Forecast disagreement about long-run macroeconomic relationships

Article

Supplemental Material

Online Appendix for “Forecast Disagreement about Long-run Macroeconomic Relationships”

Pei Kuang, Li Tang, Renbin Zhang, and Tongbin Zhang

June 12, 2022

A Stochastic growth models

This section reviews some one- and two-sector stochastic growth models and shows Proposition 1 holds for these models. We use a version of the two-sector model of Schmitt-Grohe and Uribe (2011) which nests many one- and two-sector stochastic growth models. For our purpose of studying cointegration relation, it is innocuous to leave out some non-essential features, such as capital adjustment cost, capacity utilization, taxes, government spending and transitory shocks. Incorporating these features does not affect the cointegration relation of interest and hence all of our theoretical and empirical results.

The economy is populated by a unit mass of identical infinite-horizon agents with preferences as

$$U = \sum_{t=0}^{\infty} \beta^t u(C_t, 1 - N_t),$$

where C_t is consumption of commodity goods, N_t is labour input, β is subjective discount factor and u is the utility function. The production function of final good is

$$Y_t = K_t^{1-\alpha} \left(X_t^z N_t \right)^{\alpha},$$

where K_t is the pre-determined capital stock and N_t is the labor input. X_t^z is a permanent neutral productivity shock. The capital stock evolves according to

$$K_{t+1} = (1-\delta)K_t + X_t^a H(I_t),$$

where δ is depreciation rate and I_t is investment. X_t^a is nonstationary investment-specific technology shocks. $H(I) = I^{\xi}$ is the production function. In a decentralized version of this economy, the relative price of investment goods in terms of consumption goods, which we denote by p_t^I, is given by

$$p_t^I = \frac{1}{X_t^a H'(I_t)}.$$

The resource constraint is $Y_t = C_t + I_t$.

In this model economy, TFP and the price of investment are given, respectively, by consumption

$$TFP_t = (X_t^z)^{1-\alpha}$$

and

$$p_t^I = \frac{1}{X_t^a \xi I_t^{\xi-1}}$$

Let $\mu_t^z \equiv X_t^z / X_{t-1}^z$ and $\mu_t^a \equiv X_t^a / X_{t-1}^a$ denote, respectively, the gross growth rates of X_t^z and X_t^a. And let $x_t = \psi \ln(X_t^z) - \ln(X_t^a)$. The joint law of motion of X_t^z and X_t^a follows the vector error correction model (VECM)

$$\begin{bmatrix} \ln(\mu_t^z/\mu^z) \\ \ln(\mu_t^a/\mu^a) \end{bmatrix} = \begin{bmatrix} \rho_{11} & \rho_{12} \\ \rho_{21} & \rho_{22} \end{bmatrix} \begin{bmatrix} \ln\left(\mu_{t-1}^z/\mu^z\right) \\ \ln\left(\mu_{t-1}^a/\mu^a\right) \end{bmatrix} + \begin{bmatrix} \kappa_1 \\ \kappa_2 \end{bmatrix} x_{t-1} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \epsilon_t^1 \\ \epsilon_t^2 \end{bmatrix} \tag{A.1}$$

where the innovations to the common trend in neutral and investment-specific productivity, ϵ_t^1 and ϵ_t^2 are *i.i.d* normal with mean zero and variances $\sigma_{\epsilon^1}^2$ and $\sigma_{\epsilon^2}^2$, respectively.

We consider three cases. **Case I:** neutral productivity shock X_t^z and investment-specific productivity shock X_t^a share a common stochastic trend; it is supported by the empirical evidence in Schmitt-Grohe and Uribe (2011).[1] Thus,

$$x_t = \psi \ln \left(X_t^z \right) - \ln \left(X_t^a \right) \text{ is stationary.}$$

Case II: assuming that TFP and the price of investment possess independent stochastic trends, i.e., $\rho_{21} = \rho_{21} = \kappa_1 = \kappa_2 = D_{21} = 0$, see e.g., Fisher (2006).[2] **Case III:** we shut down the investment specific shocks by setting $X_t^a = 1$ for all t. Moreover, let $\rho_{11} = \rho_{12} = 0$, and $\kappa_1 = \psi = 0$. The productivity process becomes $\ln \left(\mu_t^z / \mu^z \right) = D_{11} \epsilon_t^1$. This becomes a version of the one-sector stochastic growth, e.g., like King, Plosser and Rebelo (1988).

The balanced growth path. For all three cases, there exists a balanced growth path along which the following variables are stationary:

$$\frac{Y_t}{X_t^Y}, \frac{C_t}{X_t^Y}, \frac{I_t}{X_t^Y}, \frac{Y_t/N_t}{X_t^Y},$$

where $X_t^Y = \left(X_t^z \right)^{\frac{1-\alpha}{1-\alpha\xi}} \left(X_t^a \right)^{\frac{\alpha}{1-\alpha\xi}}$. Hence Y_t, C_t and I_t have a common trend and are cointegrated with each other with cointegrating vector $(1, -1)$. Thus, Proposition 1 holds for all three models.

B Rebasing forecasts data

Since the Survey of Professional Forecasters (SPF) began, there have been a number of changes of the base year in the national income and product accounts (NIPA). The forecasts for levels of consumption, investment and output use the base year that was in effect when the forecasters received the survey questionnaire. This Appendix explains how the forecasts data are rebased.

Table A.1 provides the base year in effect for NIPA variables (including consumption expenditures), reproduced from Table 4 of the documentation of Survey of Professional Forecasters (p. 23). For rebasing, we use real consumption, investment and output data of different vin-

[1] They estimate a two-sector stochastic growth model which contains this feature and find that innovations in the common stochastic trend explain a sizable fraction of the unconditional variances of output, consumption, investment and hours.

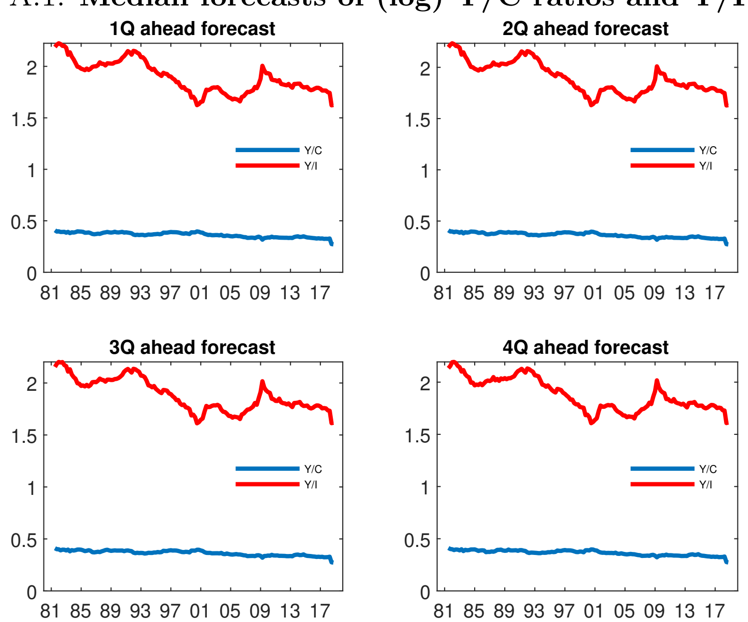
[2] Nevertheless, Schmitt-Grohe and Uribe(2011) argue that this formulation is strongly rejected by the data (see their Section 2).

Table A.1: **Base years and ratios for rebasing**

Range of Survey Dates	Base Year	Ratio
1976:Q1 to 1985:Q4	1972	3.31
1986:Q1 to 1991:Q4	1982	1.48
1992:Q1 to 1995:Q4	1987	1.23
1996:Q1 to 1999:Q3	1992	1.04
1999:Q4 to 2003:Q4	1996	1
2004:Q1 to 2009:Q2	2000	0.94
2009:Q3 to 2013:Q2	2005	0.84
2013:Q3 to present	2009	0.79

tages from the Real-Time Data Set for Macroeconomists managed by the Federal Reserve Bank of Philadelphia. Year 1996 is used as the common base year for all forecast data. The data in each window needs to be rebased by multiplying a base ratio. For instance the 2000:Q1 real consumption at the window from 1996:Q1 to 1999:Q3 is 1409.5 while it is 1469.5 at 2000:Q1 and hence the ratio is 1469.5/1409.5. Figure A.1 plots the (normalized) rebased median forecasts of (log) output-consumption ratios and (log) output-investment ratios for all four forecasting horizons, respectively.

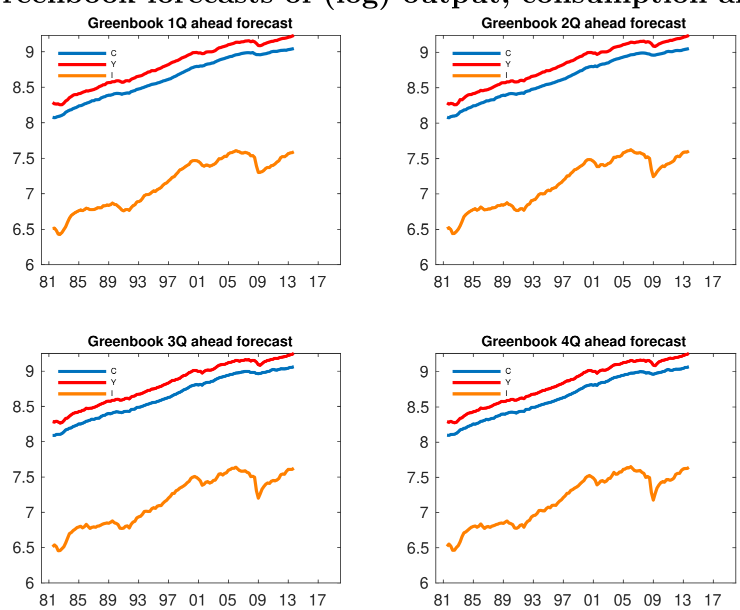
Figure A.1: **Median forecasts of (log) Y/C ratios and Y/I ratios**



C Testing using the Greenbook forecasts

This Appendix shows the result of no cointegration between forecasts of output and consumption (or investment) still holds when we use the Greenbook forecast dataset in lieu of SPF data. The Greenbook contains projections on the US economy forwards (and backwards) and is produced before each meeting of the Federal Open Market Committee. It includes projections for a large number of macroeconomic variables including real consumption growth, real GDP growth and real investment. Four forecasting horizons are reported in each projection: 1- to 4-quarter ahead (while more horizons are issued from time to time). The dataset is published with a five-year lag. The sample of Greenbook growth forecast we use spans from 1981:Q3 to 2013:Q4.

Figure A.2: **Greenbook forecasts of (log) output, consumption and investment**



Real consumption level forecast is obtained by multiplying the consumption growth forecast (gRPCE) by (rebased) real-time estimate of consumption level. Real investment level forecast is obtained by summing the forecast of the level of real residential investment and the level of real non-residential investment. The forecast of the level of real residential investment is calculated as the real residential investment growth forecast (gRRES) multiplied by (rebased) real residential investment level. The forecast of the level of real non-residential investment is calculated as the real non-residential investment growth forecast (gRBF) multiplied by (rebased) real non-residential investment level. Real total government spending forecast is subtracted from real GDP level forecast. All level data comes from real-time datasets for the US economy maintained by the Philadelphia Fed.

Figure A.3: **Greenbook forecasts of (log) Y/C ratios and Y/I ratios**

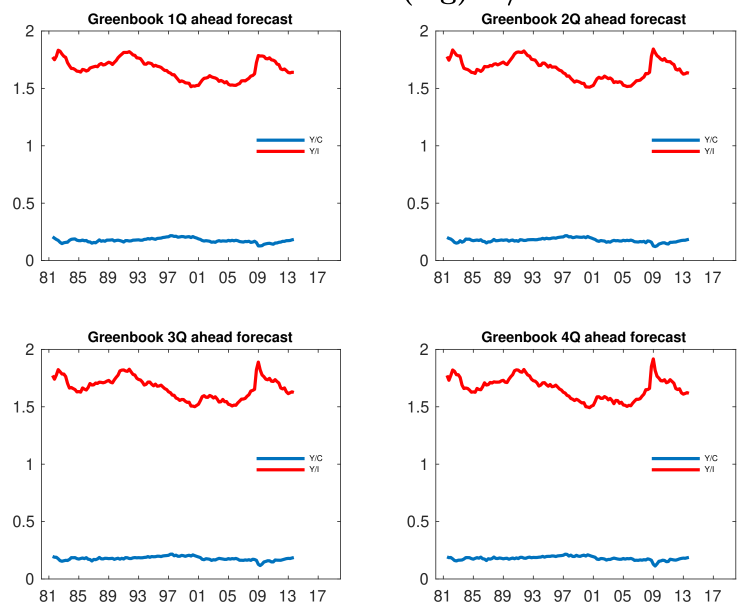


Figure A.2 plots the normalized and rebased Greenbook forecasts of log output, consumption and investment. Table A.2 reports the integration properties of Greenbook forecasts. Similar to SPF median (or mean) forecast testing results, Greenbook forecasts of consumption, output and investment are integrated of order 1, i.e. I(1), but not integrated of order 2, i.e. I(2).

Table A.2: **Integration properties of Greenbook forecasts**

P values				
	1Q ahead	2Q ahead	3Q ahead	4Q ahead
Panel A: I(1) test				
Consumption forecasts				
PP (Z_t test)	0.9903	0.9883	0.9864	0.9862
Dickey-Fuller	0.9966	0.9950	0.9930	0.9923
Output forecasts				
PP (Z_t test)	0.9287	0.9250	0.9189	0.9169
Dickey-Fuller	0.9830	0.9688	0.9604	0.9553
Investment forecasts				
PP (Z_t test)	0.7887	0.7423	0.7036	0.6845
Dickey-Fuller	0.9501	0.9107	0.8685	0.8345
Panel B: I(2) test				
Consumption forecasts				
PP (Z_t test)	0.000	0.000	0.000	0.000
Dickey-Fuller	0.000	0.000	0.000	0.000
Output forecasts				
PP (Z_t test)	0.000	0.000	0.000	0.000
Dickey-Fuller	0.000	0.000	0.000	0.000
Investment forecasts				
PP (Z_t test)	0.000	0.000	0.000	0.000
Dickey-Fuller	0.000	0.000	0.000	0.000

Figure A.3 plots Greenbook forecasts of (log) output-to-consumption and output-to-investment ratios. Table A.3 reports cointegration test results between forecasts of output and consumption (or investment) when the theoretical $(1, -1)$ cointegration relation is imposed.[3] Both PP and DF-GLS tests suggest that the forecast of output is not cointegrated with consumption (or investment) at standard critical level, when the theoretical $(1, -1)$ cointegration relation is imposed. Therefore, this result is consistent with SPF forecast testing results.

