

Sequential Variable Selection as Bayesian Pragmatism in Linear Factor Models

John Knight¹, Stephen Satchell², Jessica Qi Zhang³

¹Department of Economics, University of Western Ontario, Ontario, Canada

²Department of Finance, University of Sydney, Sydney, Australia

³Department of Accounting and Finance, University of Greenwich, London, UK

Email: jknight@uwo.ca, ses999gb@yahoo.co.uk, zq03@gre.ac.uk

Received January 10, 2013; revised February 9, 2013; accepted February 20, 2013

ABSTRACT

We examine a popular practitioner methodology used in the construction of linear factor models whereby particular factors are increased or decreased in relative importance within the model. This allows model builders to customise models and, as such, reflect those factors that the client and modeller may think important. We call this process Pragmatic Bayesianism (or prag-Bayes for short) and we provide analysis which shows when such a procedure is likely to be successful.

Keywords: Linear Factor Models; Bayesian Statistics; Sequential Regression

1. Introduction

The purpose of this paper is to investigate statistical procedures frequently used by practitioners to build factor models. In particular, we are interested in the variable selection methodologies that are used to give a particular returns model a particular style and nature. For example, in the context of global models, one may wish the model to depend more or less upon domestic factors such as country's indices rather than, say, global factors such as currency or world equity and bond markets. Likewise at the domestic level, one may want one's model to be built around styles (value, growth etc.) rather than industries or sectors—alternatively, the opposite may be preferred. The literature on this topic is very sparse. We present a brief survey of alternative approaches. The problem can be viewed as a practical alternative to well-known Bayesian procedures, such as Jorion's (1986) [1] Bayes Stein adjustment and Black-Litterman's BL model (1991, 1992) [2,3]. These models are both examples of Bayesian adjustment which effectively updates currently held opinions with data to form new opinions. Satchell and Scowcroft (2000) [4] also present details of Bayesian portfolio construction procedures based on Black-Litterman models. The essential idea in this process is to have a prior distribution over expected returns or over the regression Betas. In either case, one needs to specify hyperparameters which are, in practice, very troublesome. The procedure we advocate, and which is used by practitioners, is to convert beliefs about the magnitude of betas into procedures of sequential regression.

In Section 2 we shall describe how this is done in practice and how it could be analysed in theory. In Section 3 we shall present conditions under which these methodologies should work. Section 4 presents some empirical results. Conclusions and further discussion are presented in Section 5.

2. Theorem

There are a number of procedures that can be used to facilitate one factor being preferred to another. Here we shall assume that our return series is denoted by the $n \times 1$ vector y , and the two factors over which we may have preferences are denoted by X_1 and X_2 respectively, both $n \times 1$ vectors.

Letting $X = [X_1, X_2]$, we will facilitate calculations later by making the following assumption:

$$X'X = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

Our "true" model is

$$y = X_1\beta_1 + X_2\beta_2 + u \quad (1)$$

where y and u are $n \times 1$ vectors, β_1 and β_2 are scalars and $u \sim N(0, \sigma^2 I_N)$.

This is obviously a simplification of the general case, but little is lost in so doing and it allows us to focus on the essential features of the problem. We now define the sequential variable selection method (SVSM), which is the essential component of the prag-Bayes approach.

Definition: The SVSM is defined by the following

procedure. If you want variable 1 to “explain” more of y asset returns than variable 2, you regress variable 1 first in a univariate regression. The coefficient for variable 2 is then calculated by regressing the residual of y on variable 1 upon the residual of variable 2 on variable 1.

The question we wish to ask is: under what circumstances will this procedure lead to a larger estimated exposure ($\hat{\beta}_1$) of variable 1 versus that of variable 2, ($\hat{\beta}_2$). A closely related question is the conditions under which the new slope estimates will be bigger or smaller than those calculated from conventional ordinary least squares (OLS).

It is worth discussing a variant on these procedures which concerns testing. Rather than just focusing on the magnitude of $\hat{\beta}$: we could also alternatively make inclusion and exclusion decisions based on t-statistics. Our results can be tilted in the desired direction by moving the critical values of our tests.

