1 Introduction

1.1 Background

This project seeks to recreate parts of the analysis done in Tests of asset pricing with time-varying factor loads (Galvao et al, 2018). Galvao et al estimated factor risk premia using time-series regressions of excess returns against factor loadings. The authors proposed a Wald-like test statistic that is claimed to be adjusted for the presence of generated regressors in the time-series regression. Using this test statistic, the author tests for the correctness of asset pricing models by testing the joint hypothesis of a zero intercept and homogeneous factor risk premia across risky assets.

When applied to US-industry portfolios over the time period from 1963 to 2024, the paper found overwhelming evidence rejecting CAPM, Fama-French 3 factors and Fama-French 5 factors.

In this project, the same approach was taken with an extended dataset covering US-industry portfolios from 1963 to 2025. Overwhelming evidence was found against the null hypotheses of zero intercept and homogenous risk premia when tested separately and jointly. However, under MC simulations, this project was unable to verify the asymptotic level of the proposed tests, and the proposed tests seemed to reject the null too often.

1.2 Original methodology

Using US-industry portfolio returns and Fama-French factor returns, Galvao et al's first stage involves estimating factor-loads. The factor loads are generated as the fitted values in a AR(1) regression of a time-series of realized covariances between asset and factor returns.

Following the estimation of factor loads, the second stage assumes the presence of some known number R of unobservable common factors. Galvao et al estimate factor risk premia, unobservable factor returns, as well as loading on unobservable factors jointly by solving a constrained minimization problem.

Galvao et al then describes an asymptotic variance estimator for the estimated factor risk premia, and derives 3 test statistics to test for zero-intercept, homogenous risk premia, and the joint null hypothesis.

1.3 Replication methodology

In replicating this study, this project seeks to 1) extend dataset with data from 1963-2025; 2) implement estimation of time-varying factor loads and verify the implementation with intermediate data provided by Galvao et al; 3) implement Galvao et al's factor risk premia estimator via two different approaches, a penalty-based unconstrained optimization approach and Galvao et al's iterative convergence approach; 4) implement estimation of asymptotic variance of the above estimator 5) implement calculation of test statistics and verify asymptotic level and power and 6) conduct empirical homogeneity tests on the extended dataset.

2 Data

2.1 Original dataset

Galvao et al published an anonymized replication dataset with 12965 rows and 53 columns. From their description, this dataset is created from Fama-French's 49-US industry portfolios (daily) data, and Fama-French's 5 factor data, with Healthcare and Software industries removed due to data availability issues. Their replication data covers the period from July 1963 to December 2014.

2.2 Replication dataset

To extend the dataset, this project retrieved updated data from Kenneth French's data library. Since Galvao et al's dataset contained no indices or column names, Appendix 1 outlines the steps taken to align the original and replication dataset. The replication set contains 15,690 daily returns across 47 industry portfolios, 5 factors and the risk free rate. This yields 15,690 daily observations to estimate factor-loadings at quarterly frequency, and 206 quarterly observations for estimating factor risk premia.

3 Models

As a starting point, Galvao et al assumes asset pricing models of the form

$$r^e_{i,t+1} = \alpha_i + \beta_{it}^\top f_{t+1} + \phi_{it}^\top h_{t+1} + e_{i,t+1} \tag{1}$$

where $r^e_{i,t+1}$ is the excess return on asset i, $f_{t+1} \in \mathbb{R}^K$ is a realized vector of known systematic risk factors, $h_{t+1} \in \mathbb{R}^R$ is a realized vector of unknown factors. $\beta_{it} \in \mathbb{R}^K$ are the factor loadings on the known factors and ϕ_{it} are the loadings on the unknown factors, and $e_{i,t+1}$ satisfies $\mathbb{E}\left[e_{i,t+1}|f_{t+1}, h_{t+1}\right] = 0$.

Using the notation that $\tilde{f}_{t+1} = f_{t+1} - \mathbb{E}\left[f_{t+1}\right], \tilde{h}_{t+1} = h_{t+1} - \mathbb{E}\left[h_{t+1}\right]$, $\tilde{g}_{t+1} = (\tilde{f}_{t+1}^\top, \tilde{h}_{t+1}^\top)^\top, \beta_i^{*\top} = \left(\beta_i^\top, \phi_i^\top\right)$, and let $\lambda_{i,t+1} \in \mathbb{R}^K$ denote a vector of factor risk premia capturing the price of risk f_{t+1} for asset i, Galvao et al describes the time-series regression model

