Linear Prediction of Stock Market Averages

Basic Problem

(a)

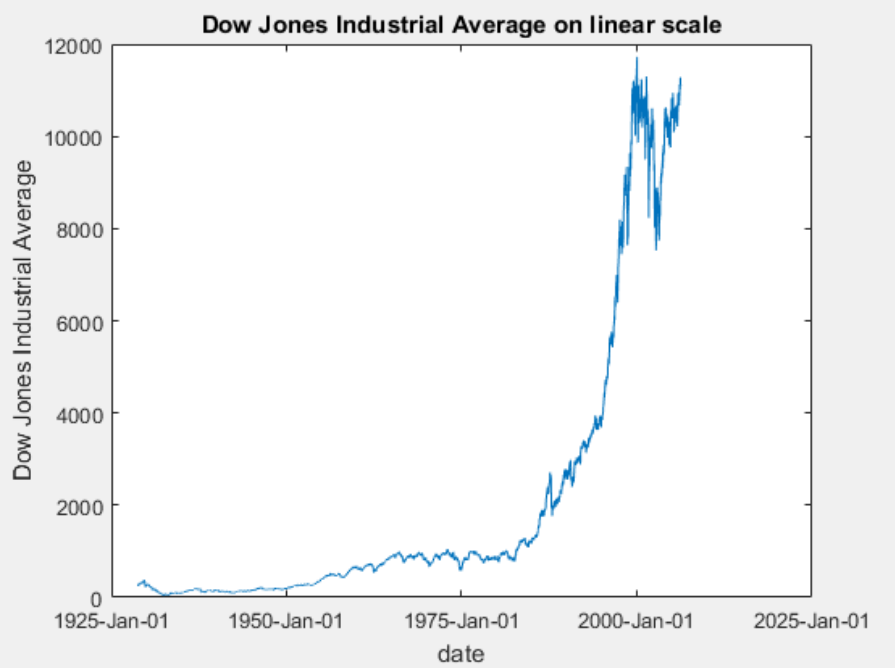

Figure 1. Dow Jones of full data set on linear scale.

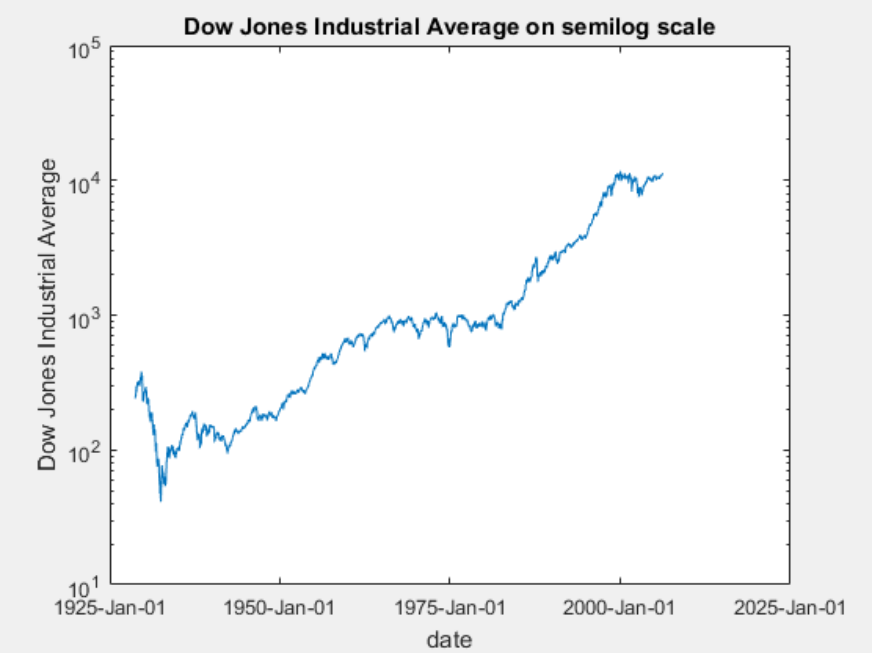

Figure 2. Dow Jones of full data set on semilog scale.

Ending amounts for $1000 investment using DJIA vs. 3% interest:

DJIA_end = invest_hold = 46337.2566960572
bank3_end = invest_interest = 10302.6812701947
equiv_rate = apr_needed_a = 0.049348(4.93%)

(b)

Predictor coefficients for first 10 years, p = 3:

a = [0.0267960474612419 0.0937731133773365 -1.11830844090208]

(c)

Plot of xhat1, xhat2, and the DJIA data. Note: xhat1 and xhat2 are identical.

xhat1/2 appears to trail the DJIA by xyz samples.