In terms of the Equation (1), we do not wish to impose $\beta_1 > \beta_2$ for all stocks. This is because we recognise that particular stocks may not be modeled subject to such a constraint. To illustrate, in the case of factor 1 being a global factor and factor 2 being a domestic factor, we can imagine cases of multinationals where $\beta_1 > \beta_2$ but there will also be Japanese railway stocks, for example, where the opposite is true. Accordingly, a Bayesian approach where β_1 and β_2 are variable allows us to approach this question in a theoretically appealing way.

We may have a prior, that $P(\beta_1 \geq \beta_2) \geq d$ where d is some threshold probability, and $P()$ denotes the probability of the event in brackets. This can be easily imposed by an adroit choice of hyper-parameters in the prior joint distribution of β_1 and β_2 . Then we can compute the likelihood in the usual way, and finally, the posterior distribution of β_1 and β_2 where the posterior probability of $\beta_1 \geq \beta_2$ can be computed in a straightforward manner. However, implementation of hierarchical Bayes models required a number of ancillary assumptions that are not particularly transparent, see Gelman (2004) [5] for example. We shall not detail how a Bayesian might proceed, but return to our SVSM method to see if it can achieve similar results and now address the second question as to whether the SVSM method will increase the magnitude, relative to OLS, of estimated β_1 .

With the above model we now consider the two estimators of β_1

1) $\hat{\beta}_1$ from $y = X_1\beta_1 + \omega$ where $\omega = X_2\beta_2 + u$

$$\begin{aligned} \hat{\beta}_1 &= (X_1'X_1)^{-1} X_1'y \\ &= \beta_1 + (X_1'X_1)^{-1} X_1'(X_2\beta_2 + u) \end{aligned}$$

2) $\tilde{\beta}_1$ from $y = X_1\beta_1 + X_2\beta_2 + u$ i.e.

$$\tilde{\beta}_1 = (X_1'\bar{P}_{x_2} X_1)^{-1} X_1'\bar{P}_{x_2} y,$$

where $\bar{P}_{x_2} = I - X_2(X_2'X_2)^{-1} X_2'$

With the assumption on $(X'X)$ we have immediately that $\hat{\beta}_1 = X_1'y$

$$\tilde{\beta}_1 = (1 - \rho^2)^{-1} (X_1'y - \rho X_2'y)$$

and since $u \sim N(0, \sigma^2 I)$

This implies $y \sim N(X\beta, \sigma^2 I)$

And $\begin{pmatrix} \hat{\beta}_1 \\ \tilde{\beta}_1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ (1 - \rho^2)^{-1} & -\rho(1 - \rho^2)^{-1} \end{pmatrix} \begin{pmatrix} X_1'y \\ X_2'y \end{pmatrix}$
 $\sim N(u, \sigma^2 \Sigma)$

Where $\mu = \begin{pmatrix} \beta_1 + \rho\beta_2 \\ \beta_1 \end{pmatrix}$ and $\Sigma = \begin{pmatrix} 1 & 1 \\ 1 & (1 - \rho^2)^{-1} \end{pmatrix}$

We now calculate the following probability illustrated in the following diagram **Figure 1**, where the horizontal axis gives values of $\hat{\beta}_1$ while the vertical gives values of $\tilde{\beta}_1$.

$$\begin{aligned} P(|\hat{\beta}_1| > |\tilde{\beta}_1|) &= P(\hat{\beta}_1 > \tilde{\beta}_1, \hat{\beta}_1 > 0, \tilde{\beta}_1 > 0) \\ &\quad + P(\hat{\beta}_1 > -\tilde{\beta}_1, \hat{\beta}_1 > 0, \tilde{\beta}_1 < 0) \\ &\quad + P(\hat{\beta}_1 < -\tilde{\beta}_1, \hat{\beta}_1 < 0, \tilde{\beta}_1 > 0) \\ &\quad + P(\hat{\beta}_1 < \tilde{\beta}_1, \hat{\beta}_1 < 0, \tilde{\beta}_1 < 0) \end{aligned}$$

The result is stated in the following Theorem.