$$r^e_{i,t+1} = \alpha_i + \beta_{it}^\top \lambda_i + v_{i,t+1} \tag{2}$$

$$v^e_{i,t+1} = \beta_{it}^{*\top} \tilde{g}_{it} + \epsilon_{i,t+1} \tag{3}$$

3.1 Estimating time-varying factor loadings

To estimate the panel data of λ_{it}, the Galvao et al first generates the necessary regressors β_{it}. Let $\triangle$ denote sampling frequency and $m = 1/\triangle$ denote number of obsevations per period t, let intra-period returns from $t+h\triangle$ to $t+(h+1)\triangle$ be $R_{t+(h+1)\triangle} = p_{t+(h+1)\triangle} - p_{t+h\triangle}$, $h = 0, \dots m-1$, and let inter-period return be $R_{t+1} = \sum_{h=0}^{m-1} R_{t+(h+1)\triangle}$. For each time step, the author stacks N risky asset returns and K factor returns in a $(N+K)$ vector denoted $R_{t+1} = (r^e_{1,t+1}, \dots r^e_{N,t+1}, f_{1,t+1}, \dots f_{K,t+1})$. The author denotes a realized covariance matrix as

$$\hat{\Omega}_{t+1} = \sum_{h=0}^{m-1} R_{t+(h+1)\triangle} R^\top_{t+(h+1)\triangle}$$

More concretely, the realized convariance matrices form a $(N+K) \times (N+K) \times T$ time-series.

Under a series of assumptions about the stochastic process that generates prices, the authors propose the following autoregressive process for each element of the realized covariacne matrix. Using $\omega_{(i,j),t+1}$ to denote the entry corresponding to the i-th asset and j-th factor

$$\omega_{(i,j),t+1} = \delta_{ij,0} + \delta_{ij,1}\omega_{(i,j),t} + v_{ij,t+1} \tag{4}$$

The authors then show that a consistent estimator of the time series $\beta_{ij,t}$ is

$$\hat{\beta}_{ij,t} = \hat{\delta}_{ij,0} + \hat{\delta}_{ij,1}\omega_{(i,j),t} \tag{5}$$

Where $\hat{\delta}_{ij,0}, \hat{\delta}_{ij,1}$ are the OLS estimates in Equation (4).

3.2 Estimation of factor risk premia

The authors adopt the notation $\hat{X}_{it} = \left(1, \beta_{it}^{\top}\right)^{\top}$ and $\eta_i = (\alpha_i, \lambda_i^{\top})^{\top}$, and fit the following model

$$r_{i,t+1}^{e} = \hat{X}_{it}\eta_i \tag{6}$$

$$r_{i,t+1}^{e} = \beta_{it}^{*\top}\tilde{g}_{it} + \epsilon_{i,t+1} \tag{7}$$

The authors propose that the estimators are

$$\hat{\eta}_i, \hat{\beta}_i^{*}, \hat{\tilde{g}}_{t+1} = \arg\min_{\eta,\beta^*,\tilde{g}} l\left(\eta_i, \beta_i^{*}, \tilde{g}_{t+1}\right) \tag{8}$$

$$= \arg\min_{\eta,\beta^*,\tilde{g}} \sum_{i=1}^{N}\sum_{i=1}^{T}\left(r_{i,t+1}^{e} - \hat{X}_{it}\eta_i - \beta_i^{*\top}\tilde{g}_{t+1}\right)^2 \tag{9}$$

Subject to $\frac{G^{\top}G}{N} = I$, $\frac{\beta^{*\top}\beta^{*}}{N}$ diagonal, where $G = (\tilde{g}_1, \ldots \tilde{g}_T)^{\top} \in \mathbb{R}^{T\times R}$, $\beta^8 = (\beta_1^{*}, \ldots \beta_N^{*})^{\top} \in \mathbb{R}^{N\times R}$. The authors argue that the solutions to this minimization problem should simultaneously solve

$$\hat{\eta}_i = \left(\hat{X}_i^{\top}M_{\hat{G}}\hat{X}_i\right)^{-1}\hat{X}_i^{\top}M_{\hat{G}}r_i^{e} \tag{10}$$

$$\left[\frac{1}{NT}\sum_{i=1}^{N}\left(r_i^{e} - \hat{X}_i\hat{\eta}_i\right)\left(r_i^{e} - \hat{X}_i\hat{\eta}_i\right)^{\top}\right]\hat{G} = \hat{G}\hat{V}_{NT} \tag{11}$$

Where $\hat{G}$ is the estimate of G consisting of $\hat{\tilde{g}}$, $M_{\hat{G}} = I - \hat{G}\left(\hat{G}^{\top}\hat{G}\right)^{-1}\hat{G}^{\top}$, and $\hat{V}_{NT}$ is a diagonal matrix of the R largest eigenvalues of $\hat{G}$. The authors propose solving this system by iterating Equation (10) and (11) until convergence.

