Additional Appendix to

Consistent Noisy Independent Component Analysis

Stéphane Bonhomme[1]
CEMFI, Madrid[2]

Jean-Marc Robin
Paris School of Economics,
Université Paris 1 Panthéon Sorbonne,[3]
and University College London[4]

Revised version: December 2008

[1]**Corresponding author:** CEMFI, Casado del Alisal, 5, 28014 Madrid, Spain. E-mail: bonhomme@cemfi.es

[2]Stéphane Bonhomme gratefully acknowledges the financial support from the Spanish Ministry of Science and Innovation through the Consolider-Ingenio 2010 Project "Consolidating Economics", and from the Spanish MEC, Grant SEJ2005-08880.

[3]Centre d'Economie de la Sorbonne, Université Paris 1 Panthéon Sorbonne, 106/112 bd de l'Hôpital, 75647 Paris Cedex 13, e-mail: jmrobin@univ-paris1.fr.

[4]Jean-Marc Robin gratefully acknowledges the financial support from the Economic and Social Research Council for the ESRC Centre for Microdata Methods and Practice, "Cemmap" (grant reference RES-589-28-0001).

1 Outline

This appendix contains four sections. In Section 2, we present some additional Monte-Carlo simulations which we refer to in the main paper (Bonhomme and Robin, 2008a).[1] In Section 3 we develop a method to estimate the number of factors K. Lastly, in Sections 4 and 5 we present a detailed analysis of the two applications, to test scores data and to financial data on stock returns.

2 Additional Monte Carlo simulations

We start in Table 1 by modifying the kurtosis of error variables while keeping factor variables log-normal. The design is the one of Table 2 in the main paper, and the sample size is $N = 1000$. Relative to the case with normal errors,[2] estimates in cases where errors are non-normal are very similar. If anything, slightly higher standard errors are obtained when errors follow normal mixtures with a high excess kurtosis.

Next, we set $K < L$ and compare quasi-JADE based on second, third and fourth-order moments (using the restrictions of Theorem 2 in the main paper) to quasi-JADE based on second and third-order moments only (using the restrictions of Theorem 3), which yields consistent estimates when all factors are skewed. Table 2 reports simulations with log-normal factors, standard normal errors with variance 1, and matrix $\mathbf{\Lambda}$ is equal to

$$\mathbf{\Lambda}_3 \equiv \begin{pmatrix} 2 & 2 \\ 2 & 1 \\ 1 & 2 \end{pmatrix}.$$

Table 2 shows that the standard deviations of factor loadings estimates do not necessarily decrease when adding fourth-order moments. This is because, in our design, the (finite-sample) imprecision of kurtosis estimates may dominate the (asymptotic) efficiency gains of using more moments. This illustrative table suggests that an algorithm based on third-order moments only, and relying on orthogonality up to the third order, is likely to do well in practice, provided that there is enough factor skewness.

Next, we consider a case with correlated errors. Table 3 displays means and standard deviations of the Monte Carlo distributions of factor loadings estimates obtained from 1000 simulations of samples of size $N = 1000$ generated by standardized log-normal factors, standard normal errors and $\mathbf{\Lambda}$ equal to

$$\mathbf{\Lambda}_4 \equiv \begin{pmatrix} 2 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 \\ 1 & 1 & 2 & 1 \\ 1 & 1 & 1 & 2 \end{pmatrix}.$$

We only allow errors U_3 and U_4 to have non-zero correlation ρ. In the simulation we let ρ vary between 0 and .9. When ρ increases, the performances of the algorithms

[1] Available at: http://www.cemfi.es/~bonhomme/

[2] Note that, as in the previous experiments, we do not impose that the excess kurtosis of error variables is zero in the estimation, even when errors are normally distributed.

Table 1: Non-normal errors

κ_4	-6/5 (uniform)	0 (normal)	5 (normal mixtures)	10 (normal mixtures)	100 (normal mixtures)	≈110 (log-normal)
λ_{11}	2.03 (.17)	2.03 (.17)	2.04 (.17)	2.03 (.19)	2.03 (.19)	2.01 (.19)
λ_{21}	.98 (.13)	.99 (.14)	.99 (.15)	.98 (.14)	.99 (.21)	.98 (.16)
λ_{31}	.99 (.15)	.99 (.15)	.98 (.14)	.98 (.14)	.99 (.21)	.98 (.16)
Var(U_1)	.87 (.43)	.87 (.43)	.85 (.42)	.85 (.44)	.85 (.47)	.89 (.44)

Note: factors are log-normal, errors follow normal mixtures with unitary variance, $\mathbf{\Lambda} = \mathbf{\Lambda}_1$, $N = 1000$. Bootstrapped standard errors in parentheses.

Table 2: Efficiency gains from using fourth order moments

N	500	500	1000	1000	5000	5000
Cumulants	2,3,4	2,3	2,3,4	2,3	2,3,4	2,3
λ_{11}	1.95 (.28)	1.93 (.32)	1.98 (.19)	1.97 (.24)	2.00 (.08)	2.00 (.08)
λ_{21}	1.96 (.30)	1.91 (.37)	1.99 (.16)	1.96 (.23)	1.00 (.09)	2.00 (.05)
λ_{31}	.97 (.23)	.98 (.25)	.98 (.17)	.98 (.20)	1.00 (.08)	1.00 (.08)
Var(U_1)	.98 (.21)	1.01 (.16)	.98 (.15)	1.00 (.13)	.97 (.09)	1.00 (.06)

Note: log-normal factors, standard normal errors, $\mathbf{\Lambda} = \mathbf{\Lambda}_3$. Bootstrapped standard errors in parentheses.

Table 3: JADE and Quasi-JADE for various correlation parameter ρ

JADE				
ρ	0	.2	.5	.9
λ_{41}	.89 (.42)	.87 (.44)	.84 (.45)	.75 (.50)
λ_{44}	2.25 (.29)	2.22 (.32)	2.23 (.28)	2.20 (.24)

Quasi-JADE, independent errors				
ρ	0	.2	.5	.9
λ_{41}	.98 (.17)	.95 (.19)	.90 (.21)	.86 (.28)
λ_{44}	2.05 (.20)	2.08 (.19)	2.15 (.16)	2.15 (.23)
Var(U_4)	.81 (.42)	.66 (.40)	.37 (.33)	.12 (.20)

Quasi-JADE, correlation allowed between U_3 and U_4				
ρ	0	.2	.5	.9
λ_{41}	.98 (.19)	.99 (.17)	.99 (.17)	.98 (.16)
λ_{44}	2.03 (.21)	2.03 (.23)	2.03 (.22)	2.05 (.22)
Var(U_4)	.84 (.54)	.82 (.55)	.81 (.55)	.79 (.56)
Cov(U_3, U_4)	-.002 (.22)	.20 (.23)	.49 (.24)	.88 (.25)

Note: log-normal factors, standard normal errors, $\mathbf{\Lambda} = \mathbf{\Lambda}_4$, $N = 1000$. Bootstrapped standard errors in parentheses.

assuming no or independent errors deteriorate. By comparison, quasi-JADE, with the right structure of error dependences, shows slightly larger standard errors when $\rho = 0$, which is consistent with the fact that it uses fewer moment conditions to estimate error moments. When ρ increases, it turns out to be remarkably robust. In all cases the performance of noise-free JADE is much worse.