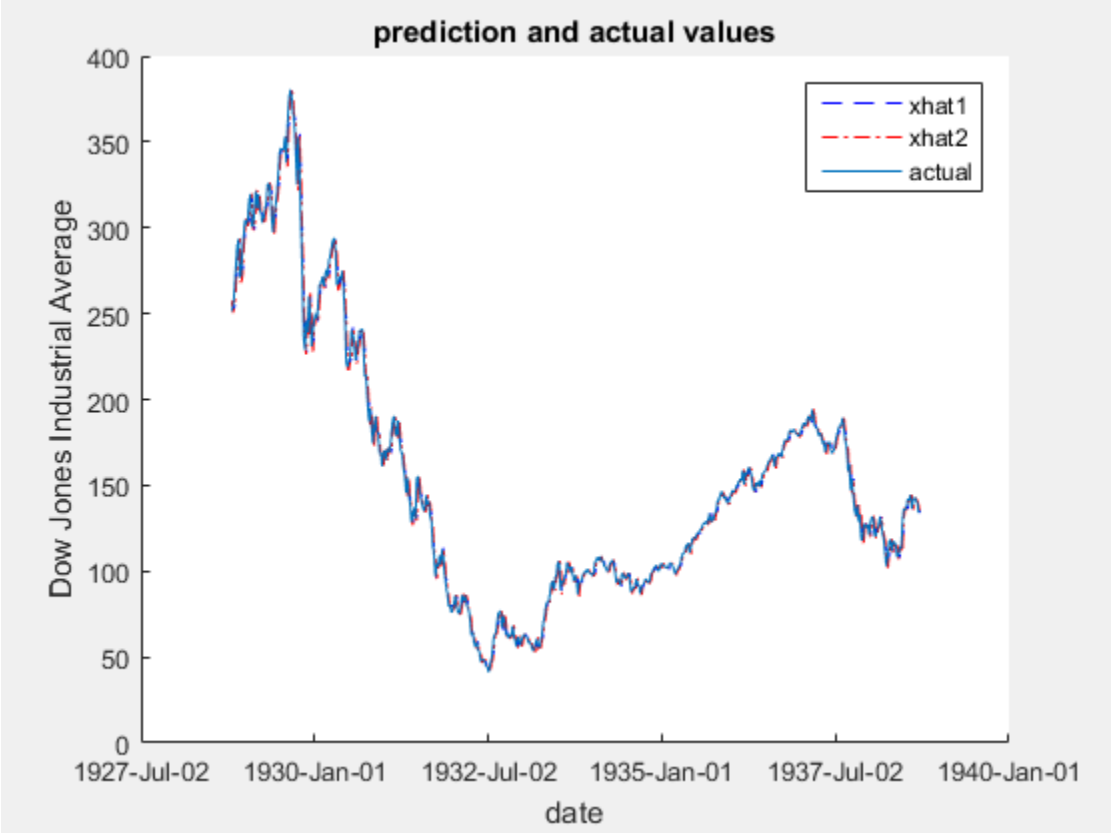

Figure 3. Prediction and actual values.

Expansion of the first “you can decide how much to zoom” or so weeks of the above data to show more detail, in particular the xyz-week “lag”:

The lag is 1 week.

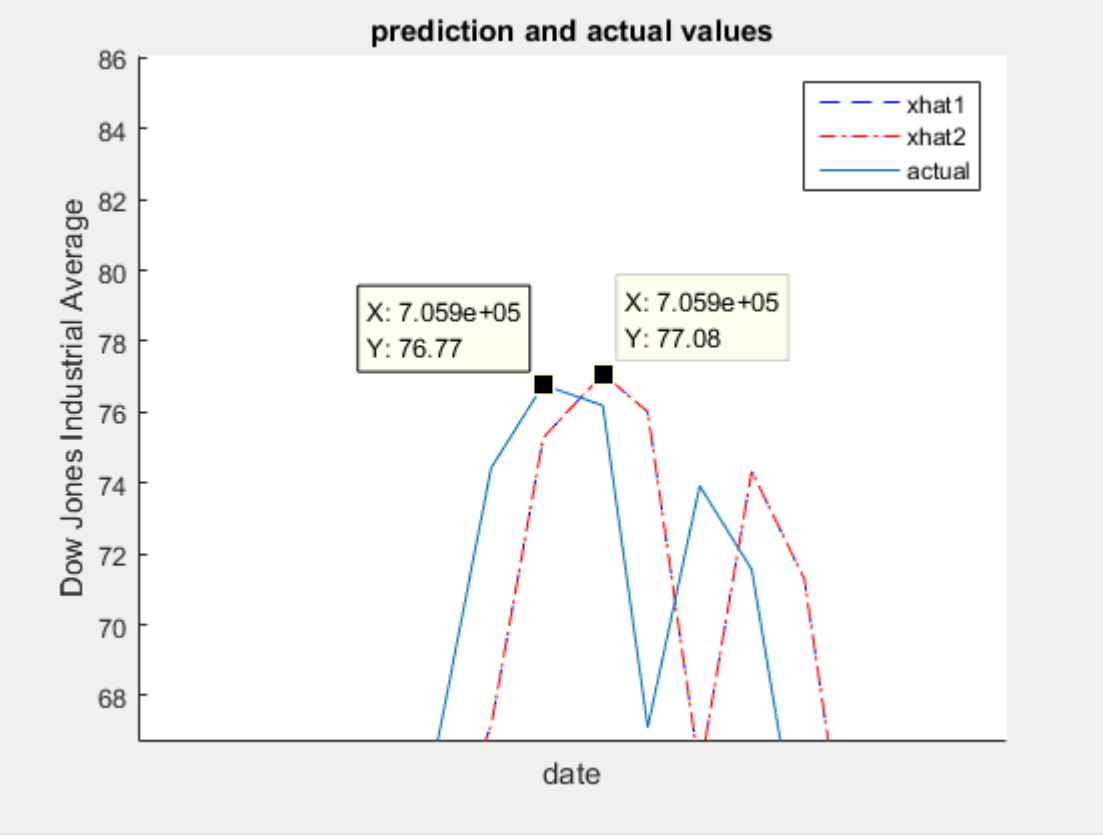

Figure 4. Zoom in to show week lag.