As an extension to the original study and for ease of computation, this project proposes an alternative solution to the system by solving the following unconstrained minization for some fixed φ

$$\hat{\eta}_i, \hat{\beta}_i^{*}, \hat{\tilde{g}}_{t+1} = \arg\min_{\eta,\beta^*,\tilde{g}} \sum_{i=1}^{N}\sum_{i=1}^{T}\left(r_{i,t+1}^{e} - \hat{X}_{it}\eta_i - \beta_i^{*\top}\tilde{g}_{t+1}\right)^2 - \varphi\|\frac{G^{\top}G}{T} - I\|_F^2 \tag{12}$$

3.3 Estimation of asymptotic variance

The author argues that a consistent estimator of the asymptotic variance of $\sqrt{T}\left(\hat{\eta} - \eta\right)$ is

$$\widehat{\text{Avar}}\left(\sqrt{T}\left(\hat{\eta} - \eta\right)\right) = \left(\hat{S} - \frac{1}{N}\hat{L}^{\top}\right)^{-1}\hat{W}\left(\hat{S} - \frac{1}{N}\hat{L}\right)^{-1}, \tag{13}$$

where:

- $\hat{\eta} = (\hat{\eta}_1^{\top}, \ldots, \hat{\eta}_N^{\top})^{\top} \in \mathbb{R}^{N(K+1)}$ stacks the individual parameter vectors $\hat{\eta}_i = (\hat{\alpha}_i, \hat{\lambda}_i^{\top})^{\top}$.
  $\hat{X}_i = (\hat{X}_{i1}, \ldots, \hat{X}_{iT})^{\top}$, where $\hat{X}_{it} = (1, \hat{\beta}_{it}^{\top})^{\top}$. Let $M_{\hat{G}} = I_T - \hat{G}(\hat{G}^{\top}\hat{G})^{-1}\hat{G}^{\top}$ denote the projection matrix onto the orthogonal complement of the space spanned by the estimated latent factors $\hat{G} = (\hat{\tilde{g}}_1, \ldots, \hat{\tilde{g}}_T)^{\top} \in \mathbb{R}^{T\times R}$.

- $\hat{S}$ is an $N(K+1) \times N(K+1)$ block-diagonal matrix with i-th diagonal block

$$\hat{S}_{ii} = \frac{1}{T}\,\hat{X}_i^\top M_{\hat{G}} \hat{X}_i \quad \text{for } i = 1, \ldots, N. \tag{14}$$

  Collecting these blocks gives $\hat{S} = \operatorname{diag}(\hat{S}_{11}, \ldots, \hat{S}_{NN})$.

- $\hat{L}$ is an $N(K+1) \times N(K+1)$ matrix with (i,j) block

$$\hat{L}_{ij} = \hat{a}_{ij}\,\frac{1}{T}\,\hat{X}_i^\top M_{\hat{G}} \hat{X}_j, \qquad \hat{a}_{ij} = (\hat{\beta}_i^*)^\top \left(\frac{\hat{G}^\top \hat{G}}{N}\right)^{-1} \hat{\beta}_j^*, \tag{15}$$

  where $\hat{\beta}^* = (\hat{\beta}_1^{*\top}, \ldots, \hat{\beta}_N^{*\top})^\top \in \mathbb{R}^{N \times R}$ collects the loadings on the latent factors.

- $\hat{W}$ is an $N(K+1) \times N(K+1)$ block-diagonal matrix $\hat{W} = \operatorname{diag}(\hat{W}_1, \ldots, \hat{W}_N)$. For each asset i,

$$\hat{W}_i = \left(\hat{\lambda}_i \odot \frac{M_{\hat{G}}\hat{X}_i}{T}\right)^\top T^{-1} \operatorname{diag}\!\big(\hat{H}_i^\top \hat{H}_i\big) \left(\hat{\lambda}_i \odot \frac{M_{\hat{G}}\hat{X}_i}{T}\right) + \left(\frac{1}{T}\,\hat{X}_i^\top M_{\hat{G}} \hat{X}_i\right) \hat{\sigma}^2. \tag{16}$$

  Here $\hat{\lambda}_i \in \mathbb{R}^K$ is the vector of estimated factor risk premia for asset i, and $\odot$ denotes the element-wise product between $\hat{\lambda}_i$ and the K slope columns of $M_{\hat{G}}\hat{X}_i/T$.