In our last experiment, we simulated an overcomplete ICA model with $L = 4$ and $K = 6$, with four restrictions on $\mathbf{\Lambda}$:

$$\mathbf{\Lambda}_5 \equiv \begin{pmatrix} 1 & 2 & 1 & 1 & 1 & 0 \\ 1 & 1 & 2 & 1 & 1 & 0 \\ 0 & 1 & 1 & 2 & 1 & 1 \\ 0 & 1 & 1 & 1 & 2 & 1 \end{pmatrix}.$$

We simulated the model 1000 times, with sample size $N = 1000$, standardized log-normal factors and standard normal errors. We obtained the following estimates (standard errors in parentheses):

$$\widehat{\mathbf{\Lambda}}_5 \equiv \begin{pmatrix} \begin{matrix}1.01\\(.38)\end{matrix} & \begin{matrix}1.93\\(.45)\end{matrix} & \begin{matrix}.93\\(.28)\end{matrix} & \begin{matrix}.99\\(.31)\end{matrix} & \begin{matrix}1.01\\(.33)\end{matrix} & 0 \\ \begin{matrix}1.00\\(.38)\end{matrix} & \begin{matrix}.93\\(.31)\end{matrix} & \begin{matrix}1.94\\(.40)\end{matrix} & \begin{matrix}.98\\(.30)\end{matrix} & \begin{matrix}1.00\\(.31)\end{matrix} & 0 \\ 0 & \begin{matrix}.97\\(.33)\end{matrix} & \begin{matrix}.98\\(.32)\end{matrix} & \begin{matrix}1.97\\(.41)\end{matrix} & \begin{matrix}.97\\(.29)\end{matrix} & \begin{matrix}.94\\(.39)\end{matrix} \\ 0 & \begin{matrix}.96\\(.33)\end{matrix} & \begin{matrix}.98\\(.30)\end{matrix} & \begin{matrix}.95\\(.29)\end{matrix} & \begin{matrix}1.98\\(.41)\end{matrix} & \begin{matrix}.95\\(.39)\end{matrix} \end{pmatrix}.$$

Finite sample biases are somewhat larger than in the case $K \leq L$. Nevertheless, this result shows that quasi-JADE can be used to estimate restricted overcomplete ICA models, even when the sample size is only moderately large.

3 Estimation of the number of factors

3.1 Estimating the number of factors K

All factors are kurtotic. Assuming that $\mathbf{Q}_{\mathcal{J}}$ has full column rank and that factor variables show excess kurtosis, then matrix $\mathbf{\Omega}_{\mathbf{Y}}$ has rank $K \leq J$ (see Theorem 1 in the main paper). We use the sequential testing procedure developed by Robin and Smith (2000) to estimate the rank of $\mathbf{\Omega}_{\mathbf{Y}}$ (see 3.2 for a description of the test).

Monte Carlo simulations show that the rank test, applied to matrix $\mathbf{\Omega}_{\mathbf{Y}}$ alone, suffers from substantial size distortions (see the simulations in the next section). Assuming $K \leq L$, the factor structure provides additional rank conditions that can be used to improve the test's properties. We propose the following refinement.

Consider matrices $\mathbf{\Omega}_{\mathbf{Y}}(\ell, m)$ with $(\ell, m) \in \mathcal{J}$. They satisfy the restrictions:

$$\mathbf{\Omega}_{\mathbf{Y}}(\ell, m) = \mathbf{\Lambda}\mathbf{D}_4 \operatorname{diag}(\mathbf{\Lambda}_\ell \odot \mathbf{\Lambda}_m)\mathbf{\Lambda}^{\mathrm{T}}.$$

Let $\mathbf{w} = (w_{\ell m}, (\ell, m) \in \mathcal{J})$ be a vector of J positive weights. Then,

$$\boldsymbol{\Omega}_{\mathbf{Y},\mathbf{w}} \equiv \sum_{(\ell,m)\in\mathcal{J}} w_{\ell,m} \boldsymbol{\Omega}_{\mathbf{Y}}(\ell, m) = \boldsymbol{\Lambda} \mathbf{D}_4 \operatorname{diag}\left(\mathbf{Q}_{\mathcal{J}}{}^{\mathrm{T}} \mathbf{w}\right) \boldsymbol{\Lambda}^{\mathrm{T}}. \tag{1}$$

As no column of $\mathbf{Q}_{\mathcal{J}}$ is identically zero, matrix $\boldsymbol{\Omega}_{\mathbf{Y},\mathbf{w}}$ has rank K for almost all $\mathbf{w}$.

It seems natural to weight cumulant matrices more if they are more precise. We therefore suggest to choose $w_{\ell,m}$ equal to the inverse of the simple average of the asymptotic variances of the components of an empirical analog $\widehat{\boldsymbol{\Omega}}_{\mathbf{Y}}(\ell, m)$ of $\boldsymbol{\Omega}_{\mathbf{Y}}(\ell, m)$. These variances can be computed by standard bootstrap.

All factors are skewed (or either skewed or kurtotic). Assuming that $\boldsymbol{\Lambda}$ and $\mathbf{Q}_{\mathcal{J}}$ have full column rank and that factor variables have non zero skewness, then matrix $\boldsymbol{\Gamma}_{\mathbf{Y}}$ has rank $K \leq \min(I_\ell, \ell \in \{1, ..., L\})$. One can thus apply the rank test to any analog estimator $\widehat{\boldsymbol{\Gamma}}_{\mathbf{Y}}$.

Assuming that each factor distribution is either skewed or kurtotic, matrix

$$\boldsymbol{\Phi}_{\mathbf{Y}} = [\boldsymbol{\Gamma}_{\mathbf{Y}}, \boldsymbol{\Omega}_{\mathbf{Y}}(1), ..., \boldsymbol{\Omega}_{\mathbf{Y}}(L)]$$

has rank K. One can thus test the rank of any consistent analog estimator $\widehat{\boldsymbol{\Phi}}_{\mathbf{Y}}$. Alternatively, in the same spirit as in the previous paragraph, remark that, under the assumption that all factors are skewed or kurtotic, all matrices

$$\boldsymbol{\Phi}_{\mathbf{Y},\mathbf{w}} = \boldsymbol{\Gamma}_{\mathbf{Y}} + \sum_{j=1}^{L} \mathbf{w}_j \boldsymbol{\Omega}_{\mathbf{Y}}(j) = \boldsymbol{\Lambda}\left[\mathbf{D}_3 + \mathbf{D}_4 \operatorname{diag}\left(\boldsymbol{\Lambda}^{\mathrm{T}} \mathbf{w}\right)\right] \mathbf{Q}_{\mathcal{J}}{}^{\mathrm{T}} \tag{2}$$

have rank K, for almost all weights $\mathbf{w} = (w_1, ..., w_L)^{\mathrm{T}} \in \mathbb{R}^L$. Matrices $\boldsymbol{\Phi}_{\mathbf{Y},\mathbf{w}}$ can therefore be used to estimate the number of factors K. We suggest to set w_j equal to the average of the variances of the components of $\widehat{\boldsymbol{\Gamma}}_{\mathbf{Y}}$ divided by the average of the variances of the components of $\widehat{\boldsymbol{\Omega}}_{\mathbf{Y}}(j)$.

