<td>0.583</td>
</tr>
<tr>
<td>q25</td>
<td>0.493</td>
<td>0.765</td>
<td>0.791</td>
<td>0.746</td>
<td>0.716</td>
</tr>
<tr>
<td>median</td>
<td>0.551</td>
<td>0.872</td>
<td>0.862</td>
<td>0.819</td>
<td>0.792</td>
</tr>
<tr>
<td>mean</td>
<td>0.549</td>
<td>0.869</td>
<td>0.855</td>
<td>0.812</td>
<td>0.781</td>
</tr>
<tr>
<td>q75</td>
<td>0.606</td>
<td>0.981</td>
<td>0.916</td>
<td>0.864</td>
<td>0.835</td>
</tr>
<tr>
<td>max</td>
<td>0.745</td>
<td>1.173</td>
<td>1.086</td>
<td>1.029</td>
<td>1.004</td>
</tr>
<tr>
<td>min</td>
<td>0.328</td>
<td>0.510</td>
<td>0.523</td>
<td>0.488</td>
<td>0.464</td>
</tr>
<tr>
<td>q25</td>
<td>0.476</td>
<td>0.708</td>
<td>0.694</td>
<td>0.641</td>
<td>0.617</td>
</tr>
<tr>
<td>median</td>
<td>0.537</td>
<td>0.804</td>
<td>0.742</td>
<td>0.685</td>
<td>0.665</td>
</tr>
<tr>
<td>mean</td>
<td>0.546</td>
<td>0.800</td>
<td>0.751</td>
<td>0.694</td>
<td>0.669</td>
</tr>
<tr>
<td>q75</td>
<td>0.613</td>
<td>0.887</td>
<td>0.824</td>
<td>0.763</td>
<td>0.739</td>
</tr>
<tr>
<td>max</td>
<td>0.809</td>
<td>1.083</td>
<td>0.981</td>
<td>0.933</td>
<td>0.899</td>
</tr>
</tbody>
</table>
Table 6: Different Measures for the beta estimation error relative to the Two-Stage Least Square estimator.

<table>
<thead>
<tr>
<th>Number of assets</th>
<th>OLS ME</th>
<th>TSLS ME</th>
<th>BHS ME</th>
<th>BLASSO ME</th>
<th>BFBS ME</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>27.539</td>
<td>1.000</td>
<td>1.655</td>
<td>2.626</td>
<td>2.997</td>
</tr>
<tr>
<td>RMSE</td>
<td>2.737</td>
<td>1.000</td>
<td>0.680</td>
<td>0.702</td>
<td>0.713</td>
</tr>
<tr>
<td>Number of assets</td>
<td>80</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ME</td>
<td>3.440</td>
<td>1.000</td>
<td>1.110</td>
<td>1.433</td>
<td>1.671</td>
</tr>
<tr>
<td>RMSE</td>
<td>2.464</td>
<td>1.000</td>
<td>0.921</td>
<td>1.112</td>
<td>1.267</td>
</tr>
<tr>
<td>Number of assets</td>
<td>140</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ME</td>
<td>2.267</td>
<td>1.000</td>
<td>1.241</td>
<td>1.529</td>
<td>1.652</td>
</tr>
<tr>
<td>RMSE</td>
<td>1.963</td>
<td>1.000</td>
<td>1.118</td>
<td>1.345</td>
<td>1.445</td>
</tr>
</tbody>
</table>

Note: Values less than one indicate that the estimator outperforms the TSLS.