Total squared error:

E = E_1 = E_2 = 23638.0642056466

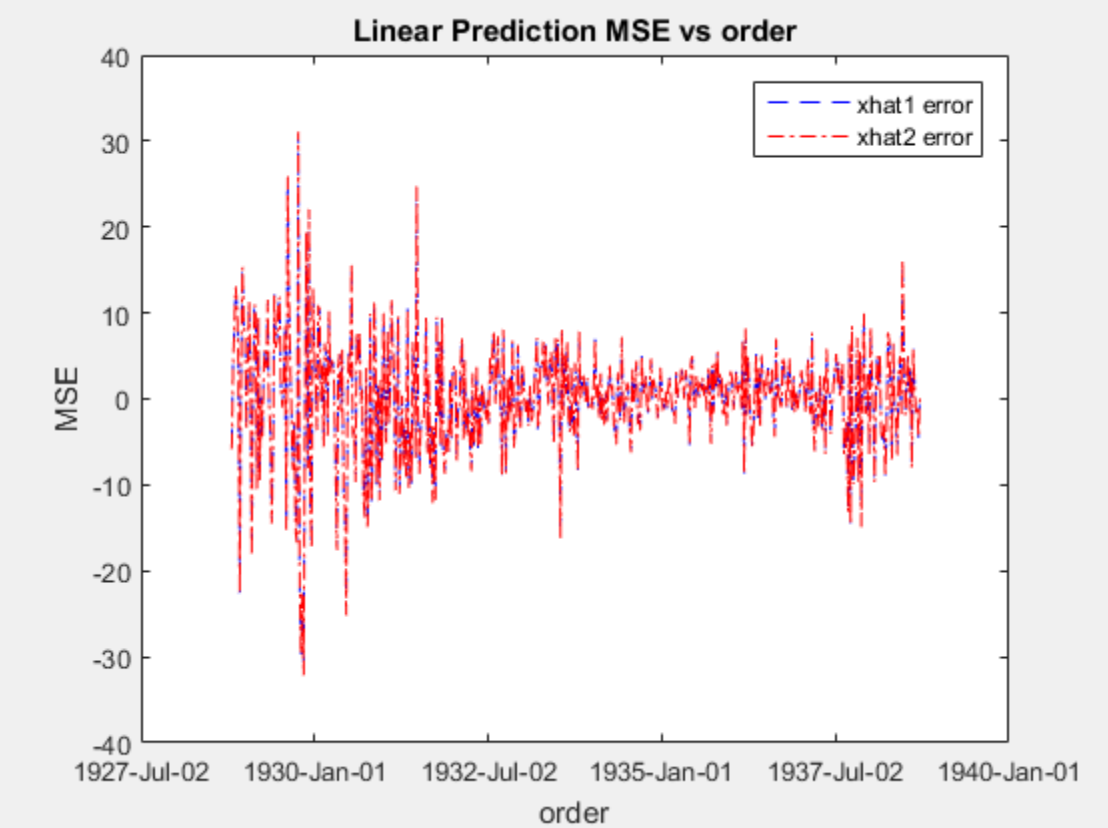

Figure 5. Identical xhat1 and xhat2 error.

(d)

Graph of error E vs. predictor order, going up to order p = 10. NOTE: this has been discussed in Piazza (Good job all of you in this discussion)

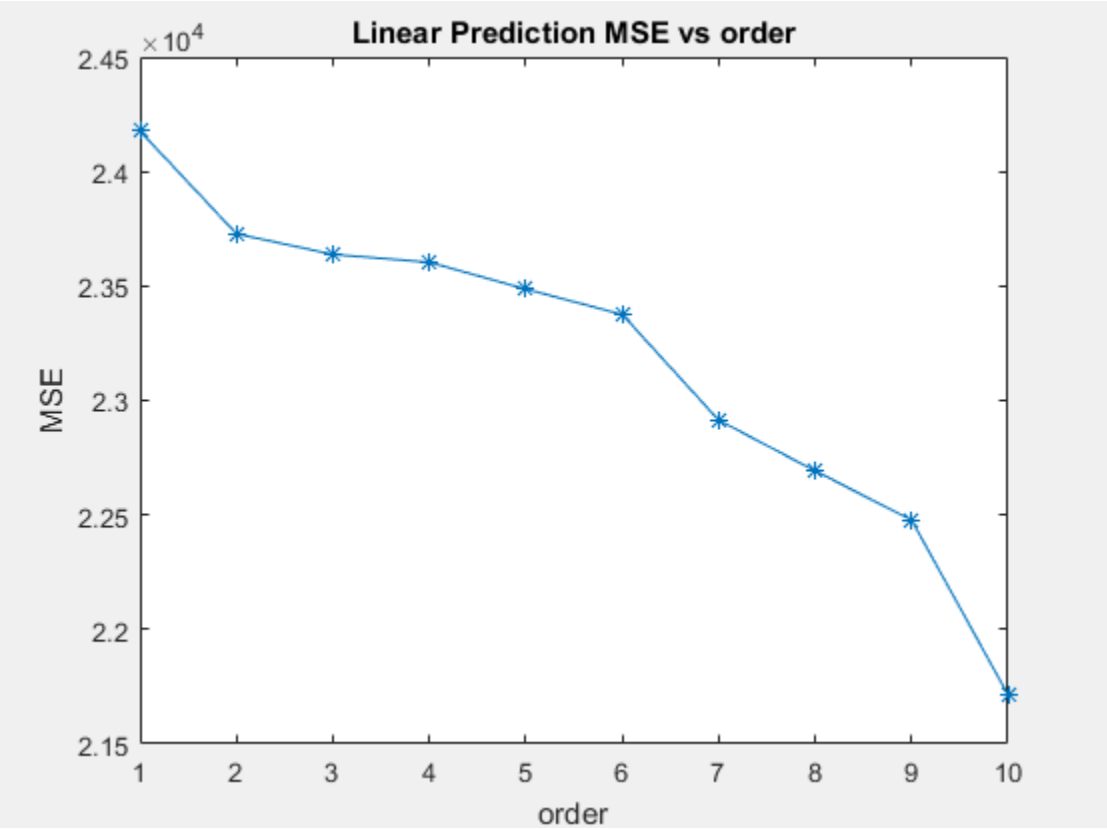

Figure 6. Linear Prediction MSE vs order.