3.2 Robin and Smith's (2000) rank test

Let $\widehat{\mathbf{B}}$ be a root-N consistent estimator of a (p, q), $p \geq q$, matrix $\mathbf{B}$, such that

$$N^{1/2} \operatorname{vec}\left(\widehat{\mathbf{B}} - \mathbf{B}\right) \stackrel{d}{\to} \mathcal{N}\left(0, \boldsymbol{\Sigma}_{\operatorname{vec}(\widehat{\mathbf{B}})}\right),$$

where $\boldsymbol{\Sigma}_{\operatorname{vec}(\widehat{\mathbf{B}})}$ is definite and $\operatorname{rank}\left(\boldsymbol{\Sigma}_{\operatorname{vec}(\widehat{\mathbf{B}})}\right) = s$, $0 < s \leq pq$. Note that $s < \dim(\mathbf{V})$ because of the symmetry properties of $\boldsymbol{\Gamma}_{\mathbf{Y}}$ and $\boldsymbol{\Omega}_{\mathbf{Y}}$. Let $\widehat{\boldsymbol{\Sigma}}_{\operatorname{vec}(\widehat{\mathbf{B}})}$ be a consistent estimate of $\boldsymbol{\Sigma}_{\operatorname{vec}(\widehat{\mathbf{B}})}$. Let $\widehat{\mathbf{B}} = \widehat{\mathbf{C}}\widehat{\mathbf{D}}\widehat{\mathbf{E}}^{\mathrm{T}}$ be the singular value decomposition of $\widehat{\mathbf{B}}$, where $\widehat{\mathbf{C}}$ and $\widehat{\mathbf{E}}$ are (p, p) and (q, q) orthogonal matrices and $\widehat{\mathbf{D}}$ is a (q, p) diagonal matrix. Let $\widehat{d}_1 \geq ... \geq \widehat{d}_K$ denote the diagonal entries of $\widehat{\mathbf{D}}^2$ (the eigenvalues of $\widehat{\mathbf{B}}^{\mathrm{T}}\widehat{\mathbf{B}}$). For a given null hypothesis: $H_0^r : K = r$, the statistics

$$\mathcal{CRT}_r \equiv N \sum_{i=r+1}^{q} \widehat{d}_i$$

Table 4: Size of the rank tests based on $\mathbf{\Omega_Y}$, $\mathbf{\Omega_{Y,w}}$ and $\mathbf{\Gamma_Y}$ for increasing kurtosis

$\kappa_4(\rho)$	-6/5 (uniform)	1/2	1	5	10	100	110	110	110
		(normal mixtures)					(log-normal)		
		Test based on $\mathbf{\Omega_Y}$						$\mathbf{\Omega_{Y,w}}$	$\mathbf{\Gamma_Y}$
α =.10	.90	.73	.82	.87	.85	.62	.56	.87	.90
α =.20	.79	.57	.67	.74	.69	.43	.34	.71	.79
α =.50	.47	.24	.32	.40	.35	.11	.08	.32	.48
α =.90	.10	.02	.04	.06	.04	.00	.00	.01	.07

Note: factors are uniform, normal mixtures or log-normal, errors are Gaussian, $\mathbf{\Lambda} = \mathbf{\Lambda}_2$, $N =$ 1000.

Table 5: Power of the improved rank test based on $\mathbf{\Omega_{Y,w}}$

$\kappa_4(\rho)$	-6/5 (uniform)	1/2	1	5	10	100
		(normal mixtures)				
α =.10	.99	.81	.81	1.00	1.00	.89
α =.20	.99	.63	.66	1.00	1.00	.80
α =.50	.96	.26	.29	.98	.99	.56
α =.90	.83	.02	.04	.72	.77	.12

Note: factors are normal mixtures, standard normal errors, $\mathbf{\Lambda} = \mathbf{\Lambda}_2$, $N = 1000$.

has the same limiting distribution as $\sum_{i=1}^{t} d_i^r Z_i^2$, where $d_1^r \geq ... \geq d_t^r$, $t \leq \min\{s, (p-r)(q-r)\}$, are the non-zero ordered eigenvalues of the matrix

$$(\widehat{\mathbf{E}}_{q-r} \otimes \widehat{\mathbf{C}}_{p-r})^{\mathrm{T}} \widehat{\mathbf{\Sigma}}_{\mathrm{vec}(\widehat{\mathbf{B}})} (\widehat{\mathbf{E}}_{q-r} \otimes \widehat{\mathbf{C}}_{p-r}),$$

where $\widehat{\mathbf{E}}_{q-r}$ and $\widehat{\mathbf{C}}_{p-r}$ are the last $q-r$ and $p-r$ columns of $\widehat{\mathbf{E}}$ and $\widehat{\mathbf{C}}$, respectively, and $\{Z_i\}_{i=1}^{\mathrm{T}}$ are independent standard normal variates.

To estimate K, we apply the following procedure. Start with $r = 0$. Test H_0^1 against $\widetilde{H}_0^1 : K > 0$. If H_0^1 is rejected, test H_0^2 against $\widetilde{H}_0^2 : K > 1$. And so on until one accepts H_0^r against $\widetilde{H}_0^r : K > r$. The test p-values can be approximated by drawing many independent values of the limiting statistics $\sum_{i=1}^{\mathrm{T}} d_i^r Z_i^2$. This procedure delivers a consistent estimate of K if the asymptotic sizes α_N^r used for the sequential tests are such that $\alpha_N^r = o(1)$ and $-N^{-1} \ln \alpha_N^r = o(1)$.

3.3 Monte carlo simulations

We here report a Monte-Carlo study of the rank tests detailed in 3.1. We first compute the empirical size of the test based on matrix $\mathbf{\Omega_Y}$ for various values of factor kurtosis. The simulation design is the same as for the results reported in Table 2. The true value of $\mathbf{\Lambda}$ is $\mathbf{\Lambda}_3$, and we test $K = 2$ against $K = 3$.

The first seven columns of Table 4 show substantial size distortion. This especially happens when excess kurtosis is low (in absolute value) – that is, when fourth-order cumulants contain very little information on the factor structure – or large – that is, when fourth-order moments are imprecisely estimated. However, for intermediate values of excess kurtosis the risk of underestimating the number of factors exists but remains limited.

In 3.1 we proposed to improve the size properties of the rank test by considering a weighted average of cumulant matrices $\mathbf{\Omega_Y}(\ell, m)$ – i.e. $\mathbf{\Omega_{Y,w}}$ in equation (1) – instead of $\mathbf{\Omega_Y}$. Column 8 in Table 4 shows that weighting scheme definitely improves the size of the test of $K = 2$ against $K = 3$. However, the rank test still under-rejects noticeably, in particular when the theoretical probability of rejection is low. Lastly, the last column refers to matrix $\mathbf{\Gamma_Y}$ (third-order cumulants). Third-order moments being more precisely estimated, the empirical size of the rank test based on $\mathbf{\Gamma_Y}$ is close to the nominal size (third column).