Table A.4 reports the Engle-Granger cointegration test outcomes when no cointegration restriction is imposed.[4] Again, the Engle-Granger test indicates that the forecasts of output

[3]We report the cointegration test results when trend is omitted. If trend is introduced, the DF-GLS test and the PP test indicate that the forecast of output-consumption ratio (or output-investment ratio) are not cointegrated at 10% critical level.

[4]We report Engle-Granger test results without incorporating the trend component. Our test results are robust when the trend is included.

Table A.3: **cointegration test for Greenbook forecasts with cointegrating vector** $(1, -1)$

Panel A: *cointegration between forecasts of consumption and output*					
Mean	1Q ahead	2Q ahead	3Q ahead	4Q ahead	1Q C & 4Q Y
PP (Z_t test)	-2.682	-2.683	-2.672	-2.738	-2.732
5% critical value	-2.888	-2.888	-2.888	-2.888	-2.888
DF-GLS	-1.593	-1.580	-1.834	-1.822	-1.888
5% critical value	-2.077	-2.062	-2.062	-2.053	-2.062
KPSS	1.19	1.30	1.35	1.40	1.03
5% critical value	0.463	0.463	0.463	0.463	0.463
Panel B: *cointegration between forecasts of investment and output*					
Mean	1Q ahead	2Q ahead	3Q ahead	4Q ahead	1Q I & 4Q Y
PP (Z_t test)	-1.856	-1.976	-2.047	-2.103	-1.871
5% critical value	-2.888	-2.888	-2.888	-2.888	-2.888
DF-GLS	-1.286	-1.492	-1.529	-1.546	-2.030
5% critical value	-2.077	-2.062	-2.062	-2.062	-2.062
KPSS	1.86	1.74	1.14	1.15	1.78
5% critical value	0.463	0.463	0.463	0.463	0.463

are not cointegrated with the forecasts of consumption (or investment) at 10% significance level, consistent with the testing results from SPF forecasts.

Table A.4: **cointegration test for Greenbook forecasts without imposing cointegrating vector** $(1, -1)$

Engle-Granger test					
Panel A: *cointegration between forecasts of output and consumption*					
Median	1Q ahead	2Q ahead	3Q ahead	4Q ahead	4Q Y &1Q C
Test stats.	-2.313	-2.575	-2.610	-2.821	-2.805
10% critical value	-3.077	-3.077	-3.077	-3.077	-3.077
Panel B: *cointegration between forecasts of output and investment*					
Median	1Q ahead	2Q ahead	3Q ahead	4Q ahead	4Q Y &1Q I
Test stats.	-2.805	-1.138	-1.446	-1.627	-1.191
10% critical value	-3.077	-3.077	-3.077	-3.077	-3.077

Conclusion (Panel A): Greenbook consumption forecasts are not cointegrated with output forecasts without imposing any cointegrating vector;
Conclusion (Panel B): Greenbook investment forecasts are not cointegrated with output forecasts without imposing any cointegrating vector.

D Integration properties of mean forecasts

This appendix reports integration properties of SPF 1- to 4-quarters ahead mean forecasts of consumption, output and investment. Panel A indicates that all forecasts over all forecasting horizons are integrated of order 1, i.e. I(1) and Panel B shows that all forecasts are not integrated of order 2, i.e. I(2) at conventional significance level. Therefore, test results for mean forecasts are consistent with median forecast results.

Table A.5: **Integration properties of mean SPF forecasts**

P values				
	1Q ahead	2Q ahead	3Q ahead	4Q ahead
Panel A: I(1) test				
Mean consumption forecasts				
PP (Z_t test)	0.9082	0.9049	0.9023	0.9023
Dickey-Fuller	0.9510	0.9478	0.9459	0.9451
Mean output forecasts				
PP (Z_t test)	0.7767	0.7796	0.7788	0.7806
Dickey-Fuller	0.8963	0.8930	0.8902	0.8884
Mean investment forecasts				
PP (Z_t test)	0.7216	0.7100	0.7097	0.7116
Dickey-Fuller	0.8916	0.8858	0.8849	0.8851
Panel B: I(2) test				
Mean consumption forecasts				
PP (Z_t test)	0.000	0.000	0.000	0.000
Dickey-Fuller	0.000	0.000	0.000	0.000
Mean output forecasts				
PP (Z_t test)	0.000	0.000	0.000	0.000
Dickey-Fuller	0.000	0.000	0.000	0.000
Mean investment forecasts				
PP (Z_t test)	0.000	0.000	0.000	0.000
Dickey-Fuller	0.000	0.000	0.000	0.000

Evidence: Mean 1-, 2-, 3- and 4-quarter ahead forecasts of aggregate consumption, output and investment are I(1) but not I(2).

E Testing using mean SPF forecasts

E.1 Testing with imposing the theoretical restriction

This section shows no cointegration between mean forecasts of output and consumption (or investment) with imposing the theory-implied cointegration vector $(1, -1)$, consistent with the testing results using median forecasts. Panel A (or B) of Table A.6 reports the testing results on cointegration between output forecasts and consumption (or investment) forecasts.

PP and DF-GLS tests fail to reject no cointegration between mean forecasts of output and consumption at 10% level with two exceptions marked by dagger (-1.825 and -1.851). The two exceptions come from DF-GLS tests between forecasts of output and consumption over forecasting horizons of 3- and 4-quarter ahead, which only marginally reject the null of no cointegration at 10% critical values. For both cases, the null hypothesis are nevertheless rejected at 5% level. The KPSS tests strongly in favor of no cointegration over all forecasting horizons at 5% level.

In Panel B, cointegration test results from PP and DF-GLS tests between mean forecasts of output and investment are almost identical to median forecast test results, in favor of no cointegration. Consistently, KPSS tests in the two panels indicate a strong rejection of its null of cointegration between mean forecasts, agreeing with other tests performed.

E.2 Testing without imposing cointegration restrictions - mean forecasts

Using the Engle-Granger test, Panel A (or B) of Table A.7 tests if mean forecasts of output are cointegrated with the mean forecasts of consumption (or investment) without imposing a cointegration vector $(1, -1)$. The tests cannot reject the null hypothesis that output forecasts are not cointegrated with consumption or investment forecasts, respectively, at 10% level over any forecasting horizon. The cointegration test results suggest that there exists no cointegrating vector, with which mean forecasts of output are cointegrated with forecasts of consumption (or

Table A.6: **cointegration test for mean SPF forecasts with cointegrating vector** $(1, -1)$

Panel A: *cointegration between forecasts of consumption and output*					
Mean	1Q ahead	2Q ahead	3Q ahead	4Q ahead	1Q C & 4Q Y
PP (Z_t test)	-1.578	-1.596	-1.628	-1.657	-1.717
10% critical value	-2.577	-2.577	-2.577	-2.577	-2.577
DF-GLS	-1.347	-1.318	-1.825†	-1.851†	-1.551
10% critical value	-1.749	-1.749	-1.737	-1.737	-1.749
KPSS	1.948	1.954	1.983	2.011	1.831
5% critical value	0.463	0.463	0.463	0.463	0.463
Panel B: *cointegration between forecasts of investment and output*					
Mean	1Q ahead	2Q ahead	3Q ahead	4Q ahead	1Q I & 4Q Y
PP (Z_t test)	-1.656	-1.616	-1.601	-1.542	-1.685
10% critical value	-2.577	-2.577	-2.577	-2.577	-2.577
DF-GLS	-0.249	-0.324	-0.319	-0.351	-0.214
10% critical value	-1.737	-1.737	-1.737	-1.737	-1.737
KPSS	2.641	2.696	2.781	2.872	2.638
5% critical value	0.463	0.463	0.463	0.463	0.463

†: Test statistics with dagger indicate that corresponding tests reject the null of unit root (no cointegration) at 10% critical values, but fail to reject the null at 5%. The 5% critical value for both tests is -2.047.

forecasts of investment) over any forecasting horizon.

E.3 Multivariate testing using mean forecasts

Using mean SPF forecasts data, Table A.8 tests if forecasts of output, consumption and investment share a common trend. Only the trace test for 1-quarter ahead forecasts rejects the null of zero cointegrating vector, in favor of the existence of cointegrating vector, but fails to reject the null of 1 cointegrating vector against the alternative of more than one cointegrating vector. However, the maximum-eigenvalue test fails to reject the null of zero cointegrating vector against the alternative of one cointegrating vector for 1-quarter ahead forecasts. The rest of the test statistics suggest that mean forecasts of output, consumption, and investment do not share a common trend, similar to median forecasts.

Table A.7: **cointegration test for mean SPF forecasts without imposing cointegrating vector** $(1, -1)$

Engle-Granger test					
Panel A: *cointegration between forecasts of consumption and output*					
Mean	1Q ahead	2Q ahead	3Q ahead	4Q ahead	1Q C & 4QY
Test stats.	-2.357	-2.373	-2.462	-2.538	-2.525
10% critical value	-3.073	-3.073	-3.073	-3.073	-3.073
Panel B: *cointegration between forecasts of investment and output*					
Mean	1Q ahead	2Q ahead	3Q ahead	4Q ahead	1Q I & 4QY
Test stats.	-2.223	-2.245	-2.272	-2.277	-2.264
10% critical value	-3.073	-3.073	-3.073	-3.073	-3.073

F Dealing with missing values

There is a small number of missing values in individual-level forecasts during 1981q3 to 2018q4. Therefore, before conducting formal unit-root/cointegration tests, we fix the missing data problem by filling in gaps. Ryan and Giles (1998) examine three natural ways of dealing with missing observations in the process of unit root testing: "ignoring" the gaps, replacing the missing observation(s) with the previously recorded observation (previous observation carried forward, POCF) and using step interpolation, i.e. linearly interpolating between the last recorded observation and the next recorded observation after to fill in the gap. They conclude that in terms of the power of the test, in addition to size distortion, ignoring the gaps is the best method among these three methods. Later works, like Ghysels and Miller (2014), examine the cointegration test results and suggest that linear interpolation of missing observation should be avoided in cointegration tests. Therefore, in line with the previous work, this paper applies the methods of "ignoring" the gap to fill in small gaps in observations and reports the relevant test results in the main text. Below, we check the robustness of test results with the methods of POCF to fill in small gaps.

We show the testing results in Section 4 are robust to an alternative method of dealing with missing values. We fill in missing gaps for individual forecasters using the method of Previous-Obervation-Carried-Forwards (POCF) and re-perform the individual-level tests. Note that

Table A.8: **Johansen trace and maximum-eigenvalue tests for the number of common trend among mean forecasts**

Johansen test				
Trace test: $J^{trace}(r)$, r = rank				
null:	no more than r cointegration vector			
Alternative:	number of the cointegrating vector > r			
Mean	r=0	5% critical	r=1	5% critical
1Q ahead	30.4	29.7	9.7*	15.4
2Q ahead	29.5*	29.7	9.7	15.4
3Q ahead	27.7*	29.7	9.1	15.4
4Q ahead	27.0*	29.7	8.4	15.4
Maximum-eigenvalue test: $max(r)$, r = rank				
null:	no more than r cointegration vector			
Alternative:	number of the cointegrating vector = r+1			
Mean	r=0	5% critical	r=1	5% critical
1Q ahead	20.7*	21.0	9.6	14.1
2Q ahead	19.8*	21.0	9.5	14.1
3Q ahead	18.5*	21.0	9.0	14.1
4Q ahead	18.7*	21.0	8.2	14.1

*: $J^{trace}(r)$ or $max(r)$ t test statistics with asterisk indicate that corresponding rank r is the lowest rank, for which trace test fails to reject its null number of cointegration equation, and is accepted as the estimated number of cointegrating vector among these three forecast variables.

when POCF is applied, we only fill in gaps that fall in the middle of forecasting periods. For example, since forecaster ID 431 starts participating in the SPF survey from 1991q1 and ends at 2013q3, only missing observations between this time interval is filled. The testing results in Table A.9, A.10, and A.11 are similar to those in Table 8, 9, and 10, respectively.

Table A.9: **Tests with** $(1, -1)$ **restriction using individual-level forecasts over the same forecasting horizon**

Total individual forecasters:	**21**, with **4** forecasts each (1Q, 2Q, 3Q & 4Q ahead)	
with $(1, -1)$ restriction	**No. of no cointegration detected out of 84**	**Proportion of no cointegration detected**
Panel A: *cointegration between forecasts of consumption and output (same horizon)*		
PP Z_t test (10% crit. value)	59	70.3%
DF-GLS (10% crit. value)	72	85.7%
KPSS (5% crit. value)	57	67.9%
Panel B: *cointegration between forecasts of investment and output (same horizon)*		
PP Z_t test (10% crit. value)	78	92.9%
DF-GLS (10% crit. value)	78	92.9%
KPSS (5% crit. value)	67	79.8%

Table A.10: **Tests using individual-level forecasts with** $(1, -1)$ **restriction over different forecasting horizons: 1Q ahead consumption (or investment) forecasts and 4Q ahead output forecasts**

with $(1, -1)$ restriction	**No. of no cointegration detected out of 21**	**Proportion of no cointegration detected**
Panel A: *cointegration between 1Q ahead consumption and 4Q ahead output*		
PP Z_t test (10% crit. value)	8	38.1%
DF-GLS (10% crit. value)	16	76.2%
KPSS (5% crit. value)	14	66.7%
Panel B: *cointegration between 1Q ahead investment and 4Q ahead output*		
PP Z_t test (10% crit. value)	20	95.2%
DF-GLS (10% crit. value)	20	95.2%
KPSS (5% crit. value)	17	81.0%

Table A.11: **Tests using individual-level forecasts without** $(1, -1)$ **restriction over same forecasting horizons**

Total individual forecasters:	**21**, with **4** forecasts each (1-, 2-, 3- & 4Q ahead)	
Engle-Granger test (10% crit. value)		
over same horizons	forecasts of Y and C	forecasts of Y and I
No. of no cointegration detected out of 84	65	82
Proportion of no cointegration detected	77.4%	97.6%

G Some testing results using individual-level forecasts

G.1 Integration properties

Table A.12 reports unit root testing results for forecasts of aggregate output, consumption and investment made by individual professional forecasters. Both ADF test and KPSS test uniformly indicate that the individual-level forecasts over all horizons are I(1) at 5% significance level.