Choose p = 10 in terms of smallest MSE

Intermediate Problem

(e)

The predictor coefficients for first 10 years, p=10:

List all the ten coefficients

[-0.1027 0.2133 -0.0338 0.0095 -0.0859 -0.0632 0.0925 0.0076 0.0686 -1.1036]

Investment results obtained:

Bank rate = 1349.7, equivalent rate = 3%

Upper bound = 4700565.98, equivalent rate = 85.25%

Buy and hold = 544.4383, equivalent rate = -6.08 %

Predictor = 1422.6, equivalent rate = 3.53%

(f)

Not sure about the definition of prediction strategy. I conducted two analysis. The first one trains a new predictor with the latest decade of data. The second one use the prediction coefficient from the first decade of data. Although the predictor from the first decade of data is not accurate for the latest decade prediction, the DJIA value is so large that even a small revenue count more than it predicts on the first decade of data.

Predictor coefficients for last 10 years, p=10:

List all the ten coefficients

[-0.0414 -0.0460 0.0221 0.0954 -0.1364 0.0610 0.0073 0.0733 -0.1101 -0.9264]

Investment results obtained:

Bank rate = 1349.7, equivalent rate = 3%

Upper bound = 167340.87, equivalent rate = 51.45%

Buy and hold = 2012.71, equivalent rate = 7.00 %

Predictor = 2620.28, equivalent rate = 9.64%

With predictor from the first decade

Investment results obtained:

Bank rate = 1349.7, equivalent rate = 3%

Upper bound = 167340.87, equivalent rate = 51.45%

Buy and hold = 2012.71, equivalent rate = 7.00 %

Predictor = 1475.54, equivalent rate = 3.89 %

Advanced Problem

(g)

In addition to the maximum gain, we also compute the gain with predictor from the first decade. We use this predictor across all data as a rule of thumb for our optimization.

We split the data into 8 pieces and the first seven of them are 520 weeks of data. We compute predictors for each block of data and the gain of using this kind of approach compared with using only the predictor from the first decade. Obviously, we see some improvement here. Detailed implementation can be shown from the code. With more blocks of data, we will likely approach the upper bound. So far, with 8 blocks of data, we can achieve half of the maximal gain.

Maximum Gain

Investment results obtained:

Bank rate = 10296.74, equivalent rate = 3%

Upper bound = 2.61e+19, equivalent rate = 48.84%

Buy and hold = 46337.26, equivalent rate = 4.93 %

Predictor from the first decade = 1.188e+19, equivalent rate = 47.81 %

Optimization

Predictor from the first decade = 1.343e+19, equivalent rate = 47.97 %

(h)

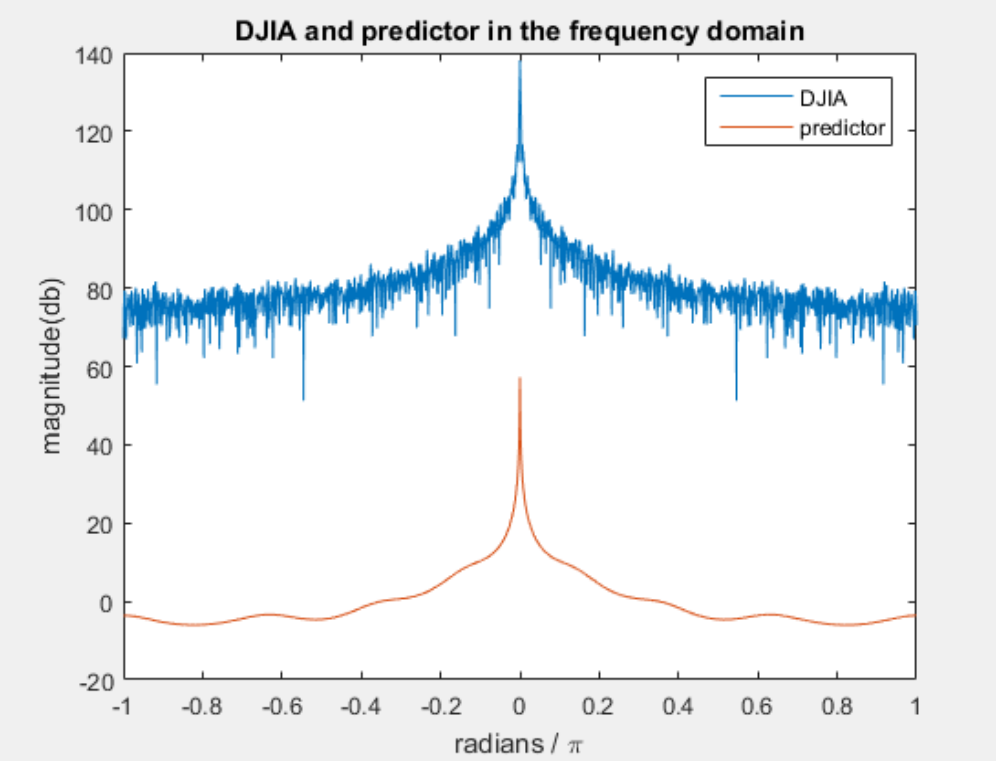

Figure 7. DJIA and predictor in the frequency domain.