This confirms that applying the characteristic root test to matrices of high-order cumulants should be done with some caution when they are too imprecisely estimated. However, the results in Table 4 show that, when skewness and excess kurtosis are not too large, the size properties of the rank test based on third and fourth-order cumulant matrices are satisfactory.

We end this section by a study of the power of the rank test based on $\mathbf{\Omega_{Y,w}}$. Table 5 displays empirical power computations for various levels of kurtosis. The true value of $\mathbf{\Lambda}$ is $\mathbf{\Lambda}_1$ and we test $K = 2$ against $K = 3$. For low levels (α less than 10%) the power of the test is good even if factors are strongly leptokurtic. For intermediate values of excess kurtosis, the power is good whatever the level.

4 Factor analysis on test scores data

In this Section we detail the application to British data on cognitive test scores that we briefly presented in the main paper.

4.1 The data

The NCDS is a longitudinal survey of a British birth cohort born in the same week of 1958. We use the following waves: 1965 (age 7), 1969 (age 11), 1974 (age 16), and 2000 (age 42). To select the sample we consider all individuals for whom we have information on test scores for the first three waves. There are seven available test measures: mathematics and reading at age 7, 11 and 16, and a verbal test at age 11 only. We also use the age at the time of leaving school and the logarithms of monthly and hourly wages measured at age 42.

Table 6 shows the first moments of the variables of interest. For interpretability, we have rescaled the test score measures so that they range between 0 and 100 points. We remark that most test scores present some skewness, either right or left, and negative excess kurtosis. In Table 7 we show the correlations between the seven test score measures and the years of education and log wage variables. We see that these correlations are all

Table 6: Descriptive statistics

	Mean	Variance	Skewness	Ex. kurtosis	N
Math (7)	53.4	589	.040	-.77	7816
Reading (7)	79.9	480	-1.17	.42	7816
Math (11)	44.4	657	.21	-1.07	7816
Reading (11)	47.8	302	.087	-.47	7816
Verbal (11)	58.2	509	-.23	-.92	7816
Math (16)	42.6	498	.46	-.68	7816
Reading (16)	75.0	329	-.89	.34	7816
Years left education	17.5	5.56	1.70	2.64	5653
Log monthly wage (2000)	4.51	.622	-.74	2.68	4012
Log hourly wage (2000)	.945	.308	-.59	8.65	3982
Female dummy	.492	.250	.03	-2.00	7816

Note: sample taken from the NCDS data, years 1965, 1969, 1974 and 2000.

positive. Moreover, more recent scores and scores in mathematics appear more strongly correlated with later outcomes. Lastly, girls do slightly better in reading/verbal, and slightly worse in mathematics, than boys.

Table 7: Correlations between test scores and education, log wage and gender

	Years education	Log monthly wage	Log hourly wage	Female
Math (7)	.26	.20	.19	-.06
Reading (7)	.29	.10	.13	.12
Math (11)	.46	.25	.27	-.03
Reading (11)	.46	.24	.26	-.01
Verbal (11)	.39	.15	.19	.11
Math (16)	.53	.30	.31	-.11
Reading (16)	.42	.25	.26	-.03

4.2 Results

We analyze the data with an independent factor model. In the present context, it is natural to allow for errors, the distributions of which are *a priori* different for each test measure. Moreover, as the exams were given on the same day (in the three interviews) it makes sense to allow for contemporaneous correlation between test scores. For instance, it could be that a child had a “bad day” and performed badly in all the tests. For this reason, we allow for correlation between the errors in the reading and mathematics scores at age 7 and age 16, and between the reading, mathematics and verbal scores at age 11.

Our approach requires that the data be sufficiently non-normal. The moments reported in Table 6 show that there is some non-normal skewness and kurtosis in the marginal distributions of the score variables. In order to check the extent of non-normality in the joint distribution of test scores we performed the tests outlined in 3.1. The results of the rank tests based on third and fourth-order moments imply that we can reject the restrictions imposed by a 5-factor model, while a 6-factor model cannot be rejected. Nevertheless, the estimates based on four-factor and five-factor models turned out to be very imprecise. So, in the following we present the results for one to three factors. In the estimation we use second, third and fourth-order moments jointly. To account for the fact that lower-order moments are better estimated, we weight the cumulant matrices as explained in Section 4 of the main paper. Relative to second-order moments, third-order moments are thus weighted by a factor .178, and fourth-order ones by .091.

Table 8 shows the estimation results. The first three columns show the factor loadings estimates that correspond to each of the seven test score measures. The last seven columns give the estimates of the variance-covariance matrix of error variables. The last two rows give the skewness and excess kurtosis of the factors. Lastly, bootstrap standard errors are given in parentheses (100 iterations). To interpret the factor variables, we give in Table 9 the correlation between the linear projection of test scores on factor loadings ($\widehat{\mathbf{X}} \equiv \mathbf{\Lambda}^{-}\mathbf{Y}$), and the variables of interest. We interpret $\widehat{\mathbf{X}}$ as an estimate of the vector of factor variables, though a more correct approach would consist in filtering $\mathbf{X}$ out using the independence assumptions (see Bonhomme and Robin, 2008b).

Table 8 shows that errors are sizeable in our application. All error variances are significantly different from zero. Moreover, errors are larger for the test scores at age 7. For instance, in the one-factor specification the error variance represents 66%, 33% and 45% of the variances of the math test scores at age 7, 11 and 16, respectively. Overall, the ratio of the sum of squares of factor loadings to total variance is 60%. This suggests that overlooking error variables in the model can have severe consequences on the results. To check that, we re-estimated the factor loadings using JADE. We found that the second and third factors were essentially driven by the math and reading test scores at age 7, respectively. This is likely to be because the large errors in the test score at age 7 are wrongly interpreted as extra factors.

We also see from Table 8 that errors are generally contemporaneously positively correlated. The only exception is for the test scores at age 16 in the specification allowing for three factors ($K = 3$, last rows of the table).

Turning to factor loadings estimates, one sees that the one-factor specification weights all test scores similarly. The factor loadings estimates are very close to the ones we obtained using second-order moments only, by applying ordinary Factor Analysis. Moreover, they are very precisely estimated. A natural interpretation of this factor could be the child's general ability. From Table 9 we see that it is positively correlated with years of education (.50) and log wages (.30), and that it is equally distributed among boys and girls.

