Exercise 1 - Global Value Factors

Jan. 21, 2026

Q1. Use annual accounting data to create the Fama-French version of the book-to-market equity ratio. Show how the book-to-market breakpoints used to create the HML factor varies across the three countries.

Q1 Breakpoints by Country and Formation Year (June)

Obs	country	form_year	size_med	bm_p30	bm_p70
1	DNK	1988	535.796	5.946	10.336
2	DNK	1989	318.590	8.028	13.373
3	DNK	1990	449.960	4.233	7.984
4	DNK	1991	393.263	4.242	8.307
5	DNK	1992	329.606	4.431	8.672
6	DNK	1993	305.105	5.445	9.446
7	DNK	1994	132.535	5.293	10.513
8	DNK	1995	183.993	4.198	9.952
25	DNK	2012	43.795	4.114	9.816
26	DNK	2013	63.150	3.791	9.774
27	DNK	2014	89.433	2.831	7.237
28	DNK	2015	80.734	2.390	8.132
29	DNK	2016	112.615	2.112	8.218
30	DNK	2017	164.415	1.955	7.813
31	DNK	2018	130.580	1.703	6.656
32	DNK	2019	112.934	2.198	7.624
33	DNK	2020	139.536	1.741	7.295
34	DNK	2021	257.925	1.228	5.971
35	DNK	2022	115.610	1.282	5.076
36	DNK	2023	110.464	2.306	6.969
37	DNK	2024	121.983	2.107	6.385
38	DNK	2025	195.889	2.550	7.108
39	JPN	1988	511.489	29.854	53.525
40	JPN	1989	555.905	27.523	48.229
41	JPN	1990	689.228	25.941	43.644
42	JPN	1991	562.455	41.768	70.412
43	JPN	1992	394.946	43.205	70.195
44	JPN	1993	562.082	58.330	96.184
45	JPN	1994	521.669	51.388	83.967
46	JPN	1995	356.405	41.552	64.126
47	JPN	1996	307.766	45.168	73.236
48	JPN	1997	193.099	57.959	94.724
49	JPN	1998	95.079	109.671	221.193
50	JPN	1999	129.347	99.045	197.524

51	JPN	2000	124.167	75.258	166.804
52	JPN	2001	97.286	94.389	207.197
53	JPN	2002	80.375	114.360	250.574
54	JPN	2003	82.941	112.709	245.193
55	JPN	2004	144.308	77.842	166.081
56	JPN	2005	158.872	61.920	126.697
57	JPN	2006	167.148	49.632	105.138
58	JPN	2007	130.786	63.828	129.483
59	JPN	2008	100.822	77.748	153.206
60	JPN	2009	90.790	95.895	182.060
61	JPN	2010	94.552	90.586	173.550
62	JPN	2011	110.273	76.661	147.000
63	JPN	2012	114.156	81.269	151.504
64	JPN	2013	121.992	82.247	153.754
65	JPN	2014	147.059	71.411	148.098
66	JPN	2015	142.967	75.259	163.099
77	USA	1960	262.703	0.671	0.671
78	USA	1961	82.261	0.767	1.266
79	USA	1962	296.735	0.372	0.464
80	USA	1963	121.983	0.433	0.885
81	USA	1964	194.357	0.414	0.788
82	USA	1965	179.322	0.385	0.749
83	USA	1966	124.235	0.377	0.669
84	USA	1967	69.118	0.485	0.939
85	USA	1968	83.125	0.316	0.660
86	USA	1969	69.868	0.273	0.549
87	USA	1970	35.526	0.396	0.838
88	USA	1971	48.544	0.512	1.008
89	USA	1972	51.812	0.456	0.908
90	USA	1973	24.008	0.445	0.948
91	USA	1974	19.247	0.817	1.746
92	USA	1975	22.830	1.348	2.705
93	USA	1976	25.711	0.989	1.902
94	USA	1977	29.345	0.787	1.470
95	USA	1978	37.515	0.797	1.455
96	USA	1979	38.906	0.793	1.486
97	USA	1980	36.902	0.639	1.356
98	USA	1981	47.215	0.480	1.297
99	USA	1982	30.604	0.536	1.291
100	USA	1983	63.883	0.488	1.146
101	USA	1984	50.309	0.405	0.895
102	USA	1985	54.871	0.487	0.984