Theorem

Under the SVSM estimation procedure we have the following probability:

When $\rho > 0$

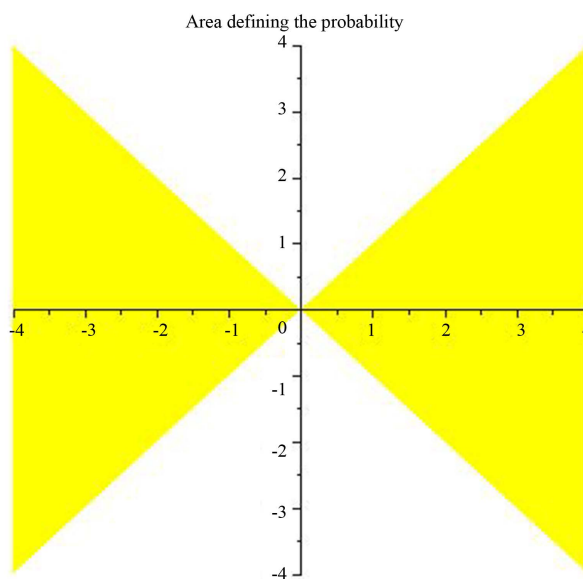


Figure 1. Area defining the probability.

$$P\left(\left|\hat{\beta}_1\right| > \left|\tilde{\beta}_1\right|\right) = \int_0^{\frac{\rho\beta_2}{2}} g(r) f_1(r) dr + \int_{\frac{\rho\beta_2}{2}}^{\infty} g(r) f_1(r) dr + \int_{-\infty}^0 g(r) f_1(r) dr$$

For $\rho < 0$

$$P\left(\left|\hat{\beta}_1\right| > \left|\tilde{\beta}_1\right|\right) = \int_0^{\infty} h(r) f_1(r) dr + \int_{-\infty}^{\frac{-\rho\beta_2}{2}} h(r) f_1(r) dr + \int_{\frac{-\rho\beta_2}{2}}^0 h(r) f_1(r) dr$$

where $g(r) = \Phi\left(\frac{\beta_2}{\sigma}\sqrt{1-\rho^2}\right) - \Phi\left(-\frac{(2r-\rho\beta_2)\sqrt{1-\rho^2}}{\rho\sigma}\right)$, $h(r) = \Phi\left(\frac{(2r+\rho\beta_2)\sqrt{1-\rho^2}}{\sigma\rho}\right) - \Phi\left(\frac{\beta_2\sqrt{1-\rho^2}}{\sigma}\right)$
 $f_1(r) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(r-(\beta_1+\rho\beta_2))^2\right)$ and $f_2(r) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(r-(\beta_1-\rho\beta_2))^2\right)$

Proof: See Appendix

3. Statistical Analysis

To illustrate our calculations, we carried out some numerical calculations; we calculated the probability that $\hat{\beta}_1$ exceeds $\tilde{\beta}_1$ for different values of σ and ρ ; we also computed the R^2 of the regression. The values of σ were 0.1, 0.2, 0.5, 1.0, and 2.0 whilst the values of ρ were -0.8, -0.5, -0.2, 0.2, 0.5 and 0.8. Different combinations of $\hat{\beta}_1$ and $\tilde{\beta}_1$ were used, namely (0.8, 0.4), (0.5, 0.4), (0.4, 0.4), (0.3, 0.4), and (0.1, 0.4). The output constitutes **Tables 1-5**.

Table 1. Probability and R-squared for $\hat{\beta}_1 = 0.80$ and $\tilde{\beta}_1 = 0.80$; (a) Probability for $\hat{\beta}_1 = 0.80$ and $\tilde{\beta}_1 = 0.40$; (b) R-squared for $\hat{\beta}_1 = 0.80$ and $\tilde{\beta}_1 = 0.40$.