Table A.12: **Unit root test results for individual forecasts**

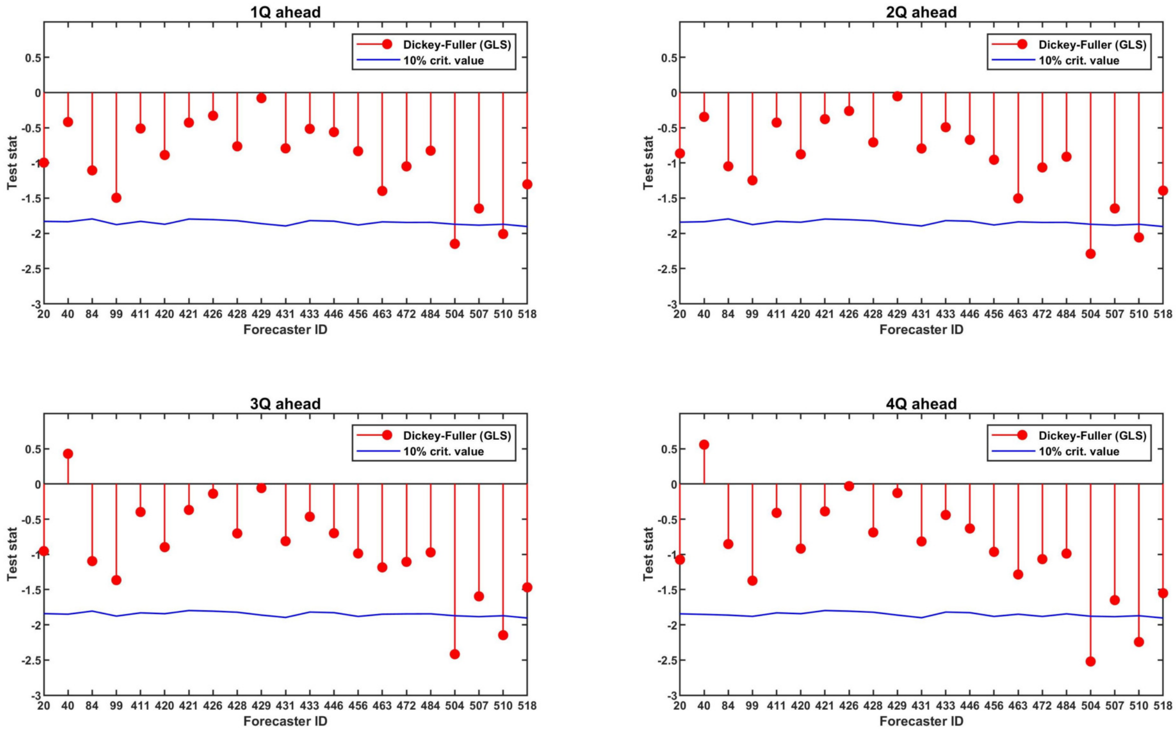
Total individual forecasters:	**21**, with **4** forecasts each (1Q, 2Q, 3Q & 4Q ahead)	
I(1) test	**Number of I(1)**	**Proportion of I(1)**
Panel A: *Consumption forecasts*		
ADF test (5% crit. value)	84	100%
KPSS (5% crit. value)	84	100%
Panel B: *Output forecasts*		
ADF test (5% crit. value)	84	100%
KPSS (5% crit. value)	84	100%
Panel C: *Investment forecasts*		
ADF test (5% crit. value)	84	100%
KPSS (5% crit. value)	84	100%

Note: Lag selection is based on AIC criterion. Results are robust to different lag selections.

G.2 DF-GLS statistics

Figure A.4 plots the DF-GLS test statistics against the corresponding critical values associated with Panel B of Table 8. The forecasts of output are cointegrated with investment forecasts with vector $(1, -1)$ for two forecasters (with ID 504 and 510) and over all forecasting horizons.

Figure A.4: **DF-GLS test statistics vs. critical values using individual-level output and investment forecasts data**



G.3 Testing overidentifying restrictions

We impose two cointegration relations implied by stochastic growth models when estimating a Vector Error Correction Model (VECM); the two cointegration vectors are $(1, -1, 0)$ and $(1, 0, -1)$ for the forecasts of output, consumption and investment. Table A.13 reports the number and the proportion of cases where the over-identifying restrictions are not rejected by the likelihood ratio test, using individual-level forecast data.

Table A.13: **Likelihood ratio test of over-identifying restrictions (individual forecasts)**

Forecasting horizon	Number of cointegration	Proportion of cointegration
1Q	3	14.3%
2Q	4	19.0%
3Q	4	19.0%
4Q	6	23.8%

H Graphical illustration of PP and KPSS test statistics

Panel A: cointegration between forecasts of output and consumption (using individual forecasts data)

Figure A.5: **Illustration of individual level Phillips-Perron test outcomes of forecasts of output-consumption ratio**

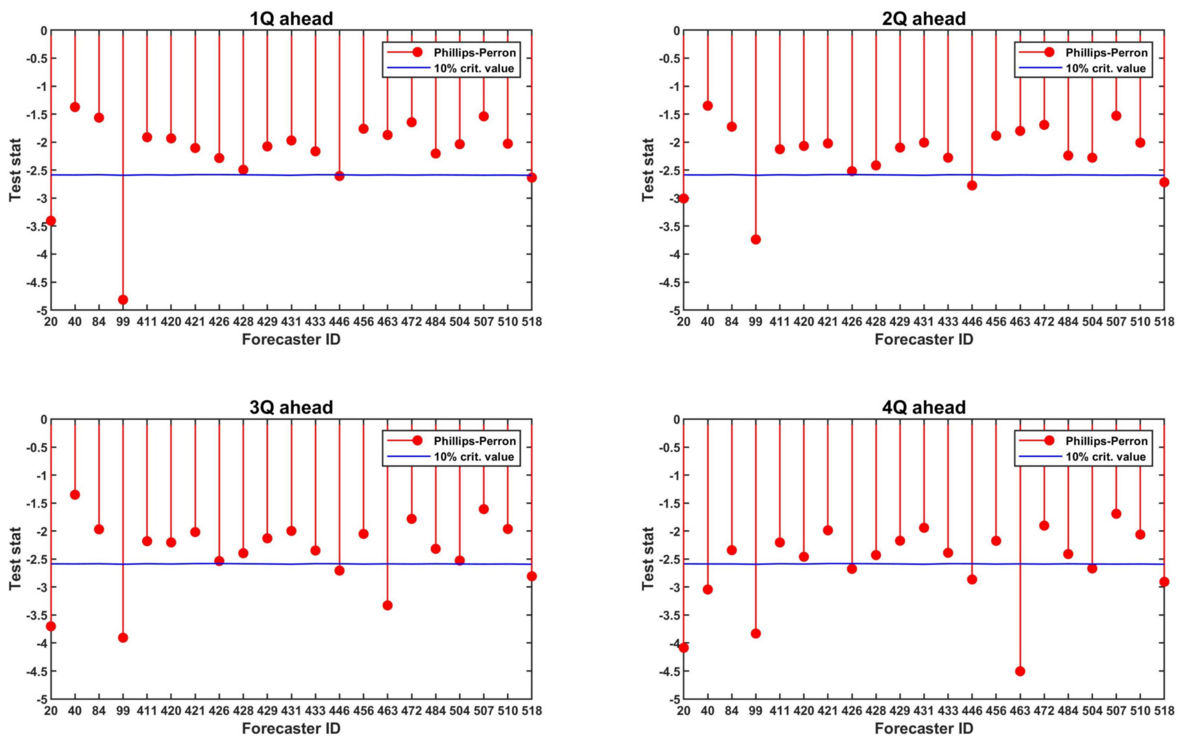
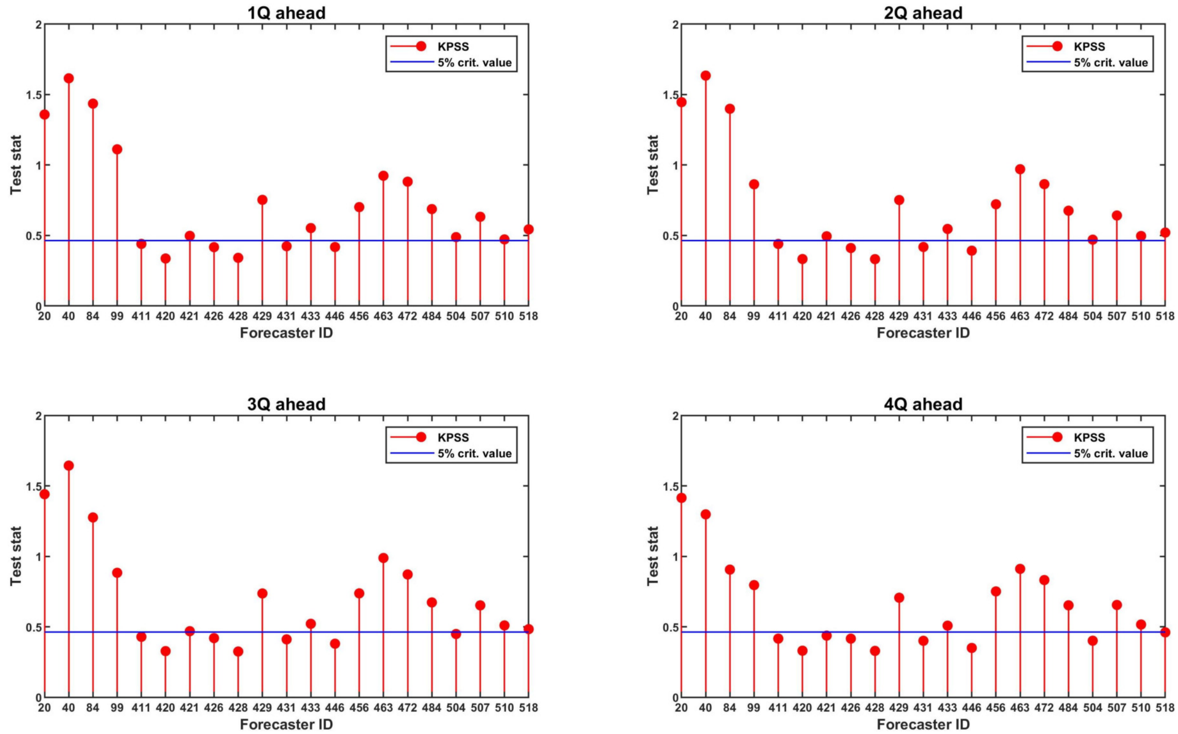


Figure A.5 and Figure A.6 visualize PP and KPSS test statistics and critical values from Panel A of Table 8 for forecasts of output-consumption ratio, respectively. The test statistics are pinned down by the circle at the end of each red stem, while the corresponding critical value locates on the blue horizontal line.

By illustrating the relationship between PP test statistics and 10% critical values, Figure A.5 shows that the majority of test statistics stay above the blue line of critical values over forecasting horizon from 1- to 4-quarter ahead, suggesting that forecasts of output-consumption ratio made by the majority of selected forecasters are non-stationary. For forecasters with ID 20, 99, 446 and 518, forecasts of output are cointegrated with forecasts of consumption with the $(1, -1)$ cointegrating vector, for all 4 forecasting horizons. Moreover, for forecasters with

Figure A.6: **Illustration of individual level KPSS test outcomes of forecasts of output-consumption ratio**



ID 463, forecasts of output-consumption ratio over 3- and 4-quarter ahead are stationary.

In Figure A.6 of KPSS test, if the circle at the higher end of a stem (signifies the test statistics) stays beyond the blue line of critical values, the corresponding test outcome indicates a rejection of the null of cointegration. For 7 individuals, with ID 411, 420, 426, 428, 431, 446 and 518, out of 21 forecasters, forecasts of output-consumption ratio are stationary over at least three forecasting horizons.

Panel B: cointegration between forecasts of output and investment (using individual forecasts data)

Figure A.7 and Figure A.8 illustrate PP and KPSS test statistics (with 10% critical values) associated with Panel B of Table 8, respectively. It is clear that the forecaster with ID 446 is the only forecaster whose test statistic circles fall below the blue line for all forecasting horizons, suggesting the PP test rejects the null of stationary forecasts of output-investment ratios over all forecasting horizons for this forecaster. Meanwhile, the forecaster with ID 40 forms stationary forecast of the ratio over 3- to 4-quarter ahead. KPSS test outcomes of forecasts of output-investment ratio illustrated in Figure A.8 show that for 5 forecasters, with ID 446, 456, 463, 472 and 484, out of 21 individuals, the test statistics over all four forecasting horizons fall short of the line of 5% critical values.