103	USA	1986	69.186	0.399	0.836
104	USA	1987	66.374	0.398	0.822
105	USA	1988	51.980	0.488	1.031
106	USA	1989	59.191	0.455	0.937
107	USA	1990	59.404	0.401	0.917
108	USA	1991	57.017	0.510	1.284
109	USA	1992	74.171	0.374	0.956
110	USA	1993	92.007	0.339	0.807
111	USA	1994	80.231	0.336	0.771
112	USA	1995	92.181	0.402	0.880
113	USA	1996	117.650	0.340	0.793
114	USA	1997	128.000	0.327	0.727
115	USA	1998	145.414	0.311	0.650
116	USA	1999	130.977	0.367	0.918
117	USA	2000	137.710	0.317	1.001
118	USA	2001	138.771	0.426	1.281
119	USA	2002	160.669	0.419	1.082
120	USA	2003	194.817	0.509	1.204
121	USA	2004	299.640	0.354	0.779
122	USA	2005	323.532	0.335	0.746
123	USA	2006	349.876	0.353	0.811
124	USA	2007	407.131	0.341	0.751
125	USA	2008	282.504	0.383	0.970
126	USA	2009	224.643	0.632	1.936
127	USA	2010	319.788	0.472	1.303
128	USA	2011	450.306	0.415	1.100

129	USA	2012	416.212	0.477	1.333
130	USA	2013	507.228	0.443	1.184
131	USA	2014	619.430	0.346	0.964
132	USA	2015	595.608	0.351	1.052
133	USA	2016	526.600	0.398	1.192
134	USA	2017	645.172	0.375	1.069
135	USA	2018	695.120	0.353	1.000
136	USA	2019	578.612	0.437	1.253
137	USA	2020	484.464	0.373	1.120
138	USA	2021	771.004	0.318	1.152
139	USA	2022	497.353	0.316	0.995
140	USA	2023	469.321	0.452	1.392
141	USA	2024	503.430	0.401	1.325
142	USA	2025	431.367	0.421	1.546

For each country, we compute book-to-market (BE/ME) using annual accounting data and December market equity from the prior year. Following the Fama–French convention, I form portfolios each June and calculate the BM breakpoints (30th and 70th percentiles) within each country-year.

