ReadMe

Replication files for

High-dimensional forecasting with known knowns and known unknowns,

M. Hashem Pesaran. University of Southern California, and Trinity College, Cambridge
Ron P. Smith, Birkbeck, University of London
APRIL 4, 2024

Forthcoming National Institute Economic Review.

The zip file contains the following files.

"TRANSFORMATIONS TO CREATE ACTIVE SET"

This is the set of transformations used to create the active set.

1. **Excel Files**:

1.1 **"UK Inflation Primary data 1979q2 2024q1 - 12Aug23.xls"**

This Excel file contains the primary data used to construct the active set.

1.2 **"UK Inflation Variable Descriptions for Active Set - 12Aug23.xls"**.

This Excel file contains in the first sheet a list of variables in the primary data, their abbreviations, and sources. The second sheet contains a list of the active set variables, abbreviations, descriptions, and the transformation used to calculate them from the primary data.

1.3 **"UK INFLATION PRIMARY DATA 1979Q2 2024Q1 - 12AUG23 ACTIVE SET.xls"**

This Excel file contains the primary data and the data on all the variables in the active set.

1.4 **"REG_PAR_NUM_SEL_LASSO_OCMT_05JAN24.xls"**

It contains the results for Tables 2 and 3.

1.5 **"Selected_Variables_from_Active_Set_q1_Ahead_05JAN24.xls"**

It contains the results for Tables S.7, S.8, and S.9.

1.6 **"Selected_Variables_from_Active_Set_q2_Ahead_05JAN24.xls"**

It contains the results for Tables S.1, S.2, S.10, S.11, and S.12.

1.7 **"Selected_Variables_from_Active_Set_q4_Ahead_05JAN24.xls"**

Tables S.3, S.4, S.5, S.6, S.13, S.14, and S.15.

1.8 **"UK_INF_FORECASTS_05JAN24.xls"**

It contains the results for Tables 5, S.16, S.17, and S.18.

1.9 “**UK_INF_FORECASTS_microfit_05JAN24.xls**”

It contains the results for S.16, S.17, and S.18 but is designed to be used in Mircrofit.

2. **MATLAB Codes**:

2.1 **Data_Importation.m**

This is a main implementation file that generates an active set file from the raw data in (1.3) Excel Files above. This file generates “ActiveSet_Data.mat.” This file has to be run first before obtaining the forecasts from (2.2), (2.3), and (2.4) below.

2.2 **UK_INF_1q_05JAN24.m**

This main implementation file generates Excel files containing all 1-quarter-ahead (forecast) results and a corresponding figure.

2.3 **UK_INF_2q_05JAN24.m**

This main implementation file generates Excel files containing all 2-quarter-ahead (forecast) results and a corresponding figure.

2.4 **UK_INF_4q_05JAN24.m**

This main implementation file generates Excel files containing all 4-quarter-ahead (forecast) results and a corresponding figure.

2.5 **ACTIVE_SEL.m**

This function file categorizes the specified active set into several subsets based on the model (e.g., Lasso, AR2-Lasso, ARX-Lasso, AR2-OCMT, ARX-OCMT): conditional sets, lagged and non-lagged active sets, and the variables' names within the active set.

2.6 **EST_AR2.m**

This function, part of the main implementation files in (2.2), (2.3), and (2.4), estimates the AR2 model.

2.7 **EST_AR2_LASSO_AVE.m**

This function, part of the main implementation files in (2.2), (2.3), and (2.4), estimates the Lasso-AR2 model.

2.8 **EST_AR2_LASSO_POOL.m**

This function, part of the main implementation files in (2.2), (2.3), and (2.4), estimates the Lasso-AR2-pooled-MSE model.

2.9 **EST_AR2_OCMT.m**

This function, part of the main implementation files in (2.2), (2.3), and (2.4), estimates the OCMT-AR2 model.

2.10 **EST_ARX.m**

This function, part of the main implementation files in (2.2), (2.3), and (2.4), estimates the ARX model.

2.11 **EST_ARX_LASSO_AVE.m**

This function, part of the main implementation files in (2.2), (2.3), and (2.4), estimates the Lasso-ARX model.