Figure A.7: **Illustration of individual level Phillips-Perron test outcomes of forecasts of output-investment ratio**

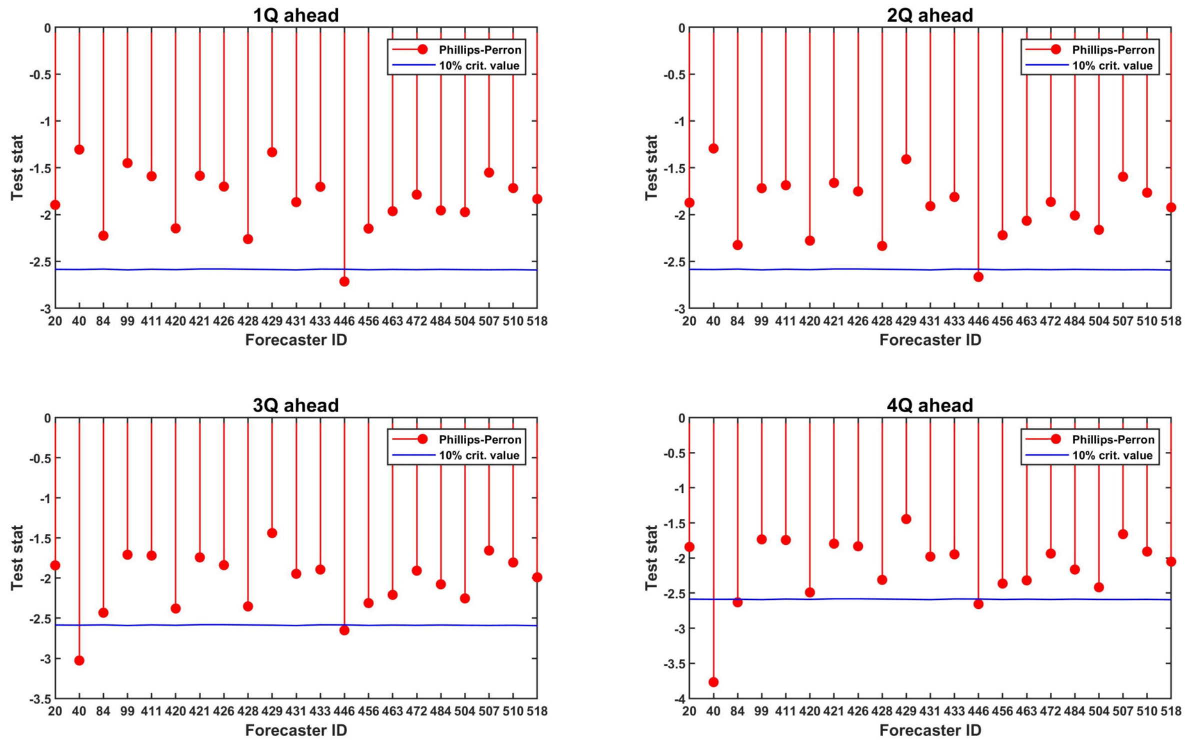
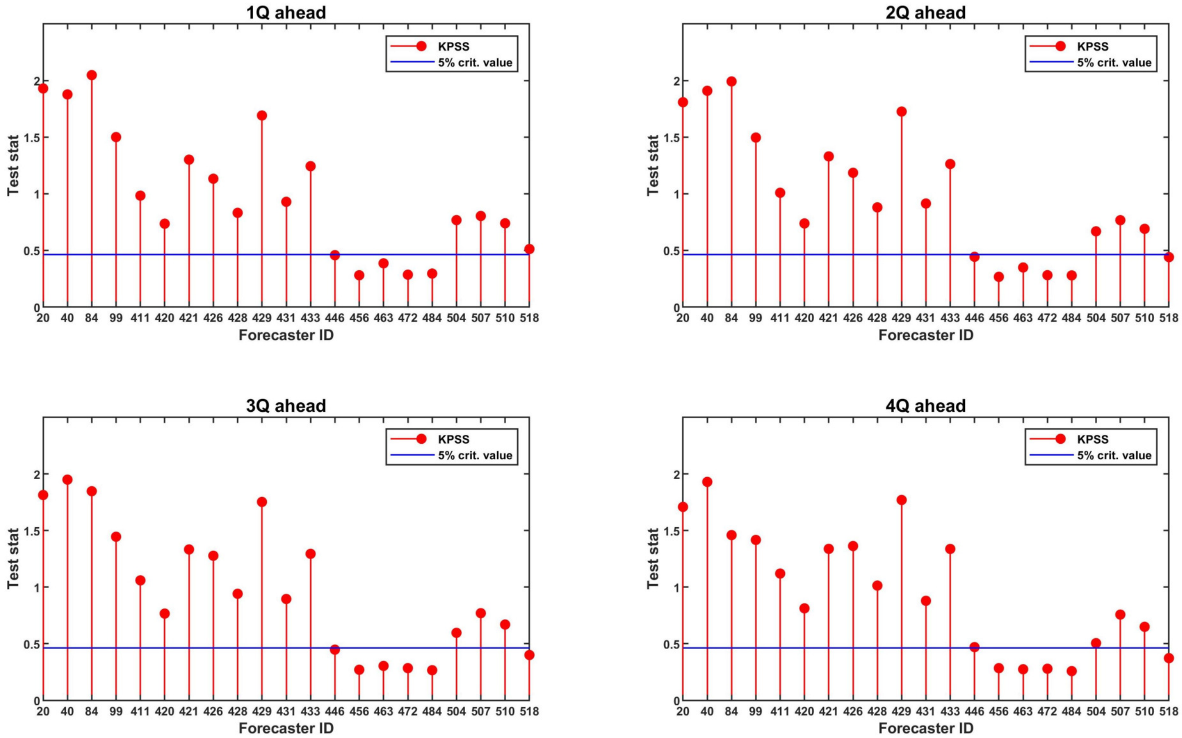


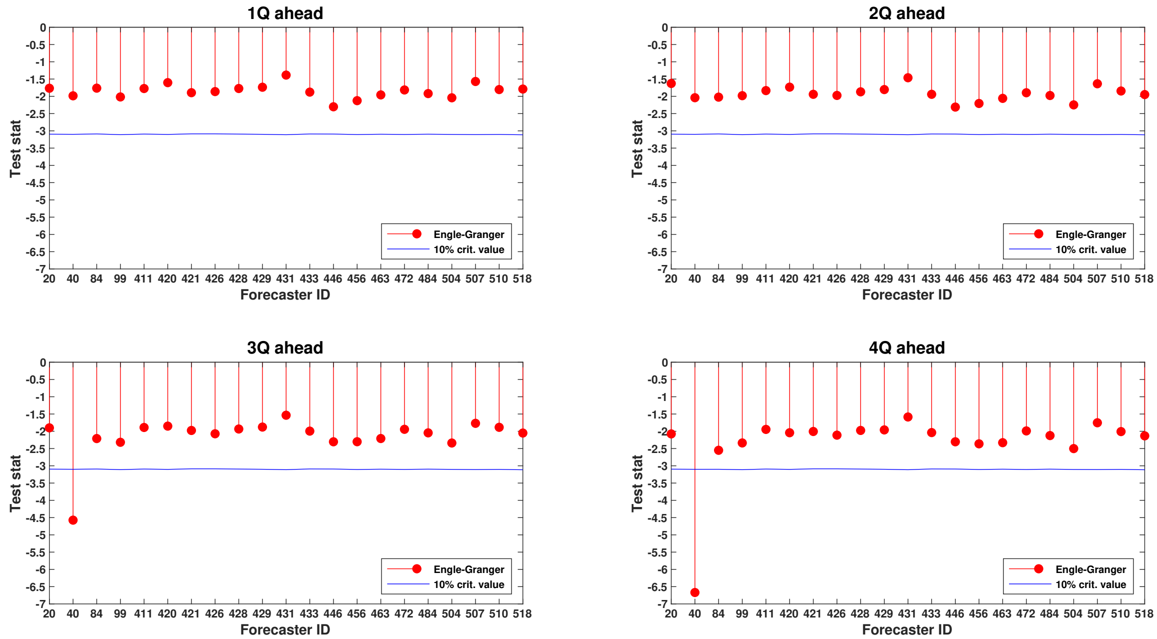
Figure A.8: **Illustration of individual level KPSS test outcomes of forecasts of output-investment ratio**



I Engle-Granger test results

Figure A.9 displays the test statistics and critical values, when the Engle-Granger test is applied to test the cointegration between individual forecasts of output and investment. For the majority of forecasters, forecasts of output and investment are not cointegrated. The test only rejects its null of no cointegration twice (for the forecaster with ID 40 when forecasting horizons are 3- and 4-quarter ahead), while the null is not rejected for remaining cases.

Figure A.9: **Engle-Granger test statistics vs critical values for testing cointegration between output and investment forecasts and without imposing $(1, -1)$ restriction**



J Multiple testing problem

Table A.14 reports corrected PP testing outcomes for both forecasts of output-consumption and output-investment ratios over 21 forecasters and four forecasting horizons, using FDR sharpened q-values (Anderson, 2008). It shows the existence of heterogeneity among forecasters in utilizing the cointegration relation in forecasting, after considering the multiple testing problem.

Table A.14: **Cointegration testing results using individual-level forecasts over the same forecasting horizon and with $(1, -1)$ restriction**

PP tests	**Using FDR sharpened q-values**	
	No. of no cointegration detected out of 84	**Proportion of no cointegration**
Forecasts of output and consumption	78	92.9%
Forecasts of output and investment	83	98.8%

Table A.15: **Test results of Pesaran panel cross-sectional dependence test**

H_0: forecasts are cross-sectionally independent.				
H_1: forecasts are cross-sectionally dependent.				
	1Q ahead	2Q ahead	3Q ahead	4Q ahead
Panel A: *cross-sectionally dependence of output-consumption forecasts*				
P-value	0.000	0.000	0.000	0.000
Average correlation coeff.	0.95	0.94	0.92	0.89
Panel B: *cross-sectionally dependence of output-investment forecasts*				
P-value	0.000	0.000	0.000	0.000
Average correlation coeff.	0.98	0.97	0.96	0.95

J.1 Cross-sectional dependence

We examine the cross-sectional dependence of forecasts of output-consumption (or output-investment) ratios across the forecasters, using the cross-sectional dependence test developed by Pesaran (2006, 2015). Table A.15 reports the p-values and average correlation coefficients of the tests over 1- to 4-quarter ahead forecasting horizons. For instance, the test shows that the p-value for 2-quarter ahead output-consumption forecast ratio is 0.000 and the average correlation coefficient is 0.94. The cross-sectional dependence tests uniformly reject the null of cross-sectional independence for both forecasts of output-consumption ratio and output-investment ratio over all horizons. And the average correlation coefficients are all close to 1, indicating the presence of highly cross-sectional dependence in our panel forecast data.

K Autocorrelations

Figure A.10 to Figure A.11 report the optimal lags using MAIC for individual forecasters' forecasts of output-consumption ratio and output-investment respectively.

Figure A.10: **Optimal lag for forecasts of output-consumption ratios**

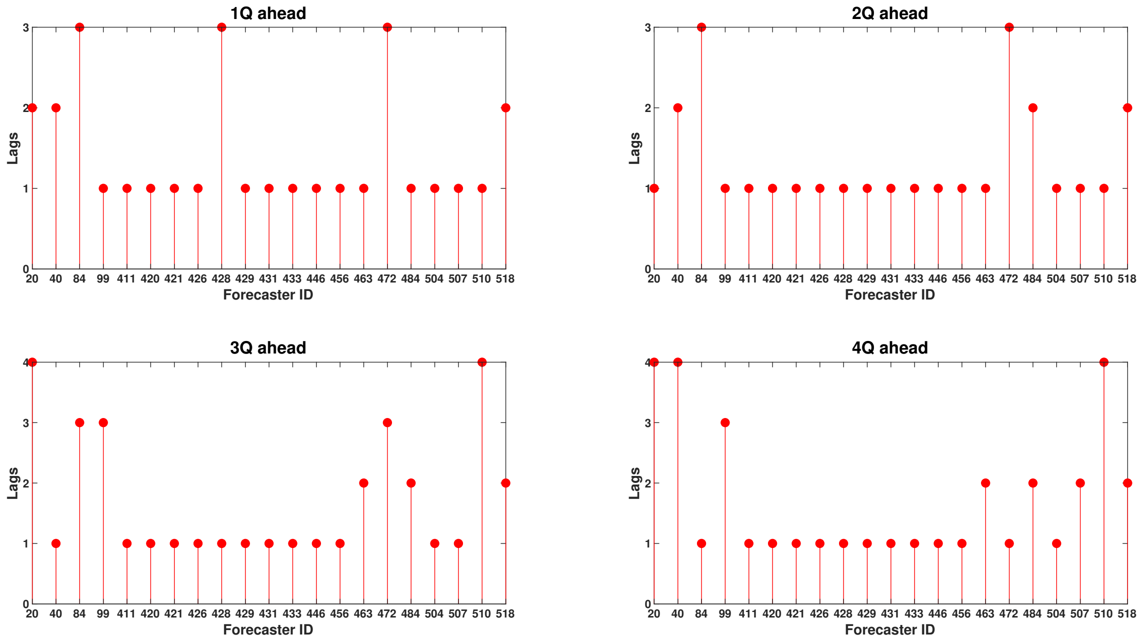


Figure A.11: **Optimal lags for forecasts of output-investment ratios**

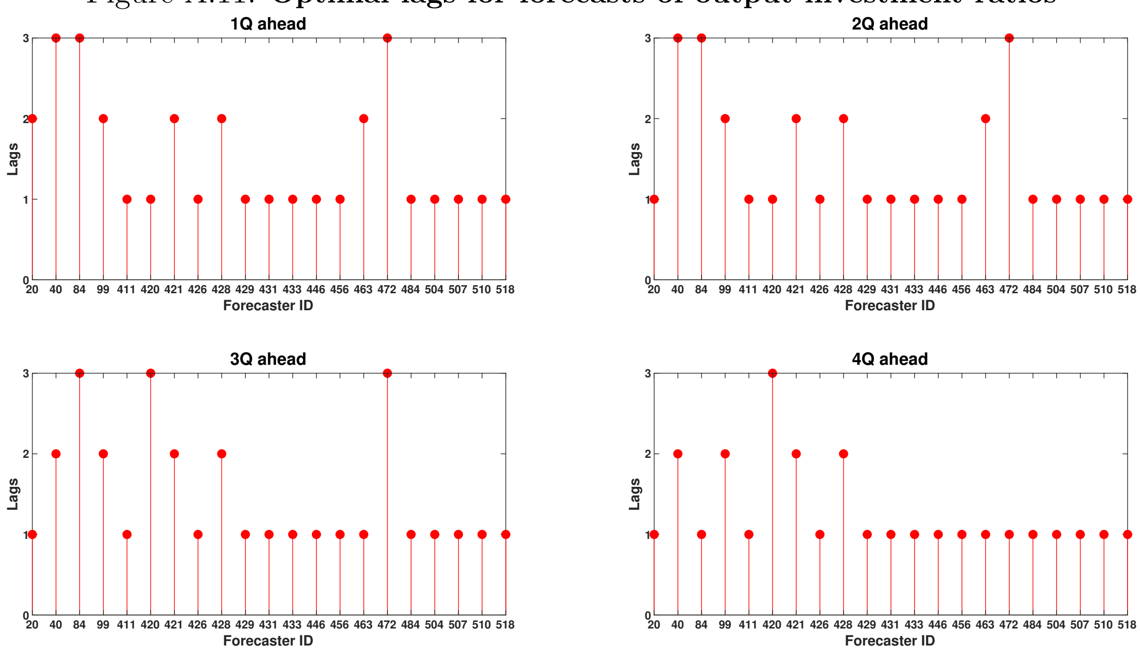
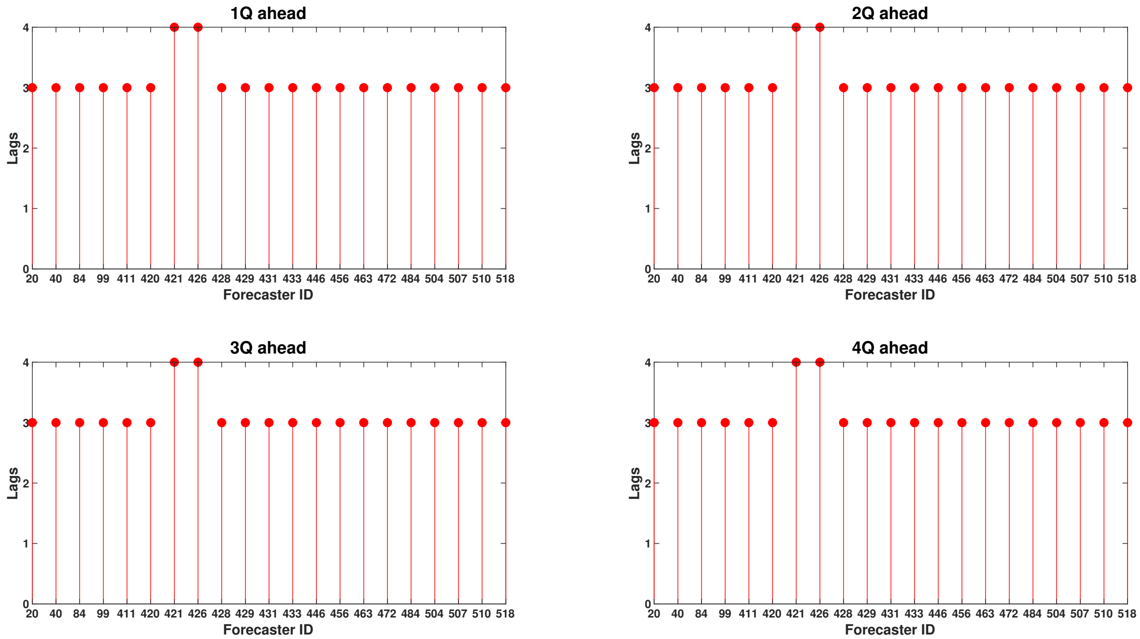


Figure A.12 plots the Newey-West optimal lags for individual-level forecasts of output-consumption ratio. The Newey-West Optimal lags for forecasts of output-investment ratio are identical to Figure A.12.