Allowing for a second factor yields a rather different picture, as none of the two factors is similar to the one estimated in the one-factor specification. The first factor is correlated with scores in reading and mathematics, the correlation being stronger with math. This factor is positively skewed, and presents negative excess kurtosis. In contrast,

Table 8: Model estimates

	Factor loadings			Error covariances						
ONE FACTOR										
Math (7)	14.3 (.26)	-	-	390 (5.7)	38.2 (3.8)	0	0	0	0	0
Reading (7)	15.5 (.22)	-	-	38.2 (3.8)	227 (4.2)	0	0	0	0	0
Math (11)	21.5 (.16)	-	-	0	0	219 (3.9)	43.3 (2.7)	69.7 (3.6)	0	0
Reading (11)	14.1 (.15)	-	-	0	0	43.3 (2.7)	114 (2.5)	38.1 (2.3)	0	0
Verbal (11)	18.5 (.17)	-	-	0	0	69.7 (3.6)	38.1 (2.3)	171 (3.5)	0	0
Math (16)	17.3 (.18)	-	-	0	0	0	0	0	228 (3.3)	8.58 (2.3)
Reading (16)	15.2 (.18)	-	-	0	0	0	0	0	8.58 (2.3)	101 (2.2)
Skewness	−.239 (.020)	-	-	.167 (.026)	−3.28 (.082)	1.88 (.076)	.203 (.047)	−.533 (.067)	1.84 (.047)	−4.34 (.17)
Ex. kurtosis	−.888 (.029)	-	-	−1.52 (.065)	2.80 (.30)	−5.84 (.31)	−.750 (.10)	−4.62 (.21)	−1.77 (.16)	7.96 (.70)
TWO FACTORS										
Math (7)	13.6 (.35)	4.61 (.48)	-	378 (6.0)	59.3 (4.0)	0	0	0	0	0
Reading (7)	10.8 (.44)	11.3 (.55)	-	59.3 (4.0)	230 (7.3)	0	0	0	0	0
Math (11)	22.3 (.35)	6.71 (.56)	-	0	0	118 (6.1)	17.3 (2.4)	45.6 (3.0)	0	0
Reading (11)	11.0 (.32)	9.92 (.32)	-	0	0	17.3 (2.4)	90.1 (2.3)	15.3 (2.1)	0	0
Verbal (11)	14.6 (.43)	12.1 (.46)	-	0	0	45.6 (3.0)	15.3 (2.1)	156 (3.4)	0	0
Math (16)	18.4 (.34)	4.24 (.50)	-	0	0	0	0	0	137 (9.3)	3.22 (3.5)
Reading (16)	11.1 (.35)	11.0 (.40)	-	0	0	0	0	0	3.42 (3.5)	85.5 (5.6)
Skewness	.600 (.047)	−1.71 (.087)	-	−.106 (.036)	−2.95 (.11)	−2.18 (.29)	.573 (.11)	−.646 (.19)	.868 (.15)	−4.62 (.27)
Ex. kurtosis	−1.29 (.045)	1.24 (.28)	-	−1.59 (.067)	1.72 (.33)	−11.1 (.88)	−4.71 (.27)	−8.71 (.56)	−1.43 (.32)	4.93 (.90)
THREE FACTORS										
Math (7)	13.6 (.25)	5.43 (.60)	−4.03 (.92)	363 (7.0)	16.4 (4.4)	0	0	0	0	0
Reading (7)	10.9 (.26)	13.4 (.90)	−6.64 (1.5)	16.4 (4.4)	141 (8.0)	0	0	0	0	0
Math (11)	22.4 (.21)	5.30 (.45)	−1.83 (1.1)	0	0	128 (5.1)	37.1 (2.9)	60.1 (3.1)	0	0
Reading (11)	11.0 (.25)	8.77 (.27)	2.06 (1.2)	0	0	37.1 (2.9)	104 (1.9)	37.2 (2.2)	0	0
Verbal (11)	14.8 (.28)	10.7 (.39)	−1.47 (1.4)	0	0	60.1 (3.1)	37.2 (2.2)	175 (3.7)	0	0
Math (16)	18.8 (.28)	4.07 (.41)	3.79 (.94)	0	0	0	0	0	114 (3.6)	−41.6 (2.1)
Reading (16)	12.2 (.37)	10.9 (.65)	7.05 (1.4)	0	0	0	0	0	−41.6 (2.1)	15.2 (1.7)
Skewness	.552 (.036)	−1.65 (.069)	.009 (.91)	−.0764 (.034)	−5.23 (.50)	−1.82 (.22)	−.0635 (.082)	−1.03 (.11)	1.18 (.16)	−68.9 (9.5)
Ex. kurtosis	−1.28 (.040)	1.28 (.21)	.520 (1.04)	−1.71 (.066)	3.52 (.61)	−8.83 (.62)	−3.07 (.17)	−6.51 (.32)	−.927 (.49)	185 (57)

Table 9: Correlation of the predicted factors with several variables

	One factor	Two factors		Three factors		
Math (7)	.682 (.008)	.685 (.031)	.035 (.048)	.662 (.018)	.070 (.059)	−.504 (.030)
Reading (7)	.751 (.005)	.352 (.034)	.654 (.033)	.367 (.016)	.685 (.054)	−.456 (.072)
Math (11)	.914 (.002)	.861 (.014)	.145 (.023)	.899 (.0072)	.088 (.020)	−.116 (.030)
Reading (11)	.842 (.004)	.538 (.022)	.521 (.022)	.593 (.017)	.476 (.019)	.135 (.063)
Verbal (11)	.877 (.003)	.575 (.025)	.521 (.022)	.640 (.015)	.452 (.020)	−.104 (.057)
Math (16)	.821 (.004)	.867 (.016)	−.011 (.024)	.870 (.013)	−.015 (.018)	.144 (.026)
Reading (16)	.821 (.004)	.492 (.023)	.555 (.024)	.356 (.021)	.542 (.030)	.291 (.067)
Years educ.	.494 (.010)	.454 (.013)	.078 (.012)	.470 (.011)	.065 (.011)	.110 (.012)
Log monthly wage	.261 (.015)	.293 (.015)	−.048 (.015)	.292 (.017)	−.046 (.017)	.093 (.013)
Log hourly wage	.281 (.017)	.290 (.015)	−.011 (.013)	.293 (.020)	−.014 (.016)	.081 (.014)
Female dummy	−.002 (.011)	−.136 (.012)	.203 (.011)	−.119 (.014)	.190 (.011)	−.128 (.011)

Note: bootstrapped standard errors in parentheses.

the second factor is correlated to reading test scores, but has small or zero correlation with the scores in mathematics. Contrary to the first one, this factor is both negatively skewed and leptokurtic. Moreover, the first and second factors account for 45% and 19% of the total variance, while errors account for 36%. Separating these two components requires to use third and fourth-order moments of the data, in order to fix the rotation matrix. This explains why standard errors are rather large compared to the one-factor specification. However, we remark that the estimates are still precise.

These results are consistent with the existence of different components of ability. Columns 2 and 3 in Table 9 show that the first factor is strongly related to math test scores, while the second only determines reading and verbal ability. Moreover, the first factor is strongly correlated with education and the log hourly wage (.45 and .30), while the second is less strongly correlated with education (.08) and is uncorrelated with the log hourly wage. This suggests that the second component of ability does not increase labor productivity. Lastly, girls are more likely to be endowed with the second factor, the negative correlation with the log monthly wage indicating that it is negatively associated with labor market participation.