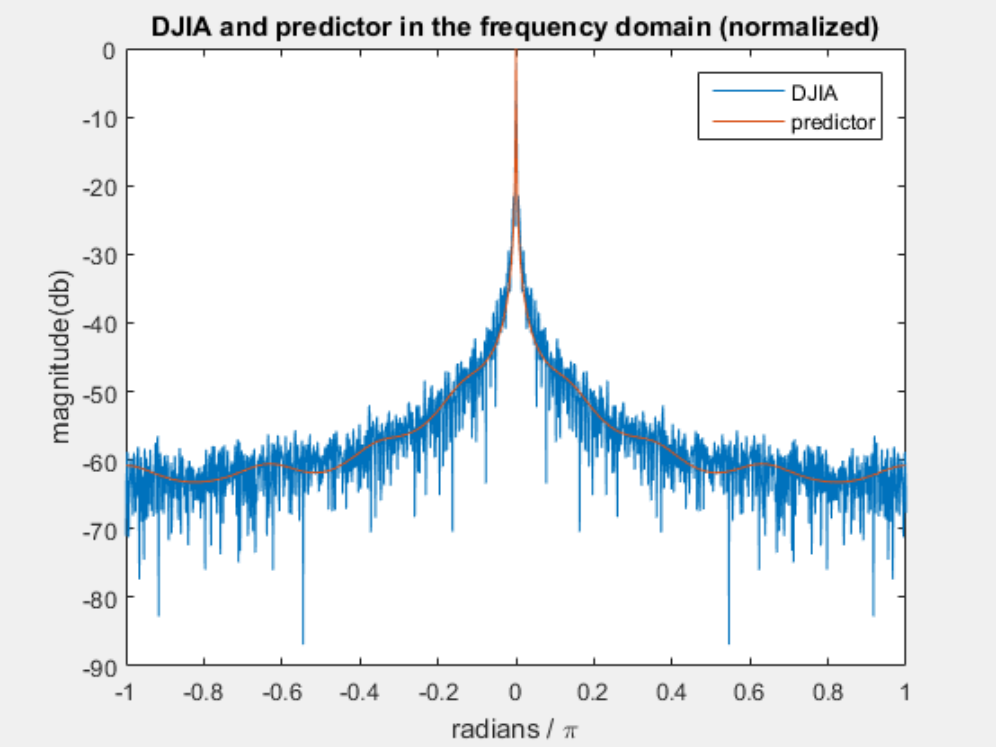

Figure 8. DJIA and predictor in the frequency domain (normalized).

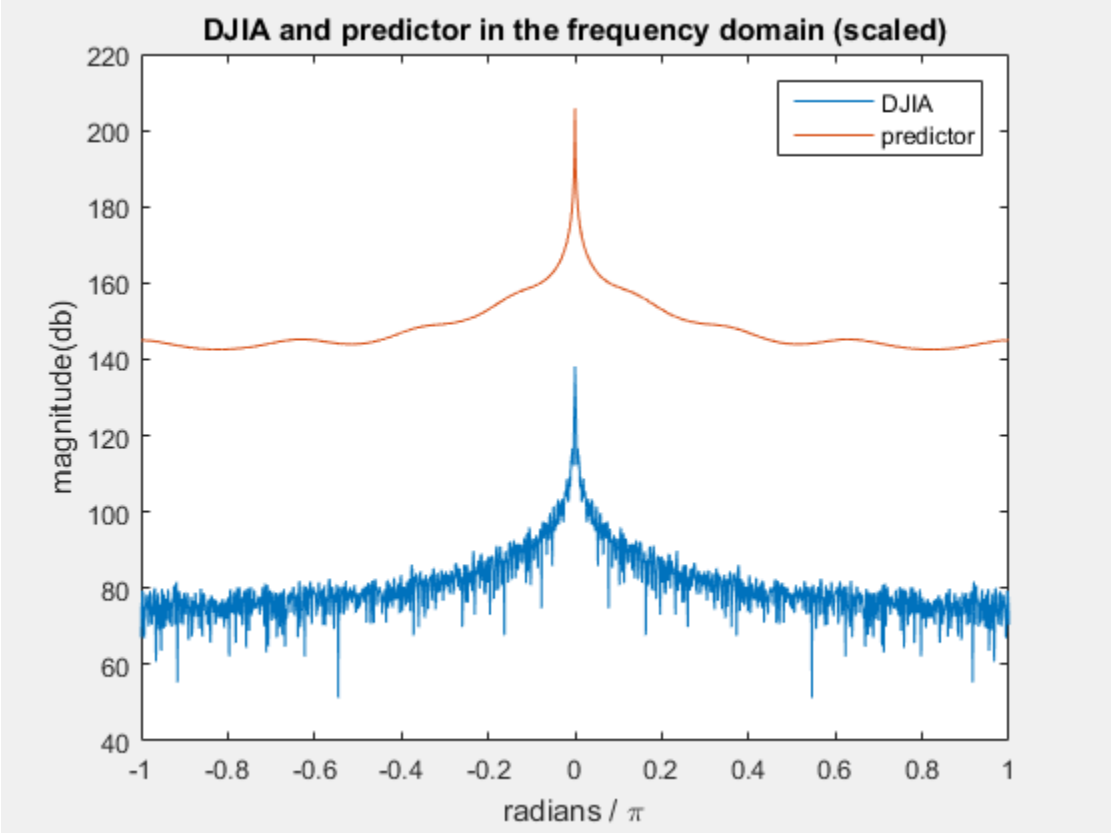

Figure 8. DJIA and predictor in the frequency domain (scaled).