- $\hat{H}_i$ is a $T \times K$ matrix

$$\hat{Z}_{ij} = \big(1,\, \hat{\omega}_{ij}\big) \in \mathbb{R}^{T \times 2}, \tag{17}$$

$$\hat{d}_{ij} = \left(\frac{1}{T}\hat{Z}_{ij}^\top \hat{Z}_{ij}\right)^{-1} \left(\frac{1}{\sqrt{T}}\hat{Z}_{ij}^\top \hat{v}_{ij}\right), \tag{18}$$

$$\hat{h}_{ij} = \hat{Z}_{ij}\hat{d}_{ij} \in \mathbb{R}^T, \tag{19}$$

  and then set $\hat{H}_i = (\hat{h}_{i1}, \ldots, \hat{h}_{iK})$.

- $\hat{\sigma}^2$ is the sample variance of the pricing error $\hat{\epsilon}_{i,t+1}$, obtained from the regression

$$\hat{\sigma^2} = \frac{\sum_{t=1}^{T}\sum_{i=1}^{N}\left(r_{i,t+1}^e - \hat{X}_{it}\hat{\eta}_i - \beta_i^{*\top}\hat{\tilde{g}}_[t+1]\right)^2}{NT - N(K+1) - (N+T)R}$$

  appropriate degrees of freedom.

3.4 Test statistics

Using the asymptotic variance estimator, Galvão et al. propose the following Wald-type test statistics for assessing the homogeneity of intercepts and slope parameters across assets.

Let $\hat{\eta}_i = (\hat{\alpha}_i, \hat{\lambda}_i^\top)^\top$, $\hat{\eta}_\cdot = \frac{1}{N}\sum_{i=1}^N \hat{\eta}_i$, $\tilde{\eta} = (\hat{\eta}_1 - \hat{\eta}_\cdot,\ \ldots,\ \hat{\eta}_N - \hat{\eta}_\cdot) \in \mathbb{R}^{N(K+1)}$, and $\widehat{V}^{-1} = \widehat{\text{Avar}}\Big(\sqrt{T}(\hat{\eta} - \eta)\Big)^{-1}$ denote the inverse of the asymptotic variance estimator defined previously.

Joint intercept and slope homogeneity

To test the joint null hypothesis

$$H_0^{\alpha,\lambda}: \quad \alpha_1 = \cdots = \alpha_N = 0, \qquad \lambda_1 = \cdots = \lambda_N,$$

the proposed statistic is

$$\hat{\Gamma}_{\alpha,\lambda} = \frac{T\,\tilde{\eta}^\top \widehat{V}^{-1}\tilde{\eta} \;-\; \big[(N-1)K + N\big]}{\sqrt{2\big[(N-1)K + N\big]}}, \tag{33}$$

which satisfies $\hat{\Gamma}_{\alpha,\lambda} \xrightarrow{d} N(0,1)$ under $H_0^{\alpha,\lambda}$.

Intercept homogeneity

For the null hypothesis

$$H_0^{\alpha}: \quad \alpha_1 = \cdots = \alpha_N = 0,$$

let $\tilde{\alpha}$ denote the subvector of $\tilde{\eta}$ containing only the intercept parameters, and let $\widehat{V}_\alpha^{-1}$ be the corresponding submatrix of $\widehat{V}^{-1}$. The test statistic is

$$\hat{\Gamma}_{\alpha} = \frac{T\,\tilde{\alpha}^\top \widehat{V}_\alpha^{-1}\tilde{\alpha} \;-\; N}{\sqrt{2N}}, \tag{34}$$

which is asymptotically standard normal under H_0^{α}.

Slope homogeneity

For the null of slope homogeneity,

$$H_0^{\lambda}: \quad \lambda_1 = \cdots = \lambda_N,$$

let $\tilde{\lambda}$ be the slope subvector of $\tilde{\eta}$ and let $\widehat{V}_\lambda^{-1}$ denote the corresponding submatrix of $\widehat{V}^{-1}$. The Wald statistic is

$$\hat{\Gamma}_{\lambda} = \frac{T\,\tilde{\lambda}^\top \widehat{V}_\lambda^{-1}\tilde{\lambda} \;-\; (N-1)K}{\sqrt{2(N-1)K}}, \tag{35}$$

which satisfies $\hat{\Gamma}_{\lambda} \xrightarrow{d} N(0,1)$ under H_0^{λ}.

4 Results

4.1 Estimating factor loading

Appendix 2 contains the replication code for estimating factor loadings based on OLS estimates of the AR(1) equations as per Equation (5). Comparing our beta estimates against intermediate data provided by Galvao et al, we verify our replication implementation.