Notice that, given $L = 7$ and $J = 16$, the bound on the number of factors that can be identified if only second-order moments are used in the prewhitening step of the algorithm is:

$$K = \frac{2L + 1 - \sqrt{(2L+1)^2 - 8J}}{2} \approx 2.58.$$

Hence, in order to identify a third factor, higher-order data moments are required in the first step. Adding a third factor, we remark that the first two factors remain unchanged: both factor loadings and moments are very similar to their values in the two-factor specification. This result confirms that the first two factors represent true dimensions of ability. As shown by Tables 8 and 9, the third factor puts positive weights on later test scores (age 16) and negative weights on earlier ones (age 7). Moreover, it accounts for an additional 4% of total variance. This factor shows some excess kurtosis, though badly estimated, and it is positively correlated with education and log wages, albeit less so than the first factor (the correlation is .10 with education and the two log wage measures). Lastly, being a girl is negatively associated with this factor. We interpret the third factor as reflecting heterogeneous learning slopes. It allows to distinguish children who learn more at the beginning (age 7) or at the end (age 16) of their schooling career.

To conclude, this application shows that our algorithm succeeds in identifying three interpretable test score factors. A first dimension of children's ability reflects mathematical skills. It has a high positive return in terms of education and wages. The second dimension of ability is only correlated to reading and verbal test scores. It contributes a little to education, but does not increase labor market productivity. Moreover, it is more frequent among girls. The third dimension reflects the learning slopes of children. This last factor accounts for a small part of total variance, and has positive returns on education and wages.

5 Application to stock returns

In an influential paper, Fama and French (1993) identify three factors explaining a large proportion of the variance of time-series of U.S. excess stock returns, $Y_\ell(t) = R_\ell(t) - R_F(t)$, $\ell = 1, ..., L$. In addition to the market return ($R_M(t) - R_F(t)$, where $R_F(t)$ is the risk-free return), which is the unique factor of the CAPM model, they identify two additional factors:

- $SMB(t)$, or "small minus big", is the difference between the average of the returns on two stock portfolios: one containing firms with market value (price time number of shares) less than the median, and one containing firms with size above the median.

- $HML(t)$, or "high minus low", is the difference between the average of the returns on two stock portfolios: one gathering firms with book-to-market ratio (book value of capital divided by market value, denoted B/M) less that the 30th percentile and another one containing all firms with B/M ratio above the 70th percentile.

Fama and French show that these three factors explain monthly data on 25 portfolios formed by intersecting size and book-to-market quintiles remarkably well. Other relevant contributions by the same authors include Fama and French (1995, 1996) and Davis, Fama and French (2000). Fama and French's factors are now widely used in applied finance to summarize the correlations between bond or stock returns.

In this Section we apply quasi-JADE to estimate a linear independent factor model with three factors. Unlike Fama and French, who construct factors on the basis of economic intuition, we shall estimate factors blindly.

Table 10: Moments of the variables

Size	B/M ratio	Mean	St. Dev.	Skewness	Kurtosis
Small	Low	.0008	1.10	-.86	13.8
	2	.0293	.91	-.87	13.3
	3	.0341	.77	-.98	15.1
	4	.0428	.71	-1.00	15.4
	High	.0479	.71	-.97	14.2
2	Low	.0123	1.17	-.50	10.0
	2	.0251	.91	-.68	12.4
	3	.0383	.81	-.67	11.6
	4	.0404	.77	-.69	12.9
	High	.0447	.87	-.55	10.2
3	Low	.0148	1.15	-.38	9..8
	2	.0302	.88	-.59	12.8
	3	.0309	.77	-.62	12.9
	4	.0372	.77	-.48	11.9
	High	.0447	.88	-.66	16.5
4	Low	.0220	1.10	-.21	12.0
	2	.0209	.87	-.75	18.5
	3	.0320	.81	-.92	20.4
	4	.0383	.80	-.63	15.7
	High	.0394	.92	-.57	15.3
Big	Low	.0198	1.06	-.37	15.8
	2	.0214	.95	-.96	27.6
	3	.0231	.91	-.90	27.3
	4	.0260	.88	-.92	32.2
	High	.0281	.99	-.57	17.9

The data. We use daily US observations between 01/07/1963 and 31/08/2005 of the returns to 25 stock portfolios formed on size and book-to-market.[3] With monthly data, we obtained similar results, though less precisely estimated. The size and book-to-market breakpoints are NYSE quintiles. There are 10,616 observations. Table 10 shows the mean, standard error, skewness and kurtosis of the returns on the 25 portfolios. Returns are net of the risk-free rate R_F, which varies between .003 and .061 over the period. All returns appear strongly leptokurtic, and somewhat skewed to the left.

Estimation results. Table 11 presents the estimates of factor loadings and error variances, under the assumption that $K = 3$. Quasi-JADE was applied using second, third and fourth-order cumulants. Interestingly, using only third or only fourth-order moments made very little difference in estimation. Moreover, bootstrapped confidence intervals

[3]These data can be downloaded from Kenneth French's website:
http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

(not reported) show that most factor loadings are rather precisely estimated. Overall, the three factors account for 90% of total variance. The first factor explains 54%, the second factor 26% and the third factor 10%, about the same as the error term. Lastly, one notices that all three factors are skewed to the left and strongly leptokurtic.

Given that the model is nearly noise-free, we predict factor levels as $\widehat{\mathbf{X}}(t) = \widehat{\mathbf{\Lambda}}^{-}\mathbf{Y}(t)$, where $\widehat{\mathbf{\Lambda}}^{-}$ is the generalized inverse of the estimated matrix of factor loadings $\widehat{\mathbf{\Lambda}}$. When the error variance cannot be neglected, predicting factor levels requires a more complicated procedure that is developed in a companion paper (Bonhomme and Robin, 2008b).

Figure 1 displays the elements of matrix $\widehat{\mathbf{\Lambda}}^{-}$ as a function of size quintiles (panels a)), and B/M quintiles (panels b)). The three columns are reported from top to bottom of the figure. Dashed lines represent OLS fit. Panels a2) and b2) show that the second factor is strongly negatively correlated with size and almost uncorrelated with book-to-market. Conversely, panels a3) and b3) show that the third factor is weakly negatively correlated with size, yet strongly positively correlated with book-to-market. Lastly, the first factor appears positively correlated with size and to book-to-market, although the correlation with the latter is weaker. These results are qualitatively the same if we estimate a three-factor model on 100 portfolios formed as the intersection of size and book-to-market deciles, as shown by Figure 2. Interestingly, unrotated PCA yields similar pictures as Figures 1 and 2. This suggests that unrotated principal components are approximately independent.

We then provide a direct comparison of these estimated factors to those used by Fama and French. To do so, we compute Fama and French's factors and correlate them to $\widehat{\mathbf{X}}_t$. Panel a) of Table 12 shows these correlations. We see that the three factors estimated by quasi-JADE are strongly correlated with the market, size and book-to-market factors constructed by Fama and French. The correlations are .84, .85 and .90, respectively. Note that Fama and French's factors are correlated. For instance, the market return has correlation $-.24$ and $-.58$ with *SMB* and *HML*, respectively. For this reason, they cannot be equal to the independent factors obtained by quasi-JADE. We then apply JADE to market return, *SMB* and *HML*, and obtain new factors which are by construction independent. Panel b) of table 12 reports the correlations between these new factors and the factors that were initially estimated by quasi-JADE on the 25 portfolios. We find very high correlations (.97, .98 and .93).