		(a)					
		ρ					
		-0.8	-0.5	-0.2	0.2	0.5	0.8
σ	0.1	0.008198	0.000266	0.000044	0.999956	0.999734	0.991802
	0.2	0.117701	0.042598	0.025148	0.974964	0.958360	0.884898
	0.5	0.279114	0.267746	0.248766	0.751119	0.721594	0.632659
	1.0	0.305628	0.367917	0.395136	0.570189	0.528138	0.436008
	2.0	0.311146	0.399377	0.448025	0.496190	0.443837	0.347957
		(b)					
		ρ					
		-0.8	-0.5	-0.2	0.2	0.5	0.8
σ	0.1	0.966443	0.979592	0.985337	0.989339	0.991150	0.992436
	0.2	0.878049	0.923077	0.943820	0.958678	0.965517	0.970414
	0.5	0.535316	0.657534	0.728850	0.787776	0.817518	0.839949
	1.0	0.223602	0.324324	0.401914	0.481328	0.528302	0.567474
	2.0	0.067164	0.107143	0.143836	0.188312	0.218750	0.246988

The results show that if the regression was a high R^2 and if the two variables are positively correlated, then this procedure leads to a high probability that $\hat{\beta}_1$ exceeds $\tilde{\beta}_1$ not just when $\hat{\beta}_1$ exceeds $\tilde{\beta}_1$, but even when $\hat{\beta}_1$ is less than $\tilde{\beta}_1$ (see **Tables 3-5**). In the case when R^2 is low or when the returns are negatively correlated, the methodology is less successful.

4. Empirical Examples

For illustrative purposes, we use six Fama-French style based portfolios formed on size and book-to-market¹. These are: Small Growth (SG), Small Neutral (SN), Small Value (SV), Big Growth (BG), Big Neutral (BN), Big Value (BV). There are two return factors, the first, SMB (Small Minus Big) is the return difference between the average of three small portfolios and the average of three large portfolios, Likewise, the second factor, HML (High Minus Low) is the return difference between the average of two value portfolios and the average of two growth portfolios.

We choose two different sample periods, where SMB and HML are either positively or negatively correlated. **Table 6** lists the regression results for the period from 1935 Jan to 1954 Dec, where SMB and HML are positively correlated with $\rho = 0.529$; **Table 7** lists the regression results for the period from 1992 Jan to 2011 Dec with $\rho = -0.348$. In our sequential variable selection model, SMB is variable 1 and HML is variable 2. $\hat{\beta}_1$ is the estimated coefficient from the univariate regression

¹The stocks are ranked based on two independent criteria: size (market capitalization) and book-to-price (the ratio of book value to market value). The median NYSE market equity is chosen to divide the stocks into two groups: big and small; the 30th and 70th percentiles of book-to-price ratio are used to split the stocks into three groups: growth, neutral and value. Six portfolios are formed from the intersection of these independent sorts. Six portfolios are formed from the intersection of these independent sorts.

Table 2. Probability and R-squared for $\hat{\beta}_1 = 0.50$ and $\tilde{\beta}_1 = 0.40$; (a) Probability for $\hat{\beta}_1 = 0.50$ and $\tilde{\beta}_1 = 0.40$ (b) R-squared for $\hat{\beta}_1 = 0.50$ and $\tilde{\beta}_1 = 0.40$.

		(a)					
		ρ					
		-0.8	-0.5	-0.2	0.2	0.5	0.8
	0.1	0.041297	0.001569	0.000057	0.999956	0.999734	0.991802
	0.2	0.196088	0.089434	0.039651	0.971446	0.956407	0.881919
σ	0.5	0.292002	0.318466	0.308932	0.690551	0.660531	0.563587
	1.0	0.307585	0.385128	0.421232	0.536673	0.490127	0.393426
	2.0	0.311519	0.404026	0.455544	0.485904	0.431795	0.334440
		(b)					
		ρ					
		-0.8	-0.5	-0.2	0.2	0.5	0.8
	0.1	0.900000	0.954545	0.970588	0.980000	0.983871	0.986486
	0.2	0.692308	0.840000	0.891892	0.924528	0.938462	0.948052
σ	0.5	0.264706	0.456522	0.568966	0.662162	0.709302	0.744898
	1.0	0.082569	0.173554	0.248120	0.328859	0.378882	0.421965
	2.0	0.022005	0.049881	0.076212	0.109131	0.132321	0.154334

Table 3. Probability and R-squared for $\tilde{\beta}_1 = 0.40$ and $\tilde{\beta}_1 = 0.40$; (a) Probability for $\tilde{\beta}_1 = 0.40$ and $\tilde{\beta}_1 = 0.40$; (b) R-squared for $\tilde{\beta}_1 = 0.40$ and $\tilde{\beta}_1 = 0.40$.