4.2 Solving for factor risk premia

We compare the two ways of estimating factor risk premia, with $\hat{\eta}_{pen}$ denoting the unconstrained penalty-based version and $\hat{\eta}_{iter}$ denoting the iterative approach discussed by Galvao et al. Appendix 3 shows that when applied to the empirical dataset, both approaches converge to the same objective value.

We also conduct MC simulations and find that both methods of computing the estimator yields similar RMSE, and that the estimators generally converge to the same value.

Table 1: Monte Carlo Results: Difference Between Estimators

	Mean	Std
$\|\hat{\eta}_{\text{pen}} - \hat{\eta}_{\text{iter}}\|$	0.135	0.063

Table 2: Monte Carlo Results: RMSE of Estimators Relative to True η

	$\hat{\eta}_{\text{pen}}$	$\hat{\eta}_{\text{iter}}$
RMSE	0.215	0.233

4.3 Empirical results

4.4 Asymptotic level and power of proposed tests

Table 3: Joint Homogeneity Test ($\Gamma_{\alpha,\lambda}$) with p-values

		K=1			K=3			K=5		
Period	Type	R=1	R=2	R=5	R=1	R=2	R=5	R=1	R=2	R = 5
1963–1983	γ	-18.91	-10.40	95.16	14.38	-4.72	10.36	-9.16	10.67	16.21
	p	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1963–2025	γ	44.21	-2.39	2441.88	-8.88	-5.89	35.92	-21.84	-6.51	-5.66
	p	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1973–1993	γ	33.95	-26.77	104.74	-11.07	-8.51	-7.71	-9.83	-8.75	31.60
	p	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1983–2003	γ	-30.28	-69.18	-156.72	1.94	46.65	622.59	-16.47	-6.37	0.81
	p	0.00	0.00	0.00	0.05	0.00	0.00	0.00	0.00	0.41
1993–2013	γ	-11.29	-62.27	55907.88	2.89	-4.10	51.99	-3.23	-3.99	246.17
	p	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2003–2023	γ	19.99	-4.49	5177.39	2.69	9.91	302.64	19.38	15.21	41.83
	p	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00

Table 4: Intercept Homogeneity Test (Γ_{α}) with p-values

		K=1			K=3			K=5		
Period	Type	R=1	R=2	R=5	R=1	R=2	R=5	R=1	R=2	R = 5
1963–1983	Γ	-5.29	-5.02	-0.80	-8.45	-8.32	-8.30	-10.73	-10.73	-10.43
	p	0.00	0.00	0.42	0.00	0.00	0.00	0.00	0.00	0.00
1963–2025	Γ	-4.71	-5.37	-4.46	-8.35	-8.33	-8.30	-10.68	-10.72	-10.71
	p	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1973–1993	Γ	-4.78	-4.86	-4.40	-8.31	-8.31	-8.30	-10.72	-10.72	-10.72
	p	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1983–2003	Γ	-4.87	-4.29	0.58	-8.29	-7.79	-1.22	-10.71	-10.72	-10.48
	p	0.00	0.00	0.56	0.00	0.00	0.22	0.00	0.00	0.00
1993–2013	Γ	-5.33	-6.50	-3.34	-8.32	-8.31	-8.32	-10.73	-10.73	-10.72
	p	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2003–2023	Γ	-3.99	-5.26	-4.41	-8.43	-8.29	-8.29	-10.73	-10.71	-10.72
	p	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 5: Slope Homogeneity Test (Γ_λ) with p-values

		K=1			K=3			K=5		
Period	Type	R=1	R=2	R=5	R=1	R=2	R=5	R=1	R=2	R = 5
1963–1983	Γ	-4.82	-4.76	-2.54	-8.68	-8.23	-7.90	-10.67	-10.51	-10.47
	p	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
1963–2025	Γ	-4.79	-4.80	9.89	-8.32	-8.30	-8.11	-10.64	-10.68	-10.67
	p	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1973–1993	Γ	-4.79	-4.80	-4.10	-8.30	-8.29	-8.28	-10.70	-10.68	-10.13
	p	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1983–2003	Γ	-4.80	-5.42	-2.94	-8.21	-7.23	-7.91	-10.77	-10.65	-10.43
	p	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1993–2013	Γ	-4.74	-5.26	992.96	-8.15	-8.19	-7.52	-10.61	-10.61	-7.19
	p	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2003–2023	Γ	-4.33	-4.76	121.80	-8.04	-8.08	-3.58	-10.29	-10.45	-10.00
	p	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00