It is interesting to evaluate the extent to which these results are driven by the *ex-ante* grouping into size and book-to-market cells. Our experiments on stock data grouped by industries[4] showed that if the first factor remained strongly linked to market return, size and book-to-market were much less strongly correlated with the two other factors. This casts some doubts on the ability of Fama and French's factors to explain very disaggregate data on stock returns with the same success.

[4]These data can be found in the section "change in industry portfolios" in French's data library.

Table 11: Factor loadings, factor moments and error variances

Size	B/M ratio	Factor 1	Factor 2	Factor 3	Error variance	R^2
Small	Low	.57	.79	-.37	.055	.95
	2	.50	.67	-.23	.054	.93
	3	.45	.56	-.14	.035	.93
	4	.41	.52	-.09	.029	.94
	High	.43	.52	-.05	.028	.94
2	Low	.70	.72	-.50	.102	.92
	2	.59	.58	-.26	.049	.94
	3	.54	.51	-.17	.066	.89
	4	.53	.47	-.10	.079	.84
	High	.61	.51	-.08	.118	.82
3	Low	.71	.61	-.58	.122	.90
	2	.62	.47	-.30	.050	.93
	3	.58	.40	-.15	.061	.88
	4	.60	.37	-.10	.080	.84
	High	.70	.40	-.06	.095	.85
4	Low	.74	.46	-.61	.094	.92
	2	.70	.33	-.28	.041	.94
	3	.69	.29	-.16	.038	.94
	4	.68	.28	-.08	.086	.84
	High	.77	.30	-.06	.146	.79
Big	Low	.83	.14	-.52	.125	.87
	2	.84	.08	-.33	.034	.96
	3	.81	.08	-.23	.077	.89
	4	.80	.07	-.12	.061	.91
	High	.83	.13	-.12	.256	.64
% Variance		.544	.258	.099	.099	
Skewness		-1.21	-.76	-.56	-	
Kurtosis		30	28	77	-	

Table 12: Fama French factors *versus* Quasi-JADE estimates, daily US data, 25 portfolios

Factors	X_1	X_2	X_3
$R_M - R_F$	.84	.24	-.41
SMB	-.49	.85	-.09
HML	-.11	.23	.90

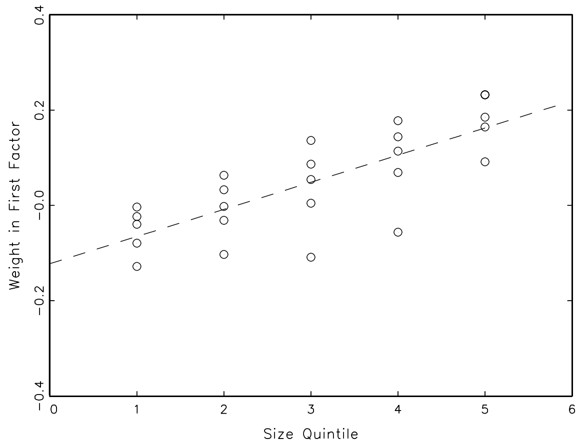
a) Fama French

Factors	X_1	X_2	X_3
$R_M - R_F$	.97	-.01	-.09
SMB	-.01	.98	-.06
HML	.09	.07	.93

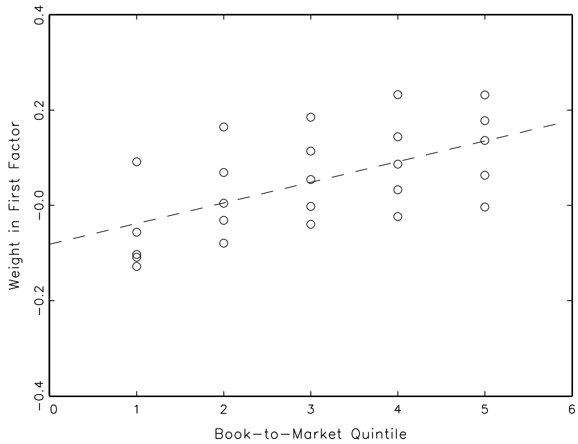
b) Independent Fama French

Figure 1: Independent factors against quintiles of size and book-to-market, daily Fama French data, 25 portfolios