		(a)					
		ρ					
		-0.8	-0.5	-0.2	0.2	0.5	0.8
	0.1	0.196464	0.021743	0.000680	0.999950	0.999733	0.991801
	0.2	0.273772	0.156281	0.070600	0.961037	0.950293	0.875036
σ	0.5	0.305567	0.345069	0.339541	0.659712	0.630060	0.532287
	1.0	0.310969	0.392804	0.431375	0.524599	0.477063	0.379741
	2.0	0.312364	0.406016	0.458267	0.482522	0.428049	0.330487
		(b)					
		ρ					
		-0.8	-0.5	-0.2	0.2	0.5	0.8
	0.1	0.864865	0.941176	0.962406	0.974619	0.979592	0.982935
	0.2	0.615385	0.800000	0.864865	0.905660	0.923077	0.935065
σ	0.5	0.203822	0.390244	0.505929	0.605678	0.657534	0.697337
	1.0	0.060150	0.137931	0.203822	0.277457	0.324324	0.365482
	2.0	0.015748	0.038462	0.060150	0.087591	0.107143	0.125874

Table 4. Probability and R-squared for $\hat{\beta}_1 = 0.30$ and $\tilde{\beta}_1 = 0.40$; (a) Probability for $\hat{\beta}_1 = 0.30$ and $\tilde{\beta}_1 = 0.40$; (b) R-squared for $\hat{\beta}_1 = 0.30$ and $\tilde{\beta}_1 = 0.40$.

		(a)					
		ρ					
		-0.8	-0.5	-0.2	0.2	0.5	0.8
	0.1	0.490914	0.145394	0.012818	0.999596	0.999673	0.991738
	0.2	0.376203	0.268794	0.139910	0.930899	0.931419	0.857201
σ	0.5	0.324386	0.376531	0.375308	0.623671	0.594845	0.497748
	1.0	0.315780	0.401432	0.442149	0.512258	0.464063	0.366637
	2.0	0.313574	0.408225	0.461091	0.479180	0.424459	0.326839
		(b)					
		ρ					
		-0.8	-0.5	-0.2	0.2	0.5	0.8
	0.1	0.852941	0.928571	0.952830	0.967532	0.973684	0.977876
	0.2	0.591837	0.764706	0.834711	0.881657	0.902439	0.917012
σ	0.5	0.188312	0.342105	0.446903	0.543796	0.596774	0.638728
	1.0	0.054820	0.115044	0.168053	0.229584	0.270073	0.306519
	2.0	0.014293	0.031477	0.048072	0.069335	0.084668	0.099505

Table 5. Probability and R-squared for $\hat{\beta}_1 = 0.10$ and $\tilde{\beta}_1 = 0.40$; (a) Probability for $\hat{\beta}_1 = 0.10$ and $\tilde{\beta}_1 = 0.40$; (b) R-squared for $\hat{\beta}_1 = 0.10$ and $\tilde{\beta}_1 = 0.40$.