The BM breakpoints vary across countries and over time, reflecting differences in the distribution of book-to-market ratios in each market.

```
libname scratch "/scratch/yale/data_output";

/* ---------- helper: check whether a variable exists ---------- */
%macro var_exists(ds, var);
  %local dsid vnum rc;
  %let dsid = %sysfunc(open(&ds));
  %if &dsid %then %do;
    %let vnum = %sysfunc(varnum(&dsid, &var));
    %let rc   = %sysfunc(close(&dsid));
    %if &vnum > 0 %then 1;
    %else 0;
  %end;
  %else 0;
%mend;
```

```
/* ---------- Step 1: Build Book Equity (BE) from WORK.FUNDA_ANNUAL_3CTY ----------
*/
/* Uses existing BE/be if present; otherwise computes:
   BE = SHE + TXDITC - PSTK
   PSTK = coalesce(pstkrv, pstkl, pstk)
   SHE  = coalesce(seq, ceq + pstk)
*/
%let has_BE = %var_exists(work.funda_annual_3cty, BE);
%let has_be = %var_exists(work.funda_annual_3cty, be);

data work.acct_be;
  set work.funda_annual_3cty;

  length country $3 gvkey_c $6;
  country = excntry;

  /* standardize gvkey to character */
  if vtype(gvkey)='N' then gvkey_c = put(gvkey, z6.);
  else gvkey_c = strip(gvkey);

  /* BE */
  %if &has_BE %then %do;
    be_ff = BE;
  %end;
  %else %if &has_be %then %do;
    be_ff = be;
  %end;
  %else %do;
    pstk_ff = coalesce(pstkrv, pstkl, pstk);
    she_ff  = coalesce(seq, (ceq + pstk_ff));
    be_ff   = she_ff + coalesce(txditc,0) - coalesce(pstk_ff,0);
  %end;

  if country in ("DNK","JPN","USA");
  if be_ff > 0;

  keep gvkey_c country fyear be_ff;
run;
```

```
/* ---------- Step 2: Pull December ME and June ME from scratch.world_msf ---------- */
data work.msf_core;
  set scratch.world_msf(keep=excntry eom me gvkey);

  length country $3 gvkey_c $6;
  country = excntry;
  date    = eom;

  if vtype(gvkey)='N' then gvkey_c = put(gvkey, z6.);
  else gvkey_c = strip(gvkey);

  if country in ("DNK","JPN","USA");
  if missing(date) then delete;
  if missing(me) then delete;

  y = year(date);
  m = month(date);

  format date yymmdd10.;
  keep country gvkey_c date y m me;
run;

/* December ME of year (t-1) */
data work.me_dec;
  set work.msf_core;
  if m=12;
  keep country gvkey_c y me;
  rename y  = calyear_dec
         me = me_dec;
run;

/* June ME of year t (formation month) */
data work.me_june;
  set work.msf_core;
  if m=6;
  keep country gvkey_c y date me;
  rename y    = form_year
         date = june_date
         me   = me_june;
run;
```

```
/* ---------- Step 3: BM at June t = BE(fyear=t-1) / ME(Dec t-1) ---------- */
proc sql;
  create table work.bm_june as
  select
    j.country,
    j.gvkey_c,
    j.form_year,
    j.june_date,
    j.me_june,
    a.be_ff,
    d.me_dec,
    (a.be_ff / d.me_dec) as bm
  from work.me_june as j
  left join work.acct_be as a
    on j.gvkey_c = a.gvkey_c
   and j.country = a.country
   and a.fyear   = (j.form_year - 1)
  left join work.me_dec as d
    on j.gvkey_c = d.gvkey_c
   and j.country = d.country
   and d.calyear_dec = (j.form_year - 1)
  where j.country in ("DNK","JPN","USA");
quit;

data work.bm_june_clean;
  set work.bm_june;
  if missing(bm) then delete;
  if missing(me_june) then delete;
run;

proc sort data=work.bm_june_clean; by country form_year; run;

/* ---------- Step 4: Breakpoints within each country-year ---------- */
/* BM 30/70 */
proc univariate data=work.bm_june_clean noprint;
  by country form_year;
  var bm;
  output out=work.bp_bm pctlpts=30 70 pctlpre=bm_p;
run;
```

```
/* Size median */
proc univariate data=work.bm_june_clean noprint;
  by country form_year;
  var me_june;
  output out=work.bp_size median=size_med;
run;

/* Merge breakpoints */
proc sql;
  create table work.q1_breakpoints as
  select
    s.country,
    s.form_year,
    s.size_med,
    b.bm_p30,
    b.bm_p70
  from work.bp_size as s
  left join work.bp_bm as b
    on s.country=b.country and s.form_year=b.form_year
  order by country, form_year;
quit;

title "Q1 Breakpoints by Country and Formation Year (June)";
proc print data=work.q1_breakpoints(obs=200); run;
```

Q2. Create a dataset of monthly return data. Show some basic summary statistics of the return data from the three countries.