Figure A.12: **Autocorrelation (Newey-West lags) for forecasts of output-consumption ratio**



L Structural break

Table A.16 reports test outcomes of Recursive Cusum test when running the augmented Dickey-Fuller regression. Panel A and B examine the stability of estimated coefficients using median forecasts of output-consumption ratio and output-investment ratio, respectively. Two types of Recursive Cusum tests are utilized: assuming recursive residuals (Brown, Durbin and Evans, 1975) or OLS residuals (Ploberger and Krämer, 1992), respectively. Both indicate that no structural break is found in the augmented Dickey-Fuller regression across samples, as test statistics are uniformly below the corresponding 5% critical value.Table A.17 reaches a similar conclusion for mean forecasts.

Table A.16: **Cumulative sum test for parameter stability of the ADF regression with median forecasts, recursive residuals and OLS residuals**

Cumulative sum with test for parameter stability **Augmented Dickey-Fuller test**				
median	1Q	2Q	3Q	4Q
Panel A: Forecast of output-consumption ratio				
Test statistics (**recursive residuals**)	0.348	0.452	0.446	0.467
5% critical value	0.948	0.948	0.948	0.948
Test statistics **(OLS residuals)**	0.774	0.831	0.860	0.850
5% critical value	1.358	1.358	1.358	1.358
Panel B: Forecast of output-investment ratio				
Test statistics (**recursive residuals**)	0.823	0.837	0.710	0.747
5% critical value	0.948	0.948	0.948	0.948
Test statistics **(OLS residuals)**	0.838	0.834	0.848	0.852
5% critical value	1.358	1.358	1.358	1.358

Note: Lag selection is based on AIC criterion. Results are robust to different lag selections.

Table A.17: **Cumulative sum test for the coefficient stability of the augmented Dickey-Fuller regression with mean forecasts, recursive residuals and OLS residuals**

Cumulative sum with test for parameter stability **Augmented Dickey-Fuller test**				
Mean	1Q	2Q	3Q	4Q
Panel A: Forecast of output-consumption ratio				
Test statistics (**recursive residuals**)	0.379	0.373	0.376	0.390
5% critical value	0.948	0.948	0.948	0.948
Test statistics **(OLS residuals)**	0.783	0.765	0.794	0.795
5% critical value	1.358	1.358	1.358	1.358
Panel B: Forecast of output-investment ratio				
Test statistics (**recursive residuals**)	0.796	0.847	0.952*	0.946
5% critical value	0.948	0.948	0.948	0.948
Test statistics **(OLS residuals)**	0.844	0.852	0.852	0.847
5% critical value	1.358	1.358	1.358	1.358

Note: Lag selection is based on AIC criterion. Results are robust to different lag selections.

Figures A.13 illustrates the individual-level Recursive Cusum test statistics and 5% critical values assuming OLS residuals for forecasts of output-investment ratios. Test statistics (red dots) are all below the corresponding 5% critical values (blue lines). This implies that Recursive

Cusum tests uniformly indicate that no structural break is found in estimated coefficients, using the individual-level forecasts. Similar results are obtained with recursive residuals.

Figure A.13: **Cusum test statistics (OLS residuals) for parameter stability using individual-level forecasts of output-investment ratios**

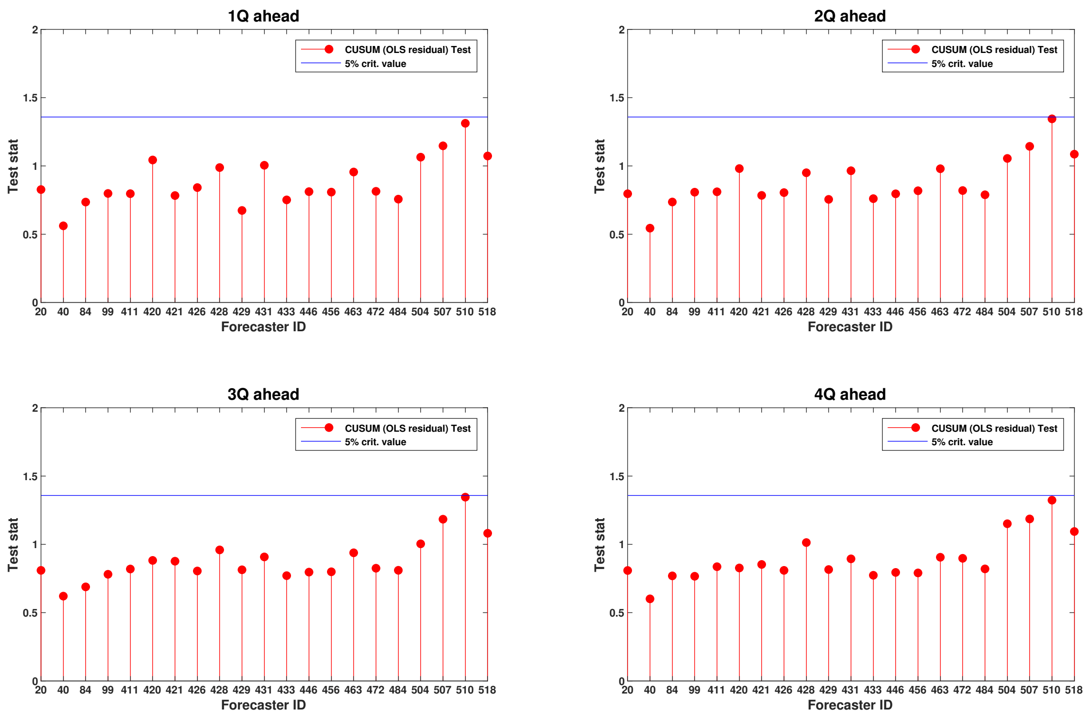


Table A.18 reports Gregory-Hansen test outcomes for median forecasts. Panel A and Panel B analyze forecasts of output-consumption ratios and output-investment ratios, respectively. As ADF and Z_t test statistics are uniformly above the corresponding 10% (and thus 5%) critical value, it implies that the forecast of output is not cointegrated with the forecast of consumption and the forecasts of investment over different forecasting horizons, even if we take the potential structure break into considerations. Similar conclusion can be derived from Table A.19 for mean forecasts.

Table A.18: **Gregory-Hansen cointegration test (ADF stats) with median output-consumption ratio forecasts**

Panel A: *cointegration between forecasts of consumption and output*				
Median	1Q ahead	2Q ahead	3Q ahead	4Q ahead
ADF test stats.	-3.99	-4.08	-4.13	-4.23
10% critical value	-4.95	-4.95	-4.95	-4.95
Z_t test stats.	-3.82	-4.03	-4.06	-4.12
10% critical value	-4.95	-4.95	-4.95	-4.95
Panel B: *cointegration between forecasts of investment and output*				
Median	1Q ahead	2Q ahead	3Q ahead	4Q ahead
ADF test stats.	-3.98	-3.93	-3.99	-4.00
10% critical value	-4.95	-4.95	-4.95	-4.95
Z_t test stats.	-3.27	-3.29	-3.36	-3.34
10% critical value	-4.95	-4.95	-4.95	-4.95

Note: lag selection is based on AIC criterion. Results are robust to different lag selections.

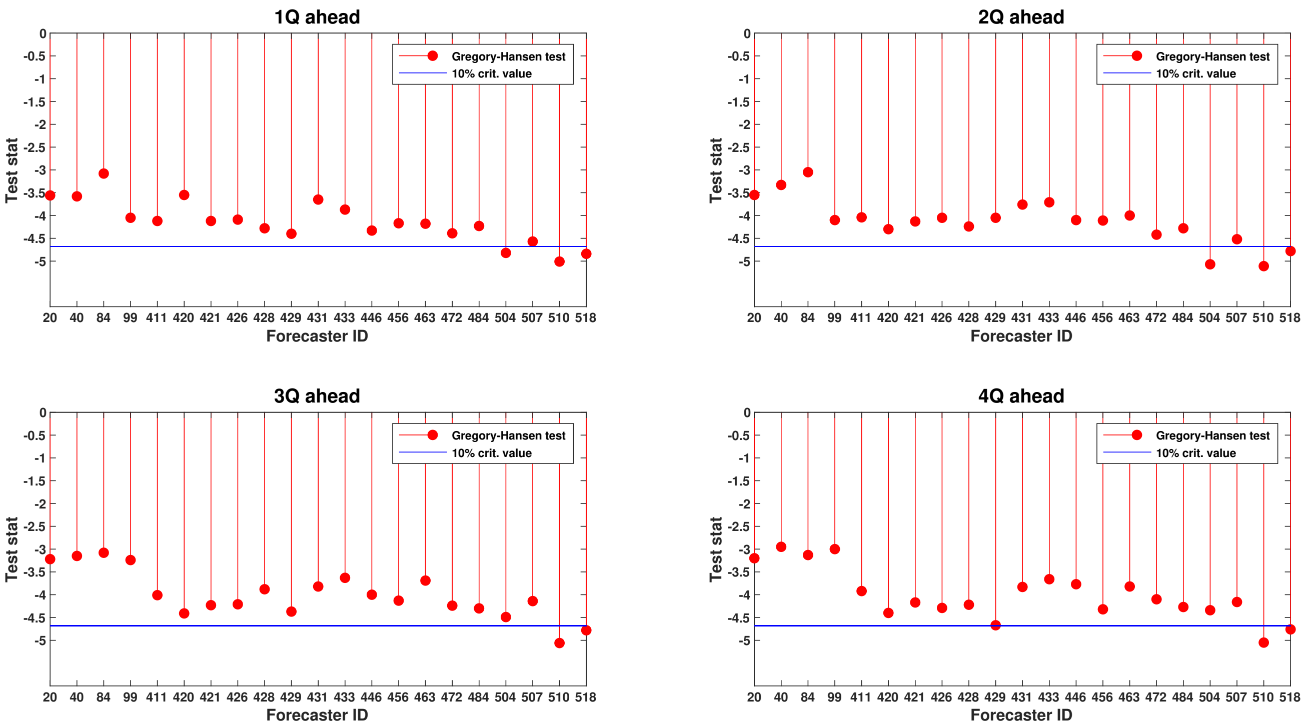
Table A.19: **Gregory-Hansen cointegration test with structural break (ADF stats) with mean output-consumption ratio forecasts**

Gregory-Hansen cointegration test with structural break				
Panel A: *cointegration between forecasts of consumption and output*				
Mean	1Q ahead	2Q ahead	3Q ahead	4Q ahead
ADF test stats.	-4.06	-4.11	-3.25	-3.14
10% critical value	-4.95	-4.95	-4.95	-4.95
Z_t test stats.	-3.88	-3.85	-3.97	-3.99
10% critical value	-4.95	-4.95	-4.95	-4.95
Panel B: *cointegration between forecasts of investment and output*				
Mean	1Q ahead	2Q ahead	3Q ahead	4Q ahead
ADF test stats.	-3.62	-3.68	-4.04	-3.65
10% critical value	-4.95	-4.95	-4.95	-4.95
Z_t test stats.	-3.28	-3.34	-3.41	-3.47
10% critical value	-4.95	-4.95	-4.95	-4.95

Note: Lag selection is based on AIC criterion. Results are robust to different lag selections.

Figure A.14 plots the Gregory-Hansen test statistics and critical values for individual forecasts of output and investment. Red dots stand for the test statistics for each individual, while the blue horizontal line corresponds to the 10% critical value. Despite of the majority of individual forecasters produce forecasts of output that are not cointegrated with their forecasts of consumption and forecasts of investment, respectively, the forecasts produced by some professional forecasters are cointegrated. There still exists heterogeneity in utilizing the long-run equilibrium relationships.

Figure A.14: **Illustration of individual level Gregory-Hansen cointegration test with output-investment ratio forecasts**



M Recursive trace tests

Figure A.15 plots the test statistics (red lines) and corresponding 5% critical values (blue lines) of recursive Johansen trace test with rank = 0 for median forecasts of output, consumption and investment. All test statistics are below the corresponding critical values, indicating that recursive trace tests fails to reject the null of no cointegrating vector against the alternative of existence of at least one cointegraing vector. Figure A.16 plots test statistics using mean forecasts.