G = 26909420.638

Since we know e[n] = x[n] +sum(ak*x[n-k]), then we can write it in z-domain as E(z) = X(z)A(z). A(z) here is just 1/(1+sum(ak*e(-jwk))). When we try to find the scale factor, we are also multiply X(z) with A(z) and the result is of course E(z) which is e[n] in the z-domain. With the Parseval's theorem, sum of square in the time domain is the same as the sum of square in the frequency domain. Then E=sum(e[n].^2) should be equal to E(z) in the frequency domain.

MATLAB CODE

```
load('djiaw_2006.mat')
initial_invest = 1000;
r = 0.03; % annual interest
interest = r/52; % weekly interest
weeks = 1:4044;
max_weeks = length(weeks);
date = djiaw(:,1);
djia = djiaw(:,2);
N = 520;
%% a
figure
plot(date,djia)
datetick('x','yyyy-mmm-dd')
xlabel('date')
ylabel('Dow Jones Industrial Average')
title('Dow Jones Industrial Average on linear scale')
figure
semilogy(date,djia)
datetick('x','yyyy-mmm-dd')
xlabel('date')
ylabel('Dow Jones Industrial Average')
title('Dow Jones Industrial Average on semilog scale')

invest_seq_hold = initial_invest;
invest_seq_interest = initial_invest;
invest_hold = initial_invest;
invest_interest = initial_invest;
for i=1:max_weeks-1
    invest_hold = invest_hold*djia(i+1)/djia(i);
    invest_interest = invest_interest*(1+interest);
    invest_seq_hold = [invest_seq_hold invest_hold];
    invest_seq_interest = [invest_seq_interest invest_interest];
end
invest_interest = invest_interest*(1+interest);
invest_seq_interest = [invest_seq_interest invest_interest];
apr_needed_a = (nthroot(invest_hold/initial_invest,max_weeks)-1)*52;

%% b
start_week = 1;
x = djia(start_week:N);
p = 3; r = 0;
[a, ~,~,X] = covpred(x, p, r);
%% c
xhat1 = -X*a(2:end);
e_1 = x(p+1+r:end) - xhat1;
xhat2 = -filter(a(2:end),1,x(1:end));
xhat2 = xhat2(p+r:end-1);
e_2 = x(p+1+r:end) - xhat2;
E_1 = sum(abs(e_1).^2);
E_2 = sum(abs(e_2).^2);
figure
hold on
```

```
plot(date(p+1+r:N),xhat1, 'b--')
plot(date(p+1+r:N),xhat2, 'r-.')
plot(date(p+1+r:N),x(p+1+r:end))
datetick('x','yyyy-mmm-dd')
legend('xhat1','xhat2','actual')
title('prediction and actual values')
xlabel('date')
ylabel('Dow Jones Industrial Average')
figure
plot(date(p+1+r:N),e_1,'b--')
hold on
plot(date(p+1+r:N),e_2, 'r-.')
datetick('x','yyyy-mmm-dd')
legend('xhat1 error','xhat2 error')
title('xhat and xhat2 error')
xlabel('date')
ylabel('MSE')
%% d
A = {};E = [];r=0;
p_arr=1:10;
start_week = 1;
for i=p_arr
    x = djia(start_week:N);
    [a, ~,~,X] = covpred(x, i, r);
    xhat1 = -X*a(2:end);
    e = x(i+1+r:end) - xhat1;
    A=[A a];
    E=[E;sum(abs(e).^2);];
end
figure
plot(p_arr,E,'*-');
title('Linear Prediction MSE vs order')
xlabel('order')
ylabel('MSE')

%% e
p=10;r=0;
start_week = 1;
x = djia(start_week:N);
[a, ~,~,~] = covpred(x, p, r);
X_e =[];
initial_invest = 1000;
for w = 0:N-1
    pred = djia(w+p:-1:w+1);
    pred_temp = - (pred')*(a(2:end));
    X_e = [X_e pred_temp];
end
invest_seq_hold_e = initial_invest;
invest_seq_interest_e = initial_invest;
invest_seq_djia_e = initial_invest;
invest_seq_pred_e = initial_invest;
invest_hold_e = initial_invest;
invest_interest_e = initial_invest;
invest_djia_e = initial_invest;
invest_pred_e = initial_invest;
for i=p+1:N+p
    invest_hold_e = invest_hold_e*djia(i)/djia(i-1);
```

```
    invest_interest_e = invest_interest_e*(1+interest);
    if(invest_djia_e*djia(i)/djia(i-1)>invest_djia_e*(1+interest))
        invest_djia_e = invest_djia_e*djia(i)/djia(i-1);
    else
        invest_djia_e = invest_djia_e*(1+interest);
    end

    if(invest_pred_e*X_e(i-p)/djia(i-1)>invest_pred_e*(1+interest))
        invest_pred_e = invest_pred_e*djia(i)/djia(i-1);
    else
        invest_pred_e = invest_pred_e*(1+interest);
    end

    invest_seq_hold_e = [invest_seq_hold_e invest_hold_e];
    invest_seq_interest_e = [invest_seq_interest_e invest_interest_e];
    invest_seq_djia_e = [invest_seq_djia_e invest_djia_e];
    invest_seq_pred_e = [invest_seq_pred_e invest_pred_e];
end
apr_needed_hold_e = (nthroot(invest_hold_e/initial_invest,N)-1)*52;