Q2 Summary stats of monthly returns (ret) by country

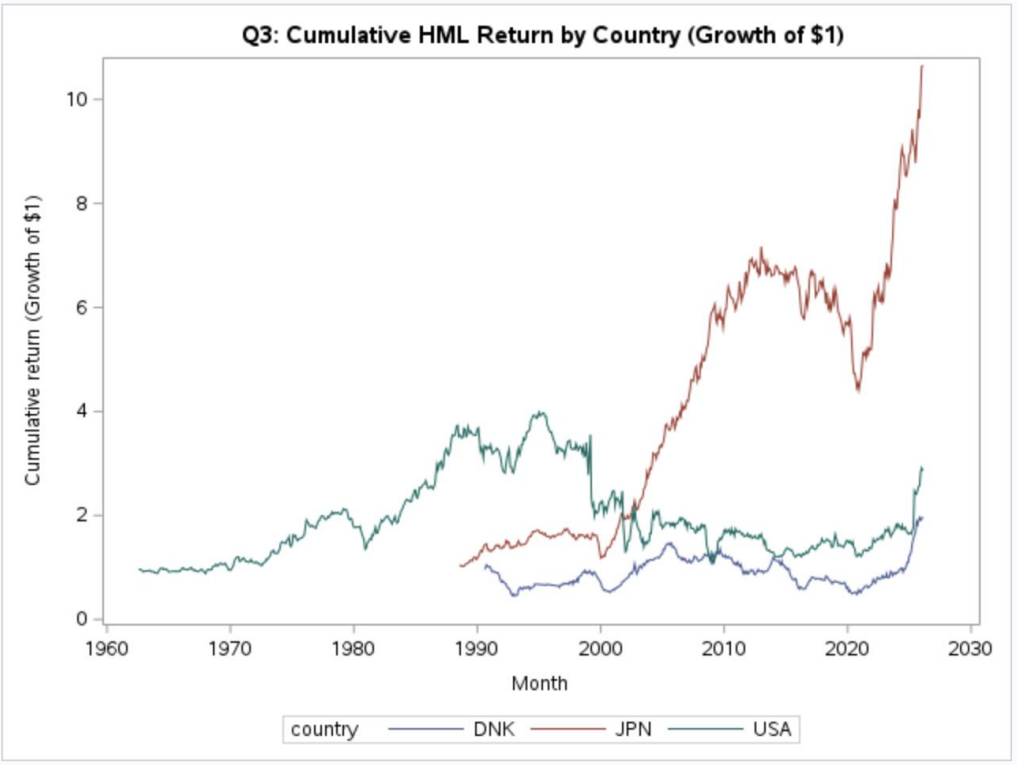
The MEANS Procedure

Analysis Variable : ret Returns											
country	N Obs	N	Mean	Std Dev	Minimum	1st Pctl	5th Pctl	50th Pctl	95th Pctl	99th Pctl	Maximum
DNK	82014	82014	0.0086386	0.1587341	-0.9998202	-0.3106560	-0.1590004	0.0043707	0.1776498	0.4006567	19.2064324
JPN	1570115	1570115	0.0085187	0.4326032	-0.9991472	-0.2722425	-0.1617418	-0.0014240	0.2005135	0.4319403	437.7032354
USA	9675385	9675385	889.6018709	2760034.16	-1.0000000	-0.3927589	-0.2019231	0.0014144	0.2344906	0.5562500	8585159999

```
libname scratch "/scratch/yale/data_output";

data work.q2_mret3;
  set scratch.world_msf(keep=excntry eom ret me gvkey permno);

  country = excntry;
  date    = eom;

  if country in ("DNK","JPN","USA");

  if missing(date) then delete;
  if missing(ret) then delete;

  format date yymmdd10.;
  keep country date ret me gvkey permno;
run;

proc print data=work.q2_mret3(obs=20);
run;

title "Q2 Summary stats of monthly returns (ret) by country";
proc means data=work.q2_mret3 n mean std min p1 p5 p50 p95 p99 max;
  class country;
  var ret;
run;

title "Q2 Sample coverage by country";
proc sql;
```

```
  select country,
        min(date) as start format=yymmdd10.,
        max(date) as end   format=yymmdd10.,
        count(*)  as n_obs
  from work.q2_mret3
  group by country;
quit;
```

Q3. Create a monthly return series of the HML factor in Denmark, Japan, and the US. Show a plot with the cumulative return of the factor in each of the three countries.



count
DNK
JPN
USA

Q3 Cumulative return of HML (DNK, JPN, USA)

