Sample B2

1. The heteroskedasticity robust error used in Regression (1) assumes serially uncorrelated error and being independently and identically distributed, hence the off-diagonal covariance terms of expansion of the variance of a weighted sum vanish.

$$\text{Cov}(v_t, v_s) = 0$$

However, this assumption clearly fails in regression (1) because it resembles to an autoregressive model where the outcome variable would depend on its past values, so error terms exhibit serial correlation in addition to potential heteroskedasticity. Therefore, HAC standard error is preferred as they remain consistent in the presence of both heteroskedasticity and autocorrelation.

2. For OOS forecasting purpose, we would choose the specification with the smallest RMSFE, which corresponds to Regression (5). The standard error of the forecast will also be the RMSFE, which is 0.381.

3. We have

$$\Delta 4\_LFPR55_{4Q15} = 0.497 + 0.138\Delta LFPR55_{4Q14} + 0.38\Delta LFPR55_{3Q14} + 0.43\Delta LFPR55_{2Q14}$$

$$\Delta 4\_LFPR55_{4Q15} = 0.50866$$

By the definition of the outcome variable

$$\widehat{LFPR}55_{4Q15} = LFPR55_{4Q14} + 0.50866 = 40.50866$$

The 95% forecast interval is $\widehat{LFPR}55_{4Q15} \pm 1.96 \times RMSFE$, which is evaluated to

$$[39.41, 41.60]$$

4. We perform a Chow test for a known break date. Let $y_t = \Delta 4\_LFPR55_t$ and $x_t = \Delta LFPR55_t$, we establish a regression that includes a binary break indicator defined as $D_i = \mathbf{1}\{\boldsymbol{t > 2008q1}\}$ plus its interaction terms with all regressors:

$$y_t = \beta_0 + \beta_1\Delta x_{t-4} + \beta_2\Delta x_{t-5} + \beta_3\Delta x_{t-6} + \gamma_0 D_i + \gamma_1(D_i \times \Delta x_{t-4}) + \gamma_2(D_i \times \Delta x_{t-5}) + \gamma_3(D_i \times \Delta x_{t-6}) + u_t$$

We test for the null hypothesis that 2008q1 is not a break date:

$$\mathcal{H}_0: \gamma_0 = \gamma_1 = \gamma_2 = \gamma_3 = 0$$

$$\mathcal{H}_1: \text{at least one of } \gamma_1, \dots, \gamma_4 \text{ is not zero}$$

We use the standard F-statistics with 4 degrees of freedom to test the null hypothesis.

5. We do not use Regression (1) as its standard error is inconsistent under autocorrelation. We also ignore regression (4) because all change in real disposable income terms are individually and jointly insignificant and therefore do not provide any useful information. Regression (2) can also be disregarded because no macroeconomic covariates are included in the model. Hence we focus on only Regression (3) and (5).

   For both (3) and (5), we observe using Table 2 that their macroeconomic covariates are both jointly significant under 1% level of significance, despite Regression (3) has smaller BIC and higher adjusted R-squared than (5), suggesting stronger in-sample model fit. Using Figure 3, the significant deviation OOS prediction of regression (3) during the eruption phase of the GFC suggests that it is more prone to unexpected exogenous shocks. However, regression (5) delivers stronger fit for OOS prediction despite displaying similar deviation during the onset of the GFC due to model set-up. However, as macro aggregates recovered—unemployment receded toward pre-recession levels, personal income rebounded, and housing starts resumed growth—the LFPR55 failed to recapture its prior momentum, suggesting that the recession and the slow recovery contributed to the plateau of late retirement, at least partially.

   Nevertheless, the weak in-sample explanatory poewr of Regression (5) and sizeable RMSFE for both regressions indicates that cyclical variables alone cannot account for the persistent plateau. Thus, although the recession precipitated an initial drop in older-worker participation, post-2010 stagnation appears driven by additional factors—demographic shifts, regulatory changes or alterations in retirement-saving behavior—beyond mere cyclical weakness.

6. We may use LASSO, Ridge regression and principal component regression:

- **LASSO (**L1 Regularization**)** augments OLS criterion by including a L1 penalty term, it minimize the penalized SSR

$$\hat{b} = \arg\min_{b \in R^k} \left\{ \sum_{i=1}^{N} \left(y_i - \sum_{j=1}^{k} b_j x_{ij} \right)^2 + \lambda_{\text{LASSO}} \sum_{j=1}^{k} \left| b_j \right| \right\}$$

  Where λ_{LASSO} governs the degree of shrinkage. LASSO yields a parsimonious model that retains only predictors with the strongest marginal contributions and eliminates some useless predictors, thereby ameliorating overfitting in high-dimensional settings and enhancing interpretability.

- **Ridge regression** is also a shrinkage model but with a different penalty term in LASSO, the principal behind LASSO and Ridge is similar. Ridge regression selects b by minimizing

$$S^{\text{Ridge}}(b; \lambda_{\text{Ridge}}) = \sum_{i=1}^{N} \underbrace{(y_i - b_1 x_{1i} - \cdots - b_k x_{ki})^2}_{\text{OLS Object Function}} + \underbrace{\lambda_{\text{Ridge}} \sum_{j=1}^{k} b_j^2}_{\text{Penalty}}$$

  The quadratic penalty uniformly shrinks all coefficients toward zero without eliminating any regressor. Ridge trades off a small increase in bias for potentially large improvement in efficiency.

- **Principal component regression** constructs orthogonal linear combinations of the original predictors that successively capture the greatest variance in the data.