		(a)					
		ρ					
		-0.8	-0.5	-0.2	0.2	0.5	0.8
	0.1	0.815601	0.751707	0.393672	0.918113	0.972403	0.976545
	0.2	0.599814	0.586064	0.431205	0.740597	0.794515	0.747375
σ	0.5	0.376591	0.451091	0.458812	0.539563	0.514171	0.423813
	1.0	0.329611	0.421336	0.465250	0.487285	0.438907	0.342919
	2.0	0.317082	0.413286	0.467019	0.472654	0.417802	0.320514
		(b)					
		ρ					
		-0.8	-0.5	-0.2	0.2	0.5	0.8
	0.1	0.913793	0.928571	0.939024	0.948980	0.954545	0.959016
	0.2	0.726027	0.764706	0.793814	0.823009	0.840000	0.854015
σ	0.5	0.297753	0.342105	0.381188	0.426606	0.456522	0.483471
	1.0	0.095841	0.115044	0.133449	0.156830	0.173554	0.189627
	2.0	0.025816	0.031477	0.037073	0.044434	0.049881	0.055267

Table 6. Regression results for six portfolios when $\rho = 0.529$.

	SG	SN	SV	BG	BN	BV
$\hat{\beta}_1$	1.8919	1.8433	2.426	0.76693	0.89019	1.5041
R^2_{indi}	0.57378	0.54156	0.62007	0.19987	0.25477	0.31972
$\tilde{\beta}_1$	1.6684	1.4202	1.6543	0.61075	0.50736	0.6249
$\hat{\beta}_2$	0.35162	0.66562	1.2141	0.24572	0.60229	1.3832
R^2_{multi}	0.59437	0.61492	0.78141	0.22119	0.37594	0.60065

Table 7. Regression results for six portfolios when $\rho = -0.348$.

	SG	SN	SV	BG	BN	BV
$\hat{\beta}_1$	1.4183	0.93986	0.90215	0.18208	0.047672	0.030638
R^2_{indi}	0.50081	0.39152	0.32264	0.02081	0.001394	0.000484
$\tilde{\beta}_1$	1.2584	1.0111	1.0787	0.027752	0.11295	0.20748
$\hat{\beta}_2$	-0.47902	0.21332	0.5288	-0.46232	0.19554	0.52974
R^2_{multi}	0.5471	0.40787	0.41246	0.12952	0.020397	0.11761

of y on SMB; $\tilde{\beta}_1$ is the coefficient on SMB from the multiple regression of y on SMB and HML; $\hat{\beta}_2$ is the coefficient on HML and calculated by regressing the residual of y on SMB upon the residual of HML on SMB.

We are interested in the following question. Under what circumstances will there be a larger estimated exposure $\hat{\beta}_1$ than $\tilde{\beta}_1$? The results show that when the two variables are positively correlated as in **Table 6**, this procedure always generates higher $\hat{\beta}_1$ than $\tilde{\beta}_1$. When the two variables are negatively correlated as in **Table 7**, we identify higher $\hat{\beta}_1$ than $\tilde{\beta}_1$ only for two portfolios SG and BG; for the other four portfolios, $\hat{\beta}_1$ is lower than $\tilde{\beta}_1$. Therefore comparing the two different cases, we find out that the methodology is more successful when ρ is positively correlated. This confirms our finding in section 3.

5. Conclusions

Bayesian methods are notoriously difficult to implement and practitioners often use tricks to allow their models to

reflect their beliefs. We discuss such a procedure, and show analytically conditions when it will work. The particular procedure we discuss is used by practitioners to build factor models. We are interested in the variable selection methodologies that are used to give a particular returns model a particular style and nature. For example, in the context of global models one may wish the model to depend more/less upon domestic factors such as country indices rather than, say, global factors such as currency or world equity/bond markets. The method we discuss allows for favorable selection of a variable by specifying the order in which variables enter a regression.

We strip the problem down to its bare essentials by considering bivariate situations. We evaluate these conditions using numerical integration and further confirm their relevance by looking at an empirical example. The examples used US equity data over 20 years period. These illustrate the efficacy of the procedure.