Figure A.15: Recursive Johansen trace test (rank = 0) for common trend, median forecasts of output, consumption and investment

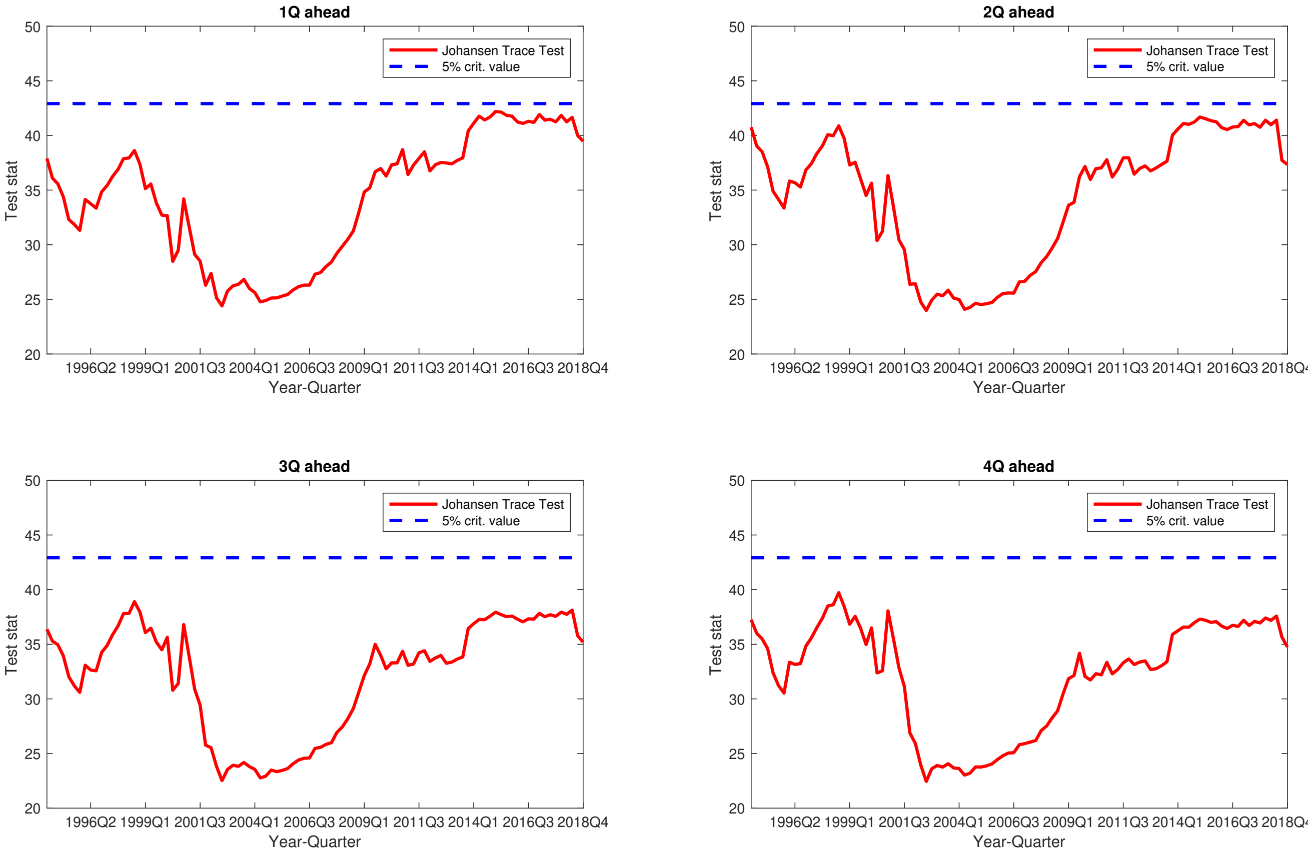


Figure A.16: Recursive Johansen trace test (rank = 0) for common trend, mean forecasts of output, consumption and investment

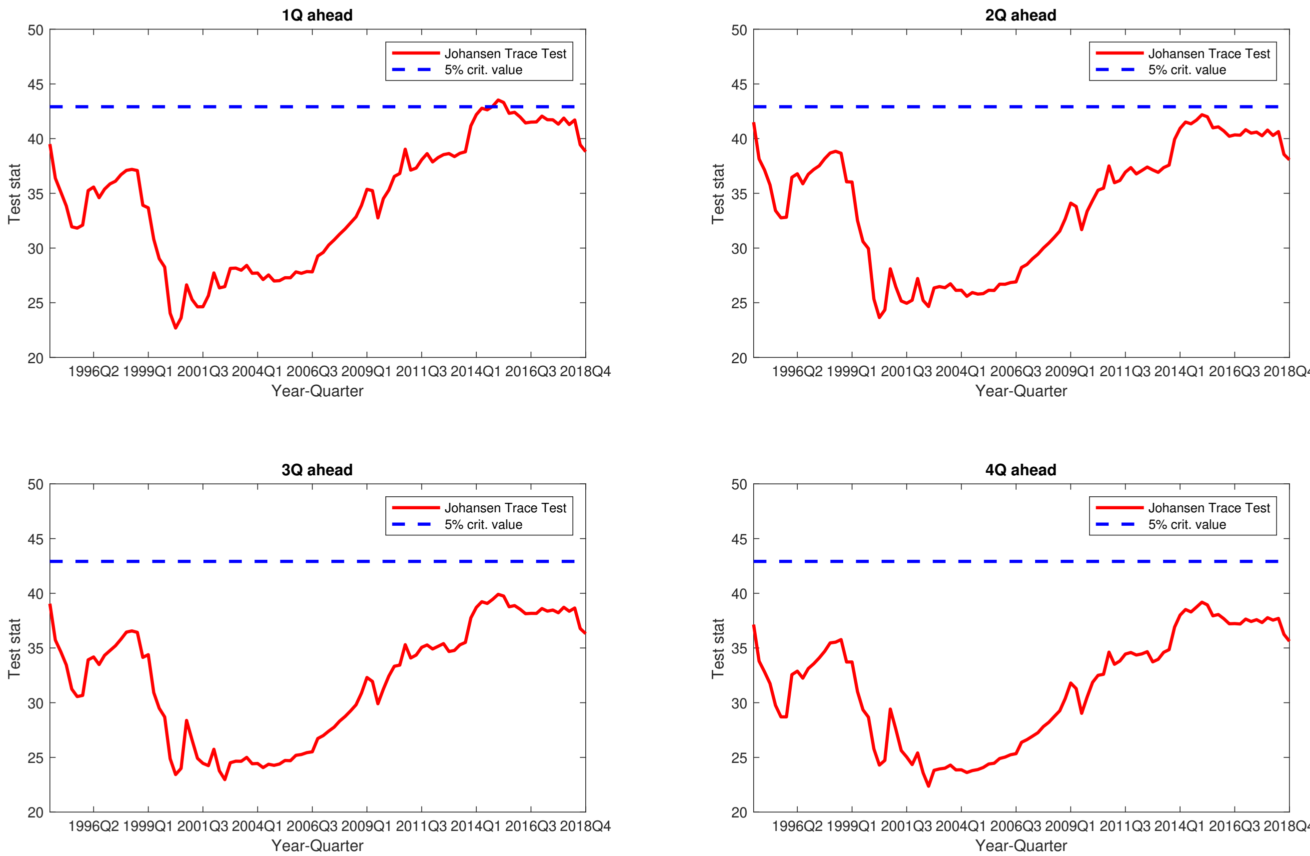
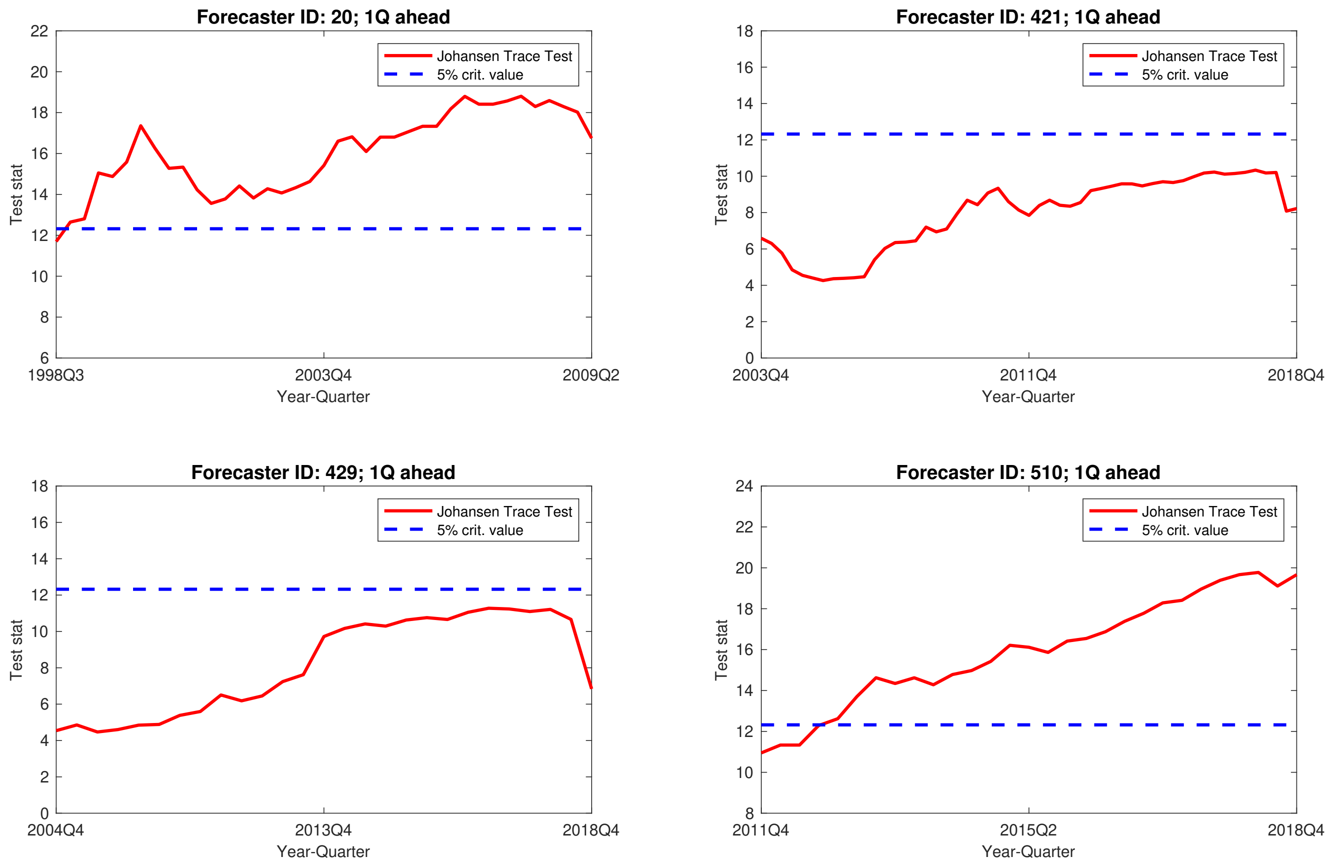


Figure A.17 illustrates the recursive trace test statistics for several forecaster IDs against the corresponding 5% critical values. For ID 20 and 510, as the sample becomes longer, the null hypothesis is rejected. For ID 421 and 429, the null hypothesis cannot be rejected for the whole rolling sample.

Figure A.17: Recursive Johansen trace test (rank = 0) for common trend, 1-quarter ahead forecasts of output, consumption and investment



N Sample size and heterogeneity by groups

The forecasters are split into three groups of different sample sizes. Table A.20 reports the proportion of non-rejection of the null hypothesis of no co-integration in different groups by the DF-GLS test and Gregory-Hansen test. Table A.21 reports the proportion of no cointegration in two groups of forecasters, i.e., those belong to financial service providers vs nonf-financial service providers.

Table A.20: **Proportions of no cointegration: by sample size**

	DF-GLS test	Gregory-Hansen test
Longest 33.3%		
Output-consumption forecasts	85.7%	89.3%
Output-investment forecasts	75.0%	85.7%
Middle 33.3%		
Output-consumption forecasts	67.9%	75.0%
Output-investment forecasts	75.0%	78.6%
Shortest 33.3%		
Output-consumption forecasts	82.1%	75.0%
Output-investment forecasts	100%	100%

Table A.21: **Proportions of no cointegration in different groups**

	DF-GLS test	Gregory-Hansen test
Financial Service Providers		
Output-consumption forecasts	65.0%	85.0%
Output-investment forecasts	80.0%	79.7%
Non-financial Service Providers		
Output-consumption forecasts	82.8%	90.0%
Output-investment forecasts	93.8%	87.5%

O Cointegration between other macroeconomic variables

O.1 Forecasts of inflation and unemployment

Table A.22 reports the integration property of mean and median forecasts of inflation and unemployment, using SPF data during 1981:Q3 to 2018:Q4. Panel A presents the Dickey-Fuller test statistics and 5% critical values for median forecasts of inflation and unemployment. Dickey-Fuller test indicates that both median forecasts of inflation and unemployment are I(0), i.e. stationary. This point is confirmed by Johansen trace and maximum-eigenvalue tests. Table A.23 reports the test outcomes. Both tests indicate that multiple cointegrating vectors are detected, suggesting the two forecasts are stationary. Similar results can be reached for mean forecasts of inflation and unemployment; the associated test outcomes are reported in Panel B of Table A.22.

Table A.22: **Integration properties of median and mean SPF forecasts, inflation and unemployment**

Stationarity test				
	1Q ahead	2Q ahead	3Q ahead	4Q ahead
Panel A: Median forecasts				
Median inflation forecasts				
Dickey-Fuller stat.	-4.280	-3.176	-3.606	-3.373
5% critical value	-1.655	-1.655	-1.655	-1.655
Median unemployment forecasts				
Dickey-Fuller stat.	-3.075	-2.973	-2.897	-2.594
5% critical value	-1.655	-1.655	-1.655	-1.655
Panel B: Mean forecasts				
Mean inflation forecasts				
Dickey-Fuller stat.	-4.225	-3.012	-3.706	-3.074
5% critical value	-1.655	-1.655	-1.655	-1.655
Mean unemployment forecasts				
Dickey-Fuller stat.	-3.099	-2.975	-2.870	-2.602
5% critical value	-1.655	-1.655	-1.655	-1.655

Note: Lag selection is based on AIC criterion. Results are robust to different lag selections.

Table A.23: **Johansen trace and maximum-eigenvalue tests for the number of common trend among median and mean forecasts of inflation and unemployment**

Johansen test				
Trace test: $J^{trace}(r)$, r = rank				
Median	r=0	5% critical	r=1	5% critical
1Q ahead	25.5	15.4	7.9	3.8
2Q ahead	21.5	15.4	8.7	3.8
3Q ahead	27.8	15.4	8.5	3.8
4Q ahead	26.9	15.4	8.9	3.8
Maximum-eigenvalue test: $max(r)$, r = rank				
Median	r=0	5% critical	r=1	5% critical
1Q ahead	17.6	14.1	7.9	3.8
2Q ahead	22.7	14.1	8.7	3.8
3Q ahead	19.3	14.1	8.5	3.8
4Q ahead	18.0	14.1	8.9	3.8

Note: Lag selection is based on AIC criterion. Results are robust to different lag selections.