apr_needed_interest_e = (nthroot(invest_interest_e/initial_invest,N)-1)*52;
apr_needed_djia_e = (nthroot(invest_djia_e/initial_invest,N)-1)*52;
apr_needed_pred_e = (nthroot(invest_pred_e/initial_invest,N)-1)*52;
%% f
p=10;r=0;
start_week = max_weeks-N-p+1;
x = djia(start_week:max_weeks-p);
[a, ~,~,~] = covpred(x, p, r);
X_f =[];
initial_invest = 1000;
for w = max_weeks-N-p:max_weeks-p-1
    pred = djia(w+p:-1:w+1);
    pred_temp = - (pred')*(a(2:end));
    X_f = [X_f pred_temp];
end
invest_seq_hold_f = initial_invest;
invest_seq_interest_f = initial_invest;
invest_seq_djia_f = initial_invest;
invest_seq_pred_f = initial_invest;
invest_hold_f = initial_invest;
invest_interest_f = initial_invest;
invest_djia_f = initial_invest;
invest_pred_f = initial_invest;
for i=max_weeks-N+1:max_weeks
    invest_hold_f = invest_hold_f*djia(i)/djia(i-1);
    invest_interest_f = invest_interest_f*(1+interest);
    if(invest_djia_f*djia(i)/djia(i-1)>invest_djia_f*(1+interest))
        invest_djia_f = invest_djia_f*djia(i)/djia(i-1);
    else
        invest_djia_f = invest_djia_f*(1+interest);
    end

    if(invest_pred_f*X_f(i-max_weeks+N)/djia(i-1)>invest_pred_f*(1+interest))
        invest_pred_f = invest_pred_f*djia(i)/djia(i-1);
    else
        invest_pred_f = invest_pred_f*(1+interest);
    end
```

```
    invest_seq_hold_f = [invest_seq_hold_f invest_hold_f];
    invest_seq_interest_f = [invest_seq_interest_f invest_interest_f];
    invest_seq_djia_f = [invest_seq_djia_f invest_djia_f];
    invest_seq_pred_f = [invest_seq_pred_f invest_pred_f];
end
apr_needed_hold_f = (nthroot(invest_hold_f/initial_invest,N)-1)*52;
apr_needed_interest_f = (nthroot(invest_interest_f/initial_invest,N)-1)*52;
apr_needed_djia_f = (nthroot(invest_djia_f/initial_invest,N)-1)*52;
apr_needed_pred_f = (nthroot(invest_pred_f/initial_invest,N)-1)*52;

%% f2
% p=10;r=0;
% start_week = 1;
% x = djia(start_week:N);
% [a, ~,~,~] = covpred(x, p, r);
% X_f =[];
% initial_invest = 1000;
% for w = max_weeks-N-p:max_weeks-p-1
%     pred = djia(w+p:-1:w+1);
%     pred_temp = - (pred')*(a(2:end));
%     X_f = [X_f pred_temp];
% end
% invest_seq_hold_f = initial_invest;
% invest_seq_interest_f = initial_invest;
% invest_seq_djia_f = initial_invest;
% invest_seq_pred_f = initial_invest;
% invest_hold_f = initial_invest;
% invest_interest_f = initial_invest;
% invest_djia_f = initial_invest;
% invest_pred_f = initial_invest;
% for i=max_weeks-N+1:max_weeks
%     invest_hold_f = invest_hold_f*djia(i)/djia(i-1);
%     invest_interest_f = invest_interest_f*(1+interest);
%     if(invest_djia_f*djia(i)/djia(i-1)>invest_djia_f*(1+interest))
%         invest_djia_f = invest_djia_f*djia(i)/djia(i-1);
%     else
%         invest_djia_f = invest_djia_f*(1+interest);
%     end
%
%     if(invest_pred_f*X_f(i-max_weeks+N)/djia(i-
1)>invest_pred_f*(1+interest))
%         invest_pred_f = invest_pred_f*djia(i)/djia(i-1);
%     else
%         invest_pred_f = invest_pred_f*(1+interest);
%     end
%
%     invest_seq_hold_f = [invest_seq_hold_f invest_hold_f];
%     invest_seq_interest_f = [invest_seq_interest_f invest_interest_f];
%     invest_seq_djia_f = [invest_seq_djia_f invest_djia_f];
%     invest_seq_pred_f = [invest_seq_pred_f invest_pred_f];
% end
% apr_needed_hold_f = (nthroot(invest_hold_f/initial_invest,N)-1)*52;
% apr_needed_interest_f = (nthroot(invest_interest_f/initial_invest,N)-1)*52;
% apr_needed_djia_f = (nthroot(invest_djia_f/initial_invest,N)-1)*52;
% apr_needed_pred_f = (nthroot(invest_pred_f/initial_invest,N)-1)*52;
```

```
%% g
p=10;r=0;
start_week = 1;
x = djia(start_week:N);
[a, ~,~,~] = covpred(x, p, r);
X_g =[];
initial_invest = 1000;
for w = 0:max_weeks-p
    pred = djia(w+p:-1:w+1);
    pred_temp = - (pred')*(a(2:end));
    X_g = [X_g pred_temp];
end
X_g = [djia(1:p-1)' X_g];
invest_seq_hold_g = initial_invest;
invest_seq_interest_g = initial_invest;
invest_seq_djia_g = initial_invest;
invest_seq_pred_g = initial_invest;
invest_hold_g = initial_invest;
invest_interest_g = initial_invest;