REFERENCES

- [1] P. Jorion, "Bayes-Stein Estimation for Portfolio Analysis," *Journal of Financial and Quantitative Analysis*, Vol. 21, No. 3, 1986, pp. 279-292. [doi:10.2307/2331042](https://doi.org/10.2307/2331042)
- [2] F. Black and R. Litterman, "Global Asset Allocation with Equities, Bonds and Currencies," Goldman Sachs and Co., New York, 1991. https://faculty.fuqua.duke.edu/~charvey/Teaching/BA453_2006/Black_Litterman_GAA_1991.pdf
- [3] F. Black, and R. Litterman, "Global Portfolio Optimization," *Financial Analysts Journal*, Vol. 48, No. 5, 1992, pp. 28-43. [doi:10.2469/faj.v48.n5.28](https://doi.org/10.2469/faj.v48.n5.28)
- [4] S. Satchell and A. Scowcroft, "A Demystification of the Black-Litterman Model: Managing Quantitative and Traditional Portfolio Construction," *Journal of Asset Management*, Vol. 1, No. 2, 2000, pp. 138-150. [doi:10.1057/palgrave.jam.2240011](https://doi.org/10.1057/palgrave.jam.2240011)
- [5] A. Gelman, J. Carlin, H. Stern and D. Rubin, "Bayesian Data Analysis," 2nd Edition, Chapter 5, Chapman & Hall/CRC, London, 2004.

Appendix: Proof of Theorem

Diagrammatically we need to calculate the two areas in **Figure 1** on either side of the origin.

$$RHS(\text{diagram}) = \int_0^r \int_{-r}^r f(r, s) ds dr$$

$$LHS(\text{diagram}) = \int_{-\infty}^0 \int_{-r}^r f(r, s) ds dr$$

Now

$$f(r, s) = f(s|r)f_1(r)$$

where

$$s|r \sim N\left(r - \rho\beta_2, \frac{\rho^2\sigma^2}{(1-\rho^2)}\right)$$

$$\text{and } r \sim N(\beta_1 + \rho\beta_2, \sigma^2)$$

First we shall calculate *RHS*.

$$RHS = \int_0^{\rho\beta_2/2} \left(\int_{\frac{-(2r-\rho\beta_2)\sqrt{1-\rho^2}}{\rho\sigma}}^{\frac{\beta_2\sqrt{1-\rho^2}}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\omega^2/2} d\omega \right) f_1(r) dr + \int_{\rho\beta_2/2}^{\infty} \left(\int_{\frac{-(2r-\rho\beta_2)\sqrt{1-\rho^2}}{\rho\sigma}}^{\frac{\beta_2\sqrt{1-\rho^2}}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\omega^2/2} d\omega \right) f_1(r) dr$$

$$\text{Letting } g(r) = \int_{\frac{-(2r-\rho\beta_2)\sqrt{1-\rho^2}}{\rho\sigma}}^{\frac{\beta_2\sqrt{1-\rho^2}}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\omega^2/2} d\omega = \Phi\left(\frac{\beta_2}{\sigma}\sqrt{1-\rho^2}\right) - \Phi\left(\frac{-(2r-\rho\beta_2)\sqrt{1-\rho^2}}{\rho\sigma}\right)$$

where $\Phi(x)$ is the cumulative distribution function of the standard normal distribution we have:

$$RHS = \int_0^{\frac{\rho\beta_2}{2}} g(r)f_1(r)dr + \int_{\frac{\rho\beta_2}{2}}^{\infty} g(r)f_1(r)dr$$

Having completed the calculation of *RHS*, we now turn to *LHS*.

$$LHS = \int_{-\infty}^0 \int_{-r}^r f(r, s) ds dr = \int_{-\infty}^0 g(r)f_1(r)dr$$

We can make further simplifications depending upon the sign of ρ .

For $\rho > 0$

$$P(|\hat{\beta}_1| > |\tilde{\beta}_1|) = \int_0^{\frac{\rho\beta_2}{2}} g(r)f_1(r)dr + \int_{\frac{\rho\beta_2}{2}}^{\infty} g(r)f_1(r)dr + \int_{-\infty}^0 g(r)f_1(r)dr$$

where

$$RHS = \int_0^{\infty} \left(\int_{-r}^r f(s|r) ds \right) f_1(r) dr$$

Transforming from s to ω ,

$$s \rightarrow \omega = (s - (r - \rho\beta_2)) \frac{\sqrt{1-\rho^2}}{\rho\sigma}$$

$$\text{we have } ds = \frac{\rho\sigma d\omega}{\sqrt{1-\rho^2}}, \text{ giving}$$

$$RHS = \int_0^{\infty} \left(\int_{\frac{-(2r-\rho\beta_2)\sqrt{1-\rho^2}}{\rho\sigma}}^{\frac{\beta_2\sqrt{1-\rho^2}}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\omega^2/2} d\omega \right) f_1(r) dr$$