2.12 **EST_ARX_LASSO_POOL.m**

This function, part of the main implementation files in (2.2), (2.3), and (2.4), estimates the Lasso-ARX-pooled-MSE model.

2.13 **EST_ARX_OCMT.m**

This function, part of the main implementation files in (2.2), (2.3), and (2.4), estimates the OCMT-ARX model.

2.14 **EST_LASSO_AVE.m**

This function, part of the main implementation files in (2.2), (2.3), and (2.4), estimates the Lasso model.

2.15 **EST_LASSO_POOL.m**

This function, part of the main implementation files in (2.2), (2.3), and (2.4), estimates the Lasso-pooled-MSE model.

2.16 **GENVARS.m**

This function, part of the main implementation files in (2.2), (2.3), and (2.4), generates the set of lagged variables.

2.17 **Lasso_fsel.m**

This function selects variables using Lasso regression by de-meaning and standardizing the predictor variables X, performing Lasso path computation, and then choosing the optimal regularization parameter lambda and the associated non-zero coefficients to identify the selected variables. The MATLAB code originates from Chudik, Kapetanios, and Pesaran (2018).

2.18 **ocmt_sel.m**

This function executes the initial stage of the OCMT procedure using a predefined delta value. The MATLAB code originates from Chudik, Kapetanios, and Pesaran (2018).

2.19 **ol1.m**

This function performs coordinate descent optimization for Lasso regression, updating coefficients beta iteratively based on the precomputed matrices, an initial value, a regularization parameter lambda, and an optional variable set until convergence within a specified tolerance or a maximum

number of iterations is reached. The MATLAB code originates from Chudik, Kapetanios, and Pesaran (2018).

2.20 **pathl1.m**

This function computes a path of Lasso solutions for a given range of regularization parameters lambda, scaling the predictor variables X for uniform L2 norms, and iteratively optimizes to find sparse coefficient vectors beta for each lambda, returning these vectors along with the corresponding lambda values and optional scaling factors. The MATLAB code originates from Chudik, Kapetanios, and Pesaran (2018).

2.21 **pathl1_l.m**

This function, similar to 'pathl1.m', calculates a path of Lasso solutions across a predefined range of regularization parameters lambda and scales predictor variables X for uniform L2 norms. It iteratively finds sparse coefficient vectors beta for each lambda, unlike 'pathl1.m', which internally generates a lambda sequence based on data characteristics. This function also returns these vectors, the specified lambda values, and optional scaling factors. The MATLAB code originates from Chudik, Kapetanios, and Pesaran (2018).

2.22 **pLasso_fsel.m**

This function follows the same procedure as “Lasso_fsel.m,” but follows the pooled-MSE procedure to choose lambda.

2.23 **select_lambda_lasso.m**

This function selects the optimal regularization parameter lambda for Lasso regression by conducting 10-fold cross-validation to compute mean squared errors for a range of lambda values, averaging the lambdas that minimize the error in each fold, and then choosing the lambda closest to this average. The MATLAB code originates from Chudik, Kapetanios, and Pesaran (2018).

2.24 **select_lambda_plasso.m**

This function determines the optimal regularization parameter lambda for penalized Lasso regression by performing 10-fold cross-validation to calculate the mean squared errors for different lambda values across folds, averaging these errors, and selecting the lambda associated with the lowest average error.

2.25 **ul1.m**

This function performs the soft thresholding operation, which is a key component of the Lasso regression algorithm. The MATLAB code originates from Chudik, Kapetanios, and Pesaran (2018).

2.26 **ActiveSet_Data.mat**

The MATLAB Data file generated by (2.1) above.

3. **PNG file:**

3.1 **UK_INF_1q_05JAN24.png**

: **Figure S.1**, which is generated by (2.2).

3.2 **UK_INF_2q_05JAN24.png**

: **Figure S.2**, which is generated by (2.3).

3.3 **UK_INF_4q_05JAN24.png**

: **Figure S.3**, which is generated by (2.4).

References

Chudik, Kapetanios, and Pesaran (2018) “A One Covariate at a Time, Multiple Testing Approach to Variable Selection in High-Dimensional Linear Regression Models,” *Econometrica*, Vol. 86, No. 4, 1479-1512.