Obs	country	date	hml
1	DNK	1990-07-31	-0.020029
2	DNK	1990-08-31	0.016609
3	DNK	1990-09-30	0.011642
4	DNK	1990-10-31	-0.001319
5	DNK	1990-11-30	-0.006373
6	DNK	1990-12-31	0.002899
7	DNK	1991-01-31	-0.009634
8	DNK	1991-02-28	-0.016651
9	DNK	1991-03-31	-0.015781
10	DNK	1991-04-30	-0.002020
11	DNK	1991-05-31	0.001045
12	DNK	1991-06-30	-0.002046

```
libname scratch "/scratch/yale/data_output";
```

```
%macro var_exists(ds, var);
  %local dsid vnum rc;
  %let dsid = %sysfunc(open(&ds));
  %if &dsid %then %do;
    %let vnum = %sysfunc(varnum(&dsid, &var));
    %let rc   = %sysfunc(close(&dsid));
    %if &vnum > 0 %then 1;
    %else 0;
  %end;
  %else 0;
%mend;

%let has_BE = %var_exists(work.funda_annual_3cty, BE);
%let has_be = %var_exists(work.funda_annual_3cty, be);

data work.acct_be;
  set work.funda_annual_3cty;

  length country $3 gvkey_c $6;
  country = excntry;

  if vtype(gvkey)='N' then gvkey_c = put(gvkey, z6.);
  else gvkey_c = strip(gvkey);

  %if &has_BE %then %do;
    be_ff = BE;
  %end;
  %else %if &has_be %then %do;
    be_ff = be;
  %end;
  %else %do;
    pstk_ff = coalesce(pstkrv, pstkl, pstk);
    she_ff  = coalesce(seq, (ceq + pstk_ff));
    be_ff   = she_ff + coalesce(txditc,0) - coalesce(pstk_ff,0);
  %end;

  if country in ("DNK","JPN","USA");
  if be_ff > 0;

  keep gvkey_c country fyear be_ff;
```

```
run;

data work.msf_raw;
  set scratch.world_msf(keep=excntry eom ret me gvkey);

  length country $3 gvkey_c $6;
  country = excntry;
  date    = eom;

  if vtype(gvkey)='N' then gvkey_c = put(gvkey, z6.);
  else gvkey_c = strip(gvkey);

  if country in ("DNK","JPN","USA");
  if missing(date) then delete;
  if missing(ret) then delete;
  if missing(me) then delete;
  if me <= 0 then delete;

  y = year(date);
  m = month(date);

  format date yymmdd10.;
  keep country gvkey_c date y m ret me;
run;

/* winsorize ret by country at 1% and 99% */
proc sort data=work.msf_raw; by country; run;

proc univariate data=work.msf_raw noprint;
  by country;
  var ret;
  output out=work.ret_cut pctlpts=1 99 pctlpre=ret_p;
run;

proc sql;
  create table work.msf_core as
  select
    a.*,
    c.ret_p1,
    c.ret_p99,
```

```
    case
      when a.ret < c.ret_p1 then c.ret_p1
      when a.ret > c.ret_p99 then c.ret_p99
      else a.ret
    end as ret_w
  from work.msf_raw as a
  inner join work.ret_cut as c
    on a.country=c.country;
quit;

data work.msf_core;
  set work.msf_core;
  drop ret_p1 ret_p99 ret;
  rename ret_w = ret;
run;

/* December ME (t-1) */
data work.me_dec;
  set work.msf_core;
  if m=12;
  keep country gvkey_c y me;
  rename y  = calyear_dec
         me = me_dec;
run;

/* June ME (t) */
data work.me_june;
  set work.msf_core;
  if m=6;
  keep country gvkey_c y date me;
  rename y    = form_year
         date = june_date
         me   = me_june;
run;

/* BM at June t */
proc sql;
  create table work.bm_june as
  select
    j.country,
```

```
    j.gvkey_c,
    j.form_year,
    j.june_date,
    j.me_june,
    a.be_ff,
    d.me_dec,
    (a.be_ff / d.me_dec) as bm
  from work.me_june as j
  left join work.acct_be as a
    on j.gvkey_c = a.gvkey_c
   and j.country = a.country
   and a.fyear   = (j.form_year - 1)
  left join work.me_dec as d
    on j.gvkey_c = d.gvkey_c
   and j.country = d.country
   and d.calyear_dec = (j.form_year - 1)
  where j.country in ("DNK","JPN","USA");
quit;

data work.bm_june_clean;
  set work.bm_june;
  if missing(bm) then delete;
  if missing(me_june) then delete;
run;

/* If you already have work.q1_breakpoints from Q1, use it.
   Otherwise, uncomment the next block to recompute breakpoints from bm_june_clean. */

/*
proc sort data=work.bm_june_clean; by country form_year; run;

proc univariate data=work.bm_june_clean noprint;
  by country form_year;
  var bm;
  output out=work.bp_bm pctlpts=30 70 pctlpre=bm_p;
run;

proc univariate data=work.bm_june_clean noprint;
  by country form_year;
  var me_june;
```

```
  output out=work.bp_size median=size_med;
run;

proc sql;
  create table work.q1_breakpoints as
  select
    s.country,
    s.form_year,
    s.size_med,
    b.bm_p30,
    b.bm_p70
  from work.bp_size as s
  left join work.bp_bm as b
    on s.country=b.country and s.form_year=b.form_year
  order by country, form_year;
quit;
*/

proc sql;
  create table work.june_labeled as
  select
    b.country,
    b.gvkey_c,
    b.form_year,
    b.june_date,
    b.me_june,
    b.bm,
    bp.size_med,
    bp.bm_p30,
    bp.bm_p70,
    case when b.me_june <= bp.size_med then "S" else "B" end as size_grp length=1,
    case
      when b.bm <= bp.bm_p30 then "L"
      when b.bm <= bp.bm_p70 then "M"
      else "H"
    end as bm_grp length=1,
    cats(calculated size_grp, "/", calculated bm_grp) as port length=3
  from work.bm_june_clean as b
  inner join work.q1_breakpoints as bp
    on b.country=bp.country and b.form_year=bp.form_year
```

```
  where b.country in ("DNK","JPN","USA");
quit;

proc sql;
  create table work.hold_panel as
  select
    m.country,
    m.gvkey_c,
    m.date,
    m.ret,
    j.port,
    j.me_june as weight_me,
    j.june_date
  from work.msf_core as m
  inner join work.june_labeled as j
    on m.country=j.country and m.gvkey_c=j.gvkey_c
  where m.date >  intnx('month', j.june_date, 0, 'e')
    and m.date <= intnx('month', j.june_date, 12, 'e');
quit;

proc sort data=work.hold_panel; by country date port; run;

proc summary data=work.hold_panel nway;
  class country date port;
  var ret;
  weight weight_me;
  output out=work.port_vwret(drop=_type_ _freq_) mean=vwret;
run;

proc sql;
  create table work.hml as
  select
    a.country,
    a.date,
    (0.5*(sh.vwret + bh.vwret) - 0.5*(sl.vwret + bl.vwret)) as hml
  from (select distinct country, date from work.port_vwret) as a
  left join work.port_vwret as sh
    on a.country=sh.country and a.date=sh.date and sh.port="S/H"
  left join work.port_vwret as bh
    on a.country=bh.country and a.date=bh.date and bh.port="B/H"
```

```
  left join work.port_vwret as sl
    on a.country=sl.country and a.date=sl.date and sl.port="S/L"
  left join work.port_vwret as bl
    on a.country=bl.country and a.date=bl.date and bl.port="B/L"
  where calculated hml is not missing
  order by country, date;
quit;

data work.hml_cum;
  set work.hml;
  by country;
  retain cum 1;
  if first.country then cum=1;
  if 1+hml <= 0 then delete;
  cum = cum * (1 + hml);
  cumret = cum - 1;
  format date yymmn6.;
run;

title "Q3 Cumulative return of HML (DNK, JPN, USA)";
proc sgplot data=work.hml_cum;
  series x=date y=cumret / group=country;
  xaxis label="Month";
  yaxis label="Cumulative return";
run;

proc print data=work.hml(obs=12);
  var country date hml;
run;

proc means data=work.hml n mean std min p5 p50 p95 max;
  class country;
  var hml;
run;
```