We proceed to test the integration property for individual-level forecasts of inflation and unemployment. Panel A and B of Table A.24 report the numbers and the proportions of individual inflation and unemployment forecasts that are I(1), respectively. No individual inflation forecast is I(1) and only a small proportion of unemployment forecasts are I(1).

Table A.24: **Integration test results for forecasts made by individual forecasters**

Total individual forecasters:	**21**, with **4** forecasts each (1Q, 2Q, 3Q & 4Q ahead)	
I(1) test	**Number of I(1)**	**Proportion of I(1)**
Panel A: *I(1) test for individual-level inflation forecasts*		
ADF test (5% crit. value)	0	0%
Panel B: *I(1) test for individual-level unemployment forecasts*		
ADF test (5% crit. value)	16	19%

Note: Lag selection is based on AIC criterion. Results are robust to different lag selections.

O.2 Forecasts of nominal interest and inflation

If real interest rate is stationary, nominal interest rate and inflation rate are cointegrated according to the Fisher equation. Particularly, they are cointegrated with vector $(1, -1)$. Table A.25 reports the integration property of median and mean forecasts of nominal interest rate. The Dickey-Fuller test indicates that median or mean nominal interest rate forecasts are I(1). Inflation forecasts are I(0), as is reported in Table A.22. Theoretically, there exists no cointegration vector between the two forecasts. This is confirmed by applying recursive Johansen trace test for rank $= 0$, as is shown in Figure A.18 and A.19.

Table A.25: **Integration properties of median and mean SPF forecasts of nominal interest rate and inflation**

Stationarity (I(1)) test				
	1Q ahead	2Q ahead	3Q ahead	4Q ahead
Median nominal interest forecasts				
Dickey-Fuller stat.	-1.862	-1.622	-1.646	-1.649
5% critical value	-2.887	-2.887	-2.887	-2.887
Mean nominal interest forecasts				
Dickey-Fuller stat.	-1.967	-1.883	-1.848	-1.919
5% critical value	-2.887	-2.887	-2.887	-2.887

Note: Lag selection is based on AIC criterion. Results are robust to different lag selections.

Table A.26: **Cointegration test for median SPF forecasts with cointegrating vector** $(1, -1)$

Panel A: *no cointegration between median forecasts of nominal interest and inflation*					
Median	1Q ahead	2Q ahead	3Q ahead	4Q ahead	4Q $R^{nominal}$ & 1Q π
PP (Z_t test)	-1.374	-1.402	-1.473	-1.689	-1.743
10% critical value	-2.577	-2.577	-2.577	-2.577	-2.577
DF-GLS	-1.535	-1.911	-1.816	-2.297	-2.104
10% critical value	-2.681	-2.636	-2.681	-2.649	-2.649
Panel B: *no cointegration between mean forecasts of nominal interest and inflation*					
Mean	1Q ahead	2Q ahead	3Q ahead	4Q ahead	4Q $R^{nominal}$ & 1Q π
PP (Z_t test)	-1.377	-1.358	-1.438	-1.577	-1.743
10% critical value	-2.577	-2.577	-2.577	-2.577	-2.577
DF-GLS	-1.517	-1.477	-1.677	-1.750	-1.809
10% critical value	-2.681	-2.681	-2.681	-2.671	-2.664

Note: Lag selection is based on AIC criterion. Results are robust to different lag selections. Lag selection for Phillips-Perron test is Min MAIC criterion. Results are robust to different lag selections.

Figure A.18: Recursive Johansen trace test (rank = 0) for common trend, median forecasts of nominal interest and inflation

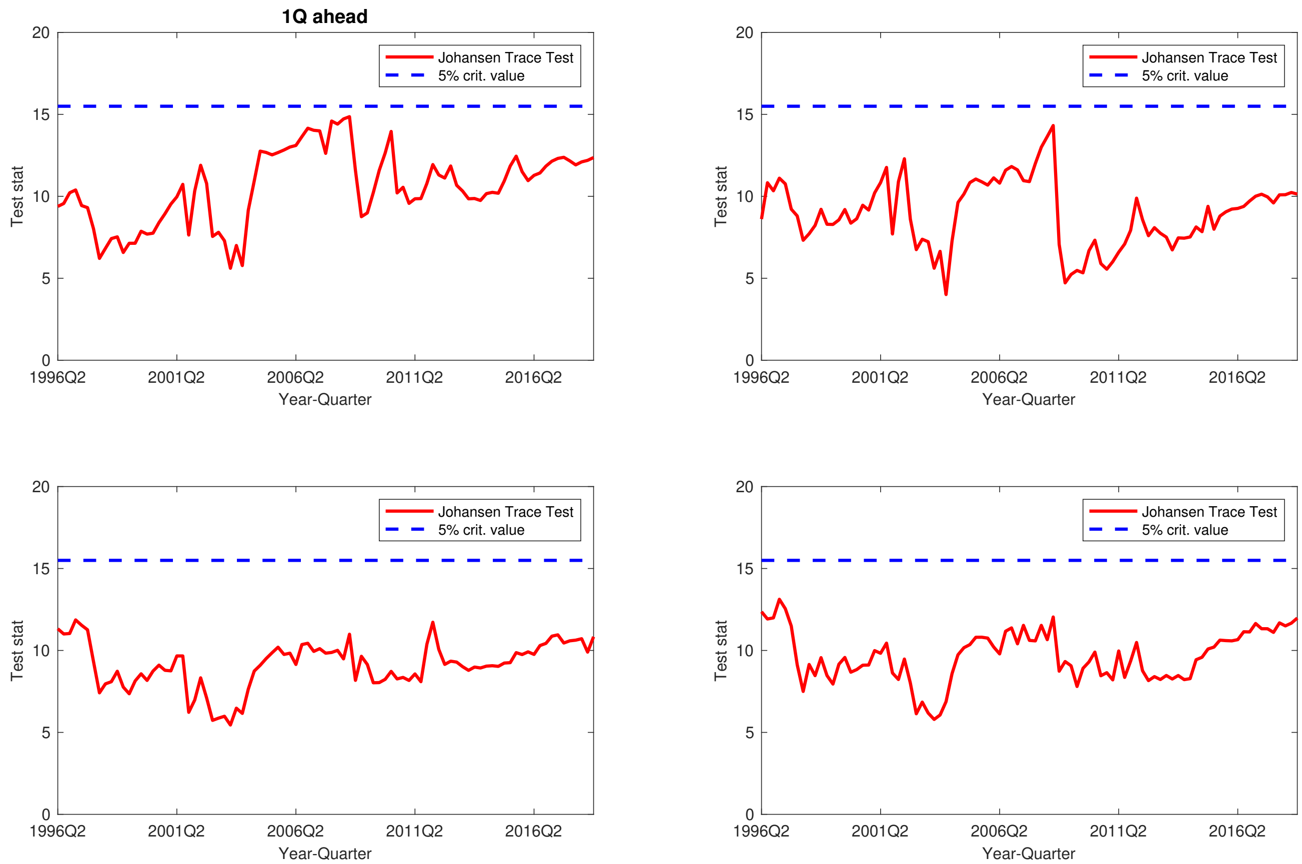
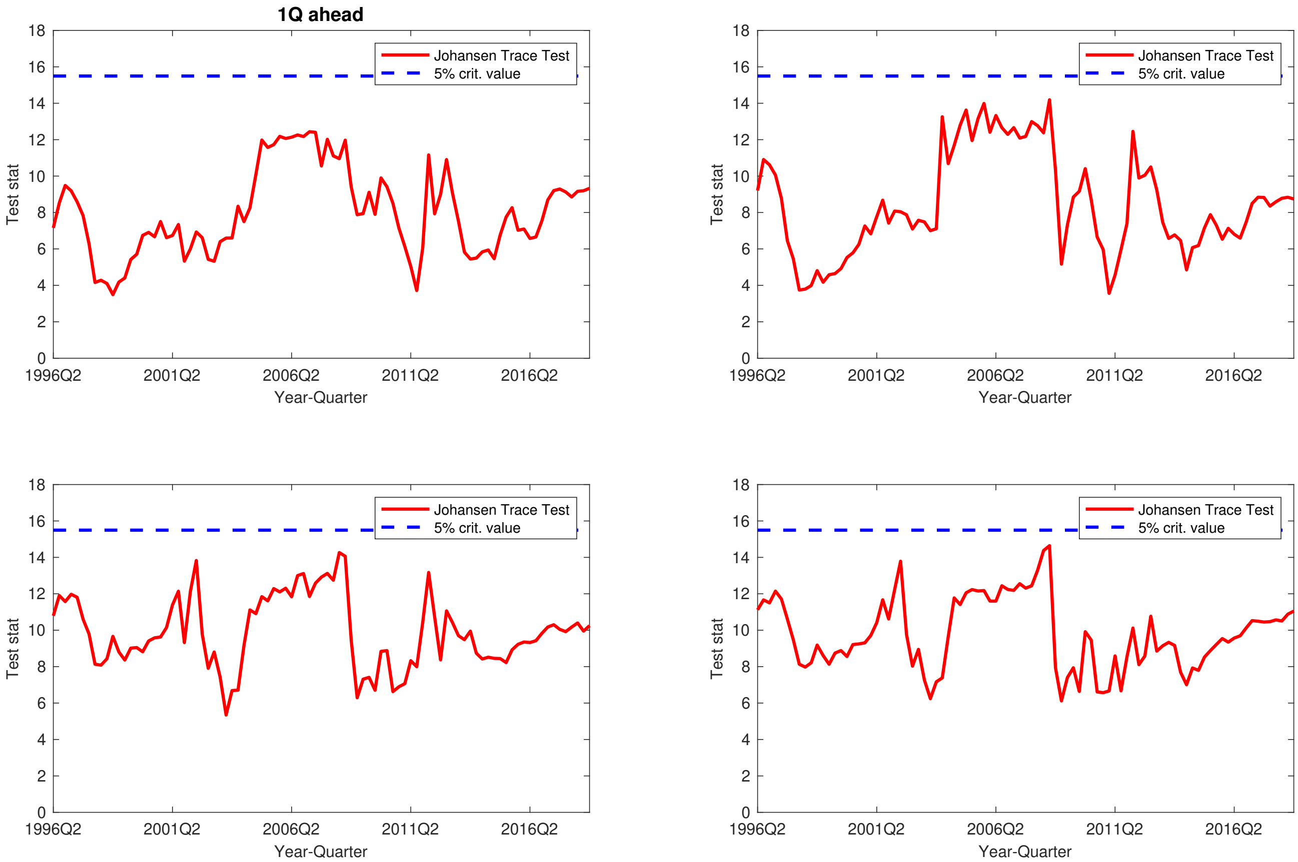


Figure A.19: Recursive Johansen trace test (rank = 0) for common trend, mean forecast of nominal interest and inflation



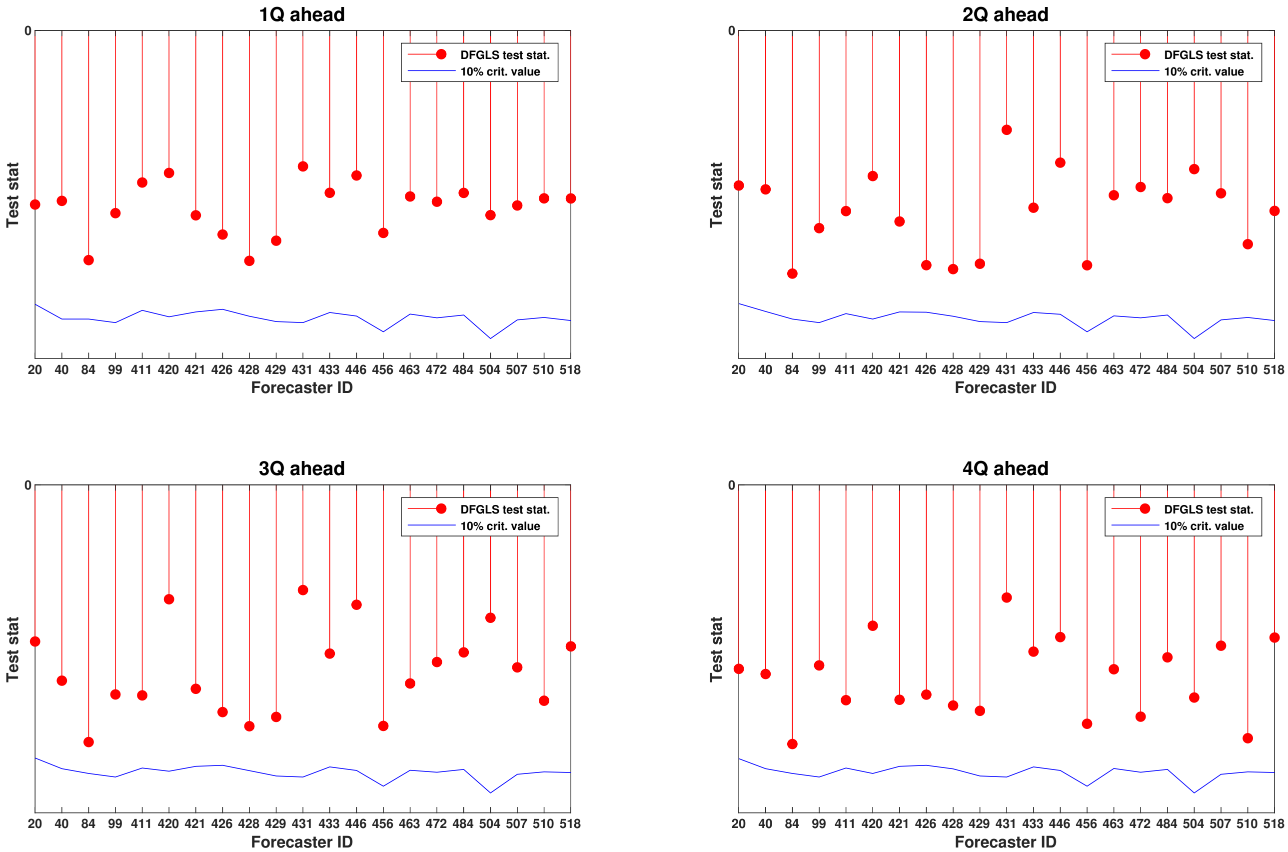
Next, we test the integration property of individual-level forecasts. Table A.27 reports the numbers and proportion of individual-level forecasts that are I(1). We find that the all individual forecasts of nominal interest rate are I(1). Again, this implies that for each forecaster, there exists no cointegrating vector between forecasts of nominal interest rate and inflation, including the theoretical vector $(1, -1)$. Using the DF-GLS test, Figure A.20 confirms that for each invidual, the forecasts of nominal interest rate and inflation are not cointegrated with vector $(1, -1)$.