invest_djia_g = initial_invest;
invest_pred_g = initial_invest;
for i=1:length(djia)-1
    invest_hold_g = invest_hold_g*djia(i+1)/djia(i);
    invest_interest_g = invest_interest_g*(1+interest);
    if(invest_djia_g*djia(i+1)/djia(i)>invest_djia_g*(1+interest))
        invest_djia_g = invest_djia_g*djia(i+1)/djia(i);
    else
        invest_djia_g = invest_djia_g*(1+interest);
    end

    if(invest_pred_g*X_g(i+1)/djia(i)>invest_pred_g*(1+interest))
        invest_pred_g = invest_pred_g*djia(i+1)/djia(i);
    else
        invest_pred_g = invest_pred_g*(1+interest);
    end

    invest_seq_hold_g = [invest_seq_hold_g invest_hold_g];
    invest_seq_interest_g = [invest_seq_interest_g invest_interest_g];
    invest_seq_djia_g = [invest_seq_djia_g invest_djia_g];
    invest_seq_pred_g = [invest_seq_pred_g invest_pred_g];
end
apr_needed_hold_g = (nthroot(invest_hold_g/initial_invest,max_weeks)-1)*52;
apr_needed_interest_g = (nthroot(invest_interest_g/initial_invest,max_weeks)-
1)*52;
apr_needed_djia_g = (nthroot(invest_djia_g/initial_invest,max_weeks)-1)*52;
apr_needed_pred_g = (nthroot(invest_pred_g/initial_invest,max_weeks)-1)*52;

p=10;r=0;
X=[];
k = floor(max_weeks/N);
for i=1:k
    x = djia((i-1)*N+1:(i)*N);
    [a, ~,~,~] = covpred(x, p, r);
    for w = (i-1)*N:(i)*(N)-1
        pred = djia(w+p:-1:w+1);
```

```
        pred_temp = - (pred')*(a(2:end));
        X = [X pred_temp];
    end
end
start_week = length(X)+1;
x = djia(start_week:max_weeks-p);
[a, ~,~,~] = covpred(x, p, r);
for w = length(X):max_weeks-p
    pred = djia(w+p:-1:w+1);
    pred_temp = - (pred')*(a(2:end));
    X = [X pred_temp];
end
X = [djia(1:p-1)' X];
invest_seq_hold_x = initial_invest;
invest_seq_interest_x = initial_invest;
invest_seq_djia_x = initial_invest;
invest_seq_pred_x = initial_invest;
invest_hold_x = initial_invest;
invest_interest_x = initial_invest;
invest_djia_x = initial_invest;
invest_pred_x = initial_invest;
for i=1:length(djia)-1
    invest_hold_x = invest_hold_x*djia(i+1)/djia(i);
    invest_interest_x = invest_interest_x*(1+interest);
    if(invest_djia_x*djia(i+1)/djia(i)>invest_djia_x*(1+interest))
        invest_djia_x = invest_djia_x*djia(i+1)/djia(i);
    else
        invest_djia_x = invest_djia_x*(1+interest);
    end

    if(invest_pred_x*X(i+1)/djia(i)>invest_pred_x*(1+interest))
        invest_pred_x = invest_pred_x*djia(i+1)/djia(i);
    else
        invest_pred_x = invest_pred_x*(1+interest);
    end

    invest_seq_hold_x = [invest_seq_hold_x invest_hold_x];
    invest_seq_interest_x = [invest_seq_interest_x invest_interest_x];
    invest_seq_djia_x = [invest_seq_djia_x invest_djia_x];
    invest_seq_pred_x = [invest_seq_pred_x invest_pred_x];
end
apr_needed_hold_x = (nthroot(invest_hold_x/initial_invest,max_weeks)-1)*52;
apr_needed_interest_x = (nthroot(invest_interest_x/initial_invest,max_weeks)-
1)*52;
apr_needed_djia_x = (nthroot(invest_djia_x/initial_invest,max_weeks)-1)*52;
apr_needed_pred_x = (nthroot(invest_pred_x/initial_invest,max_weeks)-1)*52;

%% h
p =10;r = 0;
[a,~,~,X] = covpred(djia,p,r);
xhat1 = -X*a(2:end);
e = djia(p+1+r:end) - xhat1;
[h,w] = freqz(1,a,max_weeks/2);
h = [-h(end:-1:1);h];
w = [-w(end:-1:1);w];
G = sum(abs(e).^2);
```

```
dtft_result = fft(djia);
X=fftshift(dtft_result);

%% plot without scale
figure
plot(w/pi,20*log10(abs(X)))
hold on
plot(w/pi,20*log10(abs(h)))
xlabel('radians / \pi')
ylabel('magnitude(db)')
legend('DJIA','predictor')
title('DJIA and predictor in the frequency domain')

figure
plot(w/pi,20*log10(abs(X)/max(abs(X))))
hold on
plot(w/pi,20*log10(abs(h)/max(abs(h))))
xlabel('radians / \pi')
ylabel('magnitude(db)')
legend('DJIA','predictor')
title('DJIA and predictor in the frequency domain (normalized)')

%% plot with scale
[h_g,~] = freqz(G,a,max_weeks/2);
h_g = [-h_g(end:-1:1);h_g];
figure
plot(w/pi,20*log10(abs(X)))
hold on
plot(w/pi,20*log10(abs(h_g)))
xlabel('radians / \pi')
ylabel('magnitude(db)')
legend('DJIA','predictor')
title('DJIA and predictor in the frequency domain (scaled)')
```