For

$$0 < r \leq \rho\beta_2/2 \Rightarrow 0 < -(2r - \rho\beta_2) \frac{\sqrt{1-\rho^2}}{\rho\sigma} < \frac{\beta_2\sqrt{1-\rho^2}}{\sigma}$$

$$f_1(r) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(r - (\beta_1 + \rho\beta_2))^2\right)$$

When $\rho < 0$ rewrite using $-\rho$ and then let $\rho > 0$. Thus we now have:

$$\begin{pmatrix} r \\ s \end{pmatrix} \sim N\left(\begin{pmatrix} \beta_1 + \rho\beta_2 \\ \beta_1 \end{pmatrix}, \sigma^2 \begin{pmatrix} 1 & 1 \\ 1 & (1-\rho^2)^{-1} \end{pmatrix}\right)$$

$$s|r \sim N\left(r + \rho\beta_2, \frac{\rho^2\sigma^2}{(1-\rho^2)}\right)$$

$$r \sim N(\beta_1 - \rho\beta_2, \sigma^2)$$

Now,

$$RHS = \int_0^{\infty} \int_{-r}^r f(s|r)f_2(r) ds dr$$

and again transforming from s to ω ,

$$s \rightarrow \omega = (s - r - \rho\beta_2) \frac{\sqrt{1-\rho^2}}{\sigma\rho} \text{ with } ds = \frac{\sigma\rho d\omega}{\sqrt{1-\rho^2}}$$

$$RHS = \int_0^\infty \left(\int_{\frac{-\beta_2\sqrt{1-\rho^2}}{\sigma\rho}}^{\frac{-\beta_2\sqrt{1-\rho^2}}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\omega^2/2} d\omega \right) f_2(r) dr$$

Letting

$$\begin{aligned} h(r) &= \int_{\frac{-(2r+\rho\beta_2)\sqrt{1-\rho^2}}{\sigma\rho}}^{\frac{-\beta_2\sqrt{1-\rho^2}}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\omega^2/2} d\omega = \Phi\left(\frac{-\beta_2\sqrt{1-\rho^2}}{\sigma}\right) - \Phi\left(\frac{-(2r+\rho\beta_2)\sqrt{1-\rho^2}}{\sigma\rho}\right) \\ &= \Phi\left(\frac{(2r+\rho\beta_2)\sqrt{1-\rho^2}}{\sigma\rho}\right) - \Phi\left(\frac{\beta_2\sqrt{1-\rho^2}}{\sigma}\right) \end{aligned}$$

$$RHS = \int_0^\infty h(r) f_2(r) dr$$

For $-\infty < r < -\rho\beta_2/2 \Rightarrow -\infty < \frac{(2r+\rho\beta_2)\sqrt{1-\rho^2}}{\rho\sigma} < 0$

and

$$\begin{aligned} LHS &= \int_{-\infty}^0 h(r) f_2(r) dr \\ &= \int_{-\infty}^{-\frac{\rho\beta_2}{2}} h(r) f_2(r) dr + \int_{-\frac{\rho\beta_2}{2}}^0 h(r) f_2(r) dr \end{aligned}$$

Thus for $\rho < 0$.

$$\begin{aligned} P\left(\left|\hat{\beta}_1\right| > \left|\hat{\beta}_1\right|\right) &= \int_0^\infty h(r) f_2(r) dr \\ &+ \int_{-\infty}^{-\frac{\rho\beta_2}{2}} h(r) f_2(r) dr + \int_{-\frac{\rho\beta_2}{2}}^0 h(r) f_2(r) dr \end{aligned}$$

where

$$f_2(r) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} (r - (\beta_1 - \rho\beta_2))^2\right).$$