Q4. Compare and contrast the HML factors you have created, with the book-to-market portfolios from Jensen et al. (2021)1. Discuss the differences.

Q4 Correlation: My HML vs Professor/Jensen series

The CORR Procedure

country=DNK

2 Variables:	my_hml jensen_ret

Pearson Correlation Coefficients, N = 414 Prob > |r| under H0: Rho=0		
	my_hml	**jensen_ret**
my_hml	1.00000	0.45787 <.0001
jensen_ret	0.45787 <.0001	1.00000

Q4 Correlation: My HML vs Professor/Jensen series

The CORR Procedure

country=JPN

2 Variables:	my_hml jensen_ret

Pearson Correlation Coefficients, N = 438 Prob > |r| under H0: Rho=0		
	my_hml	**jensen_ret**
my_hml	1.00000	0.59845 <.0001
jensen_ret	0.59845 <.0001	1.00000

Q4 Correlation: My HML vs Professor/Jensen series

The CORR Procedure

country=USA

2 Variables:	my_hml jensen_ret

Pearson Correlation Coefficients, N = 750 Prob > |r| under H0: Rho=0		
	my_hml	**jensen_ret**
my_hml	1.00000	0.48834 <.0001
jensen_ret	0.48834 <.0001	1.00000

We found that our HML factor is positively related to professors' series in all three countries, where there exist moderate correlations for DNK (0.458), JPN (0.598), and USA (0.488), and we found these relationships are statistically significant ($p < 0.0001$). This indicates that our construction captures a similar value-related signal, but it does not replicate professors' series exactly.