Table A.27: **Integration test results for forecasts made by individual forecasters**

Total individual forecasters:	**21**, with **4** forecasts each (1Q, 2Q, 3Q & 4Q ahead)	
I(1) test	**Number of I(1)**	**Proportion of I(1)**
I(1) test for individual-level nominal interest rate forecasts		
ADF test (5% crit. value)	84	100%

Note: Lag selection is based on AIC criterion. Results are robust to different lag selections.

Figure A.20: DF-GLS test outcomes for individual-level forecasts of nominal interest rate and inflation



P Forecasting accuracy of SPF and fitted models

P.1 Forecasting accuracy of SPF: utilizing vs. without utilizing long-run relationships

This Appendix firstly evaluates the accuracy of SPF forecasts (of output, consumption and investment) made by forecasters who utilize (or do not utilize) the long-run relationships. Forecasters are divided into two groups: those who utilize a cointegration relationship and those who do not.[5] Table A.28 reports the accuracy of forecasts which is measured by root-mean-square errors (RMSEs) over 1-, 2-, 3-, and 4-quarter horizons.

In Panel A, the block "YC cointegrated" (or "YC not cointegrated") reports the average

[5]This division is based on the DF-GLS test results.

root-mean-square errors among forecasters who utilize (or do not utilize) the cointegration relation between consumption (C) and output (Y) in forecasting. Moreover, The row "Number of forecasters" reports the number of forecasters in each group. For example, the statistic 0.00511 is the average RMSE for 1Q-ahead forecasts of consumption growth rates among the group of forecasters who does not utilize the cointegration relation between output and consumption in forecasting. And the number 18 is the number of forecasters in this group. The results suggest forecasters who do not utilize this cointegration relation in forecasting make slightly more accurate forecasts of output and consumption. Similarly, in Panel B, forecasters who do not use the cointegration relation between output and investment (I) in forecasting generally make slightly more accurate forecasts than those who use them with three exceptions (1-, 3-, 4-quarter ahead forecasts of output growth rates).

Table A.28: **Average root-mean-square errors for each group**

Average root-mean-square errors, 1981:Q3 - 2018:Q4				
	1Q	2Q	3Q	4Q
Panel A: *Output and consumption forecasts*				
YC cointegrated				
C growth forecasts	0.0060	0.0096	0.01349	0.01921
Y growth forecasts	0.0093	0.0216	0.02051	0.02654
Number of forecasters	3	4	5	4
YC not cointegrated				
C growth forecasts	0.00511*	0.00861*	0.01212*	0.01596*
Y growth forecasts	0.00783*	0.01796*	0.01804*	0.02335*
Number of forecasters	18	17	16	17
Panel B: *Output and investment forecasts*				
YI cointegrated				
I growth forecasts	0.03378	0.05678	0.07550	0.04189
Y growth forecasts	0.00798*	0.02163	0.01829*	0.02278*
Number of forecasters	2	2	2	2
YI not cointegrated				
I growth forecasts	0.03063*	0.05068*	0.06921*	0.03558*
Y growth forecasts	0.00804	0.01796*	0.01869	0.02394
Number of forecasters	19	19	19	19

*: asterisk indicates the corresponding RMSE statistic is smaller, comparing to the other group.

P.2 Fitting recursive forecasting models and out-of-sample evaluations

This Appendix approximates the modeling of expectation formation process of the forecasters who use or do not use the long-run relationships in forecasting. One way to approximate is fitting parsimonious recursive forecasting models (constant gain learning algorithms) to the data, as in e.g., Branch and Evans (2006). The recursive forecasting models we estimate might contribute to the setup of structural business cycle models with heterogeneous expectations for future studies. Moreover, the section examines the out-of-sample forecasting properties of the fitted forecasting models.

Denote by ΔY_t, ΔC_t, and ΔI_t the growth rate of output, consumption and investment from time $t-1$ to t. We firstly introduce the forecasting models to approximate the expectation formation processes and then the methodology of the empirical exercise. There are, of course, many alternative forecasting models which can be fitted to the data. For illustration, the section considers some simple parsimonious forecasting models.

P.2.1 A parsimonious forecasting model with utilizing cointegration relationships (Model A)

We consider a parsimonious forecasting model which features cointegration among output, consumption and investment, labeled as "Model A". The model approximates the expectation formation process of the forecasters who utilize the long-run relationships. Mathematically, Model A is

$$\begin{pmatrix} \Delta Y_t \\ \Delta C_t \\ \Delta I_t \end{pmatrix} = \begin{pmatrix} \theta_{\Delta Y,t} & \phi_{\Delta Y1,t} & \phi_{\Delta Y2,t} & \phi_{\Delta Y3,t} \\ \theta_{\Delta C,t} & \phi_{\Delta C1,t} & \phi_{\Delta C2,t} & \phi_{\Delta C3,t} \\ \theta_{\Delta I,t} & \phi_{\Delta I1,t} & \phi_{\Delta I2,t} & \phi_{\Delta I3,t} \end{pmatrix} \begin{pmatrix} 1 \\ \Delta Y_{t-1} \\ \Delta C_{t-1} \\ \Delta I_{t-1} \end{pmatrix} + \begin{pmatrix} \alpha_{\Delta Y,t} & \beta_{\Delta Y,t} \\ \alpha_{\Delta C,t} & \beta_{\Delta C,t} \\ \alpha_{\Delta I,t} & \beta_{\Delta I,t} \end{pmatrix} \begin{pmatrix} Y_{t-1} - C_{t-1} \\ Y_{t-1} - I_{t-1} \end{pmatrix} + \begin{pmatrix} z_{1,t} \\ z_{2,t} \\ z_{3,t} \end{pmatrix}. \tag{A.2}$$

The parameter vector $A_{Z,t} = \begin{pmatrix} \theta_{Z,t} & \phi_{Z1,t} & \phi_{Z2,t} & \phi_{Z3,t} & \alpha_{Z,t} & \beta_{Z,t} \end{pmatrix}'$ is recursively updated by the learning algorithm

$$A_{Z,t} = A_{Z,t-1} + \gamma_Z R_t^{-1} X_t (Z_t - A'_{Z,t-1} X_{t-1}), \tag{A.3}$$

$$R_t = R_{t-1} + \gamma_Z (X_{t-1} X'_{t-1} - R_{t-1}), \tag{A.4}$$

where $X_t = \begin{pmatrix} 1 & \Delta Y_{t-1} & \Delta C_{t-1} & \Delta I_{t-1} & Y_{t-1} - C_{t-1} & Y_{t-1} - I_{t-1} \end{pmatrix}'$ and $Z = \Delta Y$, ΔC, or ΔI. The gain parameters γ_Z are assumed to be constant because constant gain learning rules are typically associated with good forecasting properties, see e.g., Branch and Evans (2006). But for generality, they are allowed to be different across equations.

P.2.2 A parsimonious forecasting model without utilizing cointegration relationships (Model B)

Here forecasters are assumed to use a simple AR(1) model for the growth rate of output, consumption and investment, labeled as "Model B". This model approximates the expectation formation process of the forecasters who do not utilize the long-run relationships among output, consumption and investment, in line with a potential cause for the survey evidence identified in Section 4.5. Mathematically, Model B is

$$\begin{pmatrix} \Delta Y_t \\ \Delta C_t \\ \Delta I_t \end{pmatrix} = \begin{pmatrix} \alpha_{\Delta Y,t} & \beta_{\Delta Y,t} & 0 & 0 \\ \alpha_{\Delta C,t} & 0 & \beta_{\Delta C,t} & 0 \\ \alpha_{\Delta I,t} & 0 & 0 & \beta_{\Delta I,t} \end{pmatrix} \begin{pmatrix} 1 \\ \Delta Y_{t-1} \\ \Delta C_{t-1} \\ \Delta I_{t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \epsilon_{3,t} \end{pmatrix}. \tag{A.5}$$

The parameter vector $b_{Z,t} = \begin{pmatrix} \alpha_{Z,t} \\ \beta_{Z,t} \end{pmatrix}$ is updated by the rule

$$b_{Z,t} = b_{Z,t-1} + \tilde{\gamma_Z} R_t^{-1} X_{t-1} (Z_t - b'_{Z,t-1} X_{t-1}), \tag{A.6}$$

$$R_t = R_{t-1} + \tilde{\gamma_Z} (X_{t-1} X'_{t-1} - R_{t-1}), \tag{A.7}$$

where $X_t = \begin{pmatrix} 1 \\ Z_t \end{pmatrix}$ and $Z = \Delta Y$, ΔC, or ΔI. Again, the gain parameters $\tilde{\gamma_Z}$ are assumed to be constant but can be different across equations.

P.3 Forecasting accuracy of fitted models

We follow the approach of Branch and Evans (2006) by dividing the sample of realized data into three periods. The initial (pre-forecasting) period, corresponding to 1947:Q2-1969:Q4, is the sample period during which agents' prior beliefs used for the forecasting models. The second period, corresponding to 1970:Q1-1981:Q2, is the in-sample period during which the optimal gain parameter γ is determined (as explained below). The last period is the out-of-sample forecasting period 1981:Q3-2018:Q4 corresponding to the sample period of the SPF data. [6]

Given a gain parameter $\gamma_j \in (0,1)$, we calculate the mean square forecast error for the in-sample period

$$MSE(Z_j) = \frac{1}{T}\textstyle\sum_{t=t_0}^{T}(Z_t - \hat{Z}_{j,t})^2,$$

where Z_t is the actual growth rate of a variable (output, consumption or investment) and $\hat{Z}_{j,t}$ is the 1-quarter ahead forecast of Z_t generated from the Model A or B given γ_j. t_0 and T denote the start and the end of the in-sample period, with $t_0 =$ 1970:Q1 and $T =$ 1981:Q2. We select the optimal in-sample parameter γ^* which minimizes the root-mean-square forecast errors.

The calibrated optimal gain parameters are reported in the following table. For model A, the optimal gain parameters for forecasting the growth rate of output, consumption and investment are 0.042, 0.031, and 0.032, respectively. For model B, the optimal gain parameters for forecasting the growth rate of output, consumption and investment are 0.010, 0.035, and 0.001, respectively. They are in the range of the values of the gain parameter found in the literature, see e.g., Branch and Evans (2006), Eusepi and Preston (2011) and Kuang and Mitra (2016). We can find that in Model A the gain parameters for different variables are closer relative to those in Model B because of the cointegration relation.

[6]The relative accuracy outcomes of the two fitted models are robust to different lengths of the pre-forecasting, the in-sample and the out-of-sample periods. Here, we demonstrate results following the same selections of the pre-forecasting and the out-of-sample periods in Branch and Evans (2006).

Estimated gain parameters	Model A	Model B
GDP growth	0.042	0.010
Consumption growth	0.031	0.035
Investment growth	0.032	0.001

.

We now compare the out-of-sample forecasting performance of Model A and B during 1981:Q3-2018:Q4. Table A.29 reports root-mean-square forecast errors for both models, with the optimal gain parameters chosen using in-sample data. The results show that Model B (without utilizing the cointegration relations) generally outperforms Model A (utilizing the cointegration relations) by generating smaller forecasting errors for output growth, consumption growth and investment growth over 1-, 2-, 3-, and 4-quarter ahead with only two exceptions (1-quarter ahead output growth forecasts and 1-quarter ahead investment growth forecasts).

Table A.29: **Comparisons of fit between models**

Out-of-sample period: 1981:Q3–2018:Q4			
Root-mean-square forecast error			
	Forecasting horizon	Model A	Model B
Output growth	1Q	0.00662*	0.00691
	2Q	0.00786	0.00741*
	3Q	0.00817	0.00777*
	4Q	0.00819	0.00780*
Consumption growth	1Q	0.00485	0.00481*
	2Q	0.00503	0.00497*
	3Q	0.00532	0.00516*
	4Q	0.00560	0.00529*
Investment growth	1Q	0.02582*	0.03122
	2Q	0.03260	0.03165*
	3Q	0.03357	0.03200*
	4Q	0.03370	0.03211*

The table reports the root-mean-square forecast error in out-of-sample forecasting of actual GDP, consumption and investment growth. *: asterisk indicates the corresponding model has smaller RMSE than the other model and generates more accurate forecasts with respect to the actual growth data.

Reference

Anderson, M., "Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects," Journal of the American Statistical Association, 2008, 103 (484), 1481–1495.

Branch, W. and G. Evans, "A simple recursive forecasting model," Economics Letters, 2006, 91 (2), 158–166.

Brown, R., J. Durbin, and J. Evans, "Techniques for testing the constancy of regression relationships over time," Journal of the Royal Statistical Society. Series B (Methodological), 1975, 37 (2), 149–192.

Eusepi, S. and B. Preston, "Expectations, learning, and business cycle fluctuations," Amer- ican Economic Review, 2011, 101 (6), 2844–2872.

Fisher, J., "The dynamic effects of neutral and investment-specific technology shocks," Journal of Political Economy, 2006, 114 (3), 413–451.

Ghysels, E. and J. Miller, "On the size distortion from linearly interpolating low frequency series for cointegration tests," Essays in honor of Peter C.B. Phillips (Advances in Econo- metrics), 2014.

King, R., C. Plosser, and S. Rebelo, "Production, growth and business cycles: I. The basic neoclassical model," Journal of Monetary Economics, 1988, 21 (2), 195–232.

Kuang, P. and K. Mitra, "Long-run growth uncertainty," Journal of Monetary Economics, 2016, 79, 67–80.

Pesaran, H., "Estimation and inference in large heterogeneous panels with cross section dependence.," Econometrica, 2006, 74, 967–1012.

Pesaran, H., "A simple unit root test in the presence of corss-section dependence," Journal of Applied Econometrics, 2007, 22, 265–312.

Ploberger, W. and W. Kramer, "The Cusum test with Ols residuals," Econometrica, 1992, 60 (2), 271–285.

Ryan, F. and D. Giles, "Testing for unit roots in economic time-series with missing observations," Advances in Econometrics, 1998.

Schmitt-Grohe, S. and M. Uribe, "Business cycles with a common trend in neutral and investment-specific productivity," Review of Economic Dynamics, 2011, 14, 122–135.