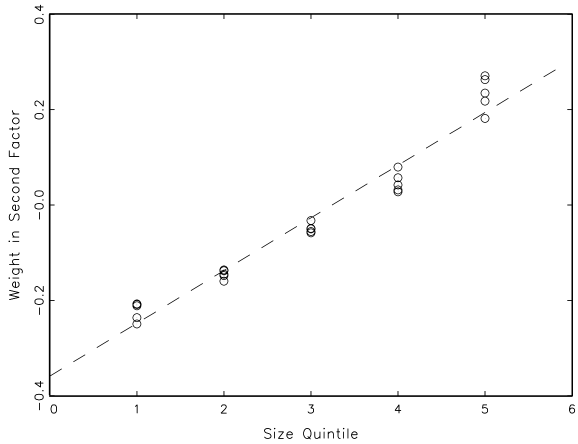
a1) Size/X_1



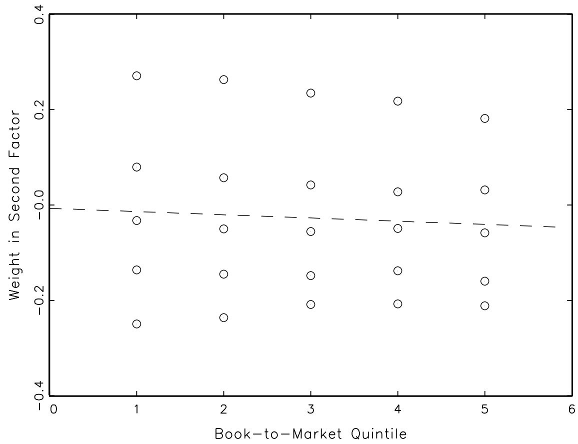
b1) Book-to-Market/X_1



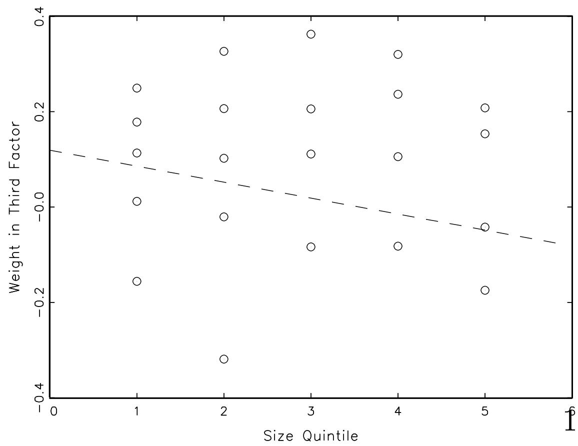
a2) Size/X_2



b2) Book-to-Market/X_2



a3) Size/X_3



b3) Book-to-Market/X_3

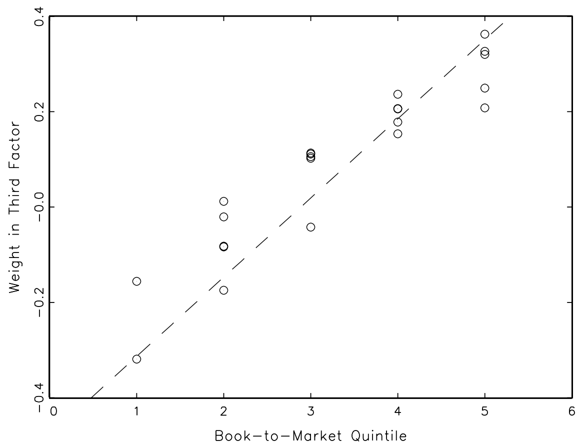
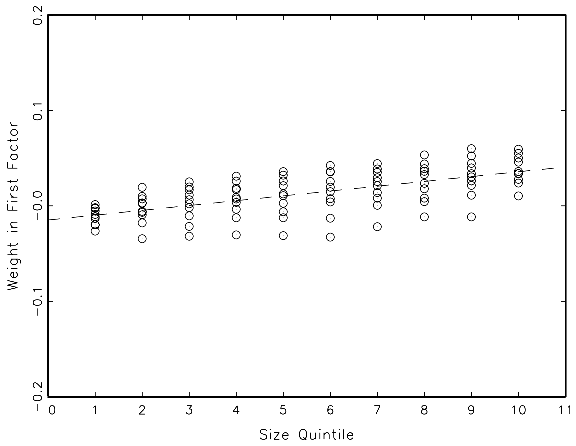
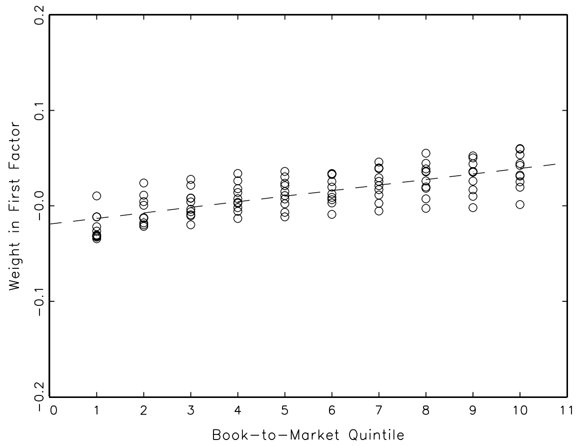


Figure 2: Independent factors against deciles of size and book-to-market, daily Fama French data, 100 portfolios

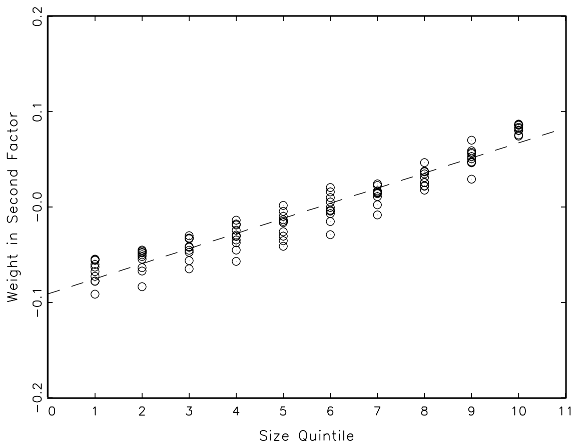
a1) Size/X_1



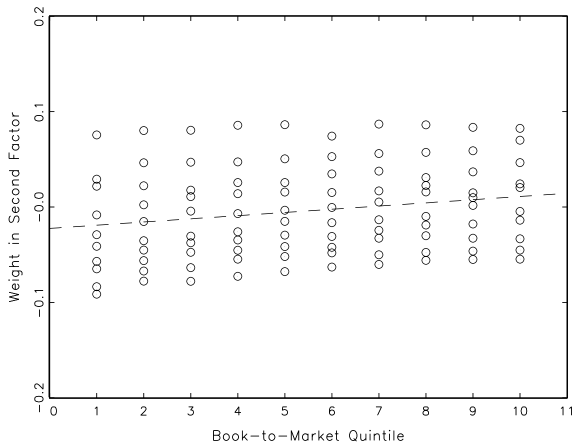
b1) Book-to-Market/X_1



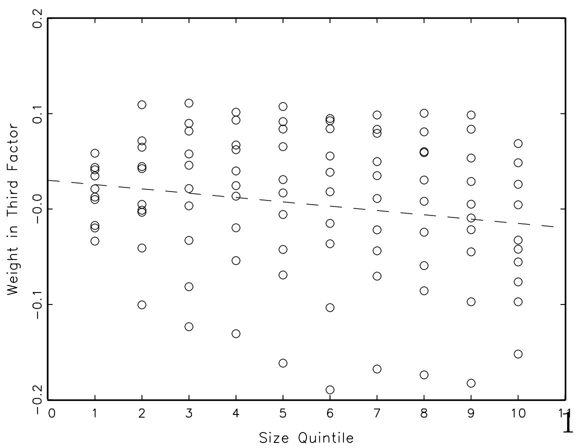
a2) Size/X_2



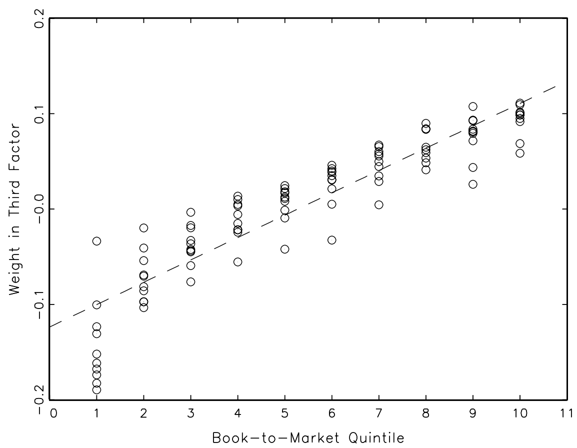
b2) Book-to-Market/X_2



a3) Size/X_3



b3) Book-to-Market/X_3



References

[1] Bonhomme, S. and J.M. Robin, 2008a, Consistent noisy independent component analysis. Mimeo.

[2] Bonhomme, S. and J.M. Robin, 2008b, Generalized nonparametric deconvolution with an application to earnings dynamics. Mimeo.

[3] Davis, J., E. Fama and K. French, 2000, Characteristics, covariances and average returns. Journal of Finance 55, 389-406.

[4] Fama, E. and K. French, 1993, Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33, 3-56.

[5] Fama, E. and K. French, 1995, Size and book-to-market factors in earnings and returns. Journal of Finance 50, 131-155.

[6] Fama, E. and K. French, 1996, Multifactor explanations of asset pricing anomalies. Journal of Finance 51, 55-84.

[7] Robin, J.M. and R.J. Smith, 2000, Tests of rank. Econometric Theory 16, 151-175.