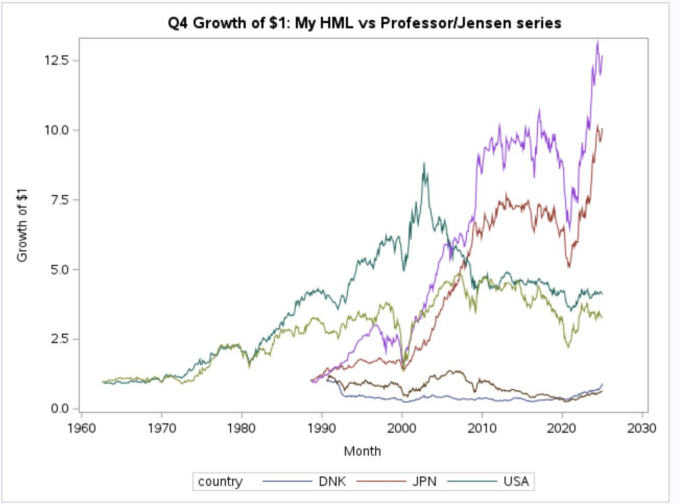
Q4 Summary stats: My HML vs Professor/Jensen series

The MEANS Procedure

country	N Obs	Variable	N	Mean	Std Dev	Minimum	5th Pctl	50th Pctl	95th Pctl	Maximum
DNK	414	my_hml	414	0.000603522	0.0408031	-0.1679981	-0.0626880	0.000238307	0.0622547	0.1162910
		jensen_ret	414	0.000281922	0.0513552	-0.1572532	-0.0785665	-0.0018158	0.0865553	0.2007694
JPN	438	my_hml	438	0.0056191	0.0258660	-0.0754107	-0.0336912	0.0052466	0.0500525	0.1155683
		jensen_ret	438	0.0064004	0.0343824	-0.1292098	-0.0479237	0.0059424	0.0616434	0.1961954
USA	750	my_hml	750	0.0021915	0.0247105	-0.0986747	-0.0354650	0.0014856	0.0448880	0.0986962
		jensen_ret	750	0.0022416	0.0362218	-0.2921344	-0.0463508	-5.345214E-6	0.0563773	0.2049141

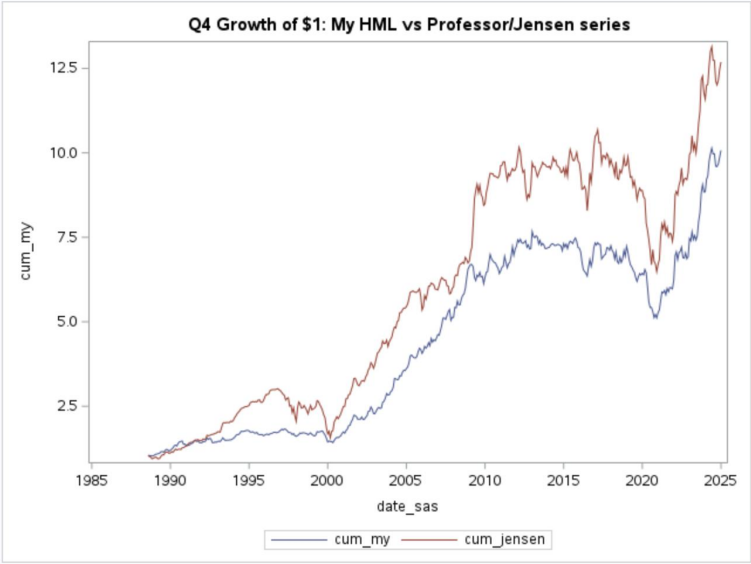
We also found that the average monthly returns are close across the two series within each country, where DNK has a mean of 0.00060 for our HML versus 0.00028 for the professors, JPN has 0.00562 versus 0.00640, and USA has 0.00219 versus 0.00224. This suggests that the long-run value premium implied by both series is similar, even though month-to-month movements differ.

A key difference is that the professors' series is consistently more volatile and more extreme. Standard deviations are higher for professors' series in Denmark (0.0514 vs 0.0408), Japan (0.0344 vs 0.0259), and the USA (0.0362 vs 0.0247). The tails are also heavier for the professors' series.



DNK

USA



JPN

The cumulative comparisons highlight two country-specific patterns that matter. In Japan, the professors' series finishes at a higher cumulative level than our series, which is consistent with its slightly higher mean and higher volatility. In the USA, the two cumulative lines track each other over long horizons but separate during volative market, which aligns with the much more negative minimum in the professors' series. These plots are most informative because they match the summary statistics and clarify where the tracking error occurs.

```
%let folder = /home/yale/seraphim/GlobalFactors;
```

```
%let ctry1 = DNK;
%let ctry2 = JPN;
%let ctry3 = USA;

/* --- Import the three CSVs --- */
proc import datafile="&folder./1.csv" out=work.j1 dbms=csv replace;
  guessingrows=max;
run;

proc import datafile="&folder./2.csv" out=work.j2 dbms=csv replace;
  guessingrows=max;
run;

proc import datafile="&folder./3.csv" out=work.j3 dbms=csv replace;
  guessingrows=max;
run;


%macro prep(in=, out=, country=);
data &out.;
  set &in.;
  length country $3;
  country = "&country.";

  /* date: handle char or numeric */
  if vtype(date)='C' then date_sas = input(date, yymmdd10.);
  else date_sas = date;

  format date_sas yymmdd10.;

  /* return: handle ret or vwret */
  jensen_ret = .;
  if nmiss(ret)=0 then jensen_ret = ret;
  else if nmiss(vwret)=0 then jensen_ret = vwret;

  keep country date_sas jensen_ret;
run;
%mend;

%prep(in=work.j1, out=work.jensen1, country=&ctry1);
```

```
%prep(in=work.j2, out=work.jensen2, country=&ctry2);
%prep(in=work.j3, out=work.jensen3, country=&ctry3);

data work.jensen_all;
  set work.jensen1 work.jensen2 work.jensen3;
  if missing(date_sas) then delete;
  if missing(jensen_ret) then delete;
run;


data work.my_hml;
  set work.hml;
  date_sas = date;
  format date_sas yymmdd10.;
  keep country date_sas hml;
run;

proc sort data=work.jensen_all; by country date_sas; run;
proc sort data=work.my_hml;     by country date_sas; run;

/* --- Merge on overlapping months --- */
data work.q4_compare;
  merge work.my_hml(in=a rename=(hml=my_hml))
        work.jensen_all(in=b);
  by country date_sas;
  if a and b;
run;


title "Q4 Overlap check";
proc sql;
  select country, count(*) as n_months
  from work.q4_compare
  group by country;
quit;

/* --- Correlation by country --- */
title "Q4 Correlation: My HML vs Professor/Jensen series";
proc sort data=work.q4_compare; by country; run;
```

```
proc corr data=work.q4_compare nosimple;
  by country;
  var my_hml jensen_ret;
run;

/* --- Mean/volatility comparison --- */
title "Q4 Summary stats: My HML vs Professor/Jensen series";
proc means data=work.q4_compare n mean std min p5 p50 p95 max;
  class country;
  var my_hml jensen_ret;
run;


data work.q4_cum;
  set work.q4_compare;
  by country;
  retain cum_my cum_jensen 1;

  if first.country then do;
    cum_my = 1;
    cum_jensen = 1;
  end;

  if (1+my_hml) > 0 then cum_my = cum_my * (1 + my_hml);
  if (1+jensen_ret) > 0 then cum_jensen = cum_jensen * (1 + jensen_ret);

  format date_sas yymmn6.;
run;

title "Q4 Growth of $1: My HML vs Professor/Jensen series";
proc sgplot data=work.q4_cum;
  series x=date_sas y=cum_my     / group=country;
  series x=date_sas y=cum_jensen / group=country;
  xaxis label="Month";
  yaxis label="Growth of $1";
Run;



proc sgplot data=work.q4_cum;
```

```
  where country="DNK";
  series x=date_sas y=cum_my;
  series x=date_sas y=cum_jensen;
run;

proc sgplot data=work.q4_cum;
  where country="JPN";
  series x=date_sas y=cum_my;
  series x=date_sas y=cum_jensen;
run;

proc sgplot data=work.q4_cum;
  where country="USA";
  series x=date_sas y=cum_my;
  series x=date_sas y=cum_jensen;
run;

title "Q4 Growth of $1: My HML vs Professor/Jensen series (by country)";
proc sgpanel data=work.q4_cum;
  panelby country / columns=3 novarname;
  series x=date_sas y=cum_my     / lineattrs=(pattern=solid)    legendlabel="My HML";
  series x=date_sas y=cum_jensen / lineattrs=(pattern=shortdash)
legendlabel="Professor/Jensen";
  colaxis label="Month";
  rowaxis label="Growth of $1";
run;
```