Forecast disagreement about long-run macroeconomic relationships

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

DOI:
10.1016/j.jebo.2022.06.002

Forecast Disagreement about Long-run Macroeconomic Relationships*

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

June 7, 2022

Abstract

Using survey forecast data, this paper studies whether professional forecasters utilize long-run cointegration relationships among macroeconomic variables to forecast the future, as postulated in stochastic growth models. Significant heterogeneity exists among forecasters. The majority of the forecasters do not use these long-run relationships. The results are robust across different groups, to addressing the multiple testing problem and to allowing for structural break.

Keywords: Survey expectation, Cointegration, Disagreement

JEL classifications: D84, G22, O41

*Author contacts: Pei Kuang, University of Birmingham, p.kuang@bham.ac.uk. Tel: (+44) 0121 4145620. Address: Birmingham Business School, University House, Edgbaston Park Road, Birmingham B15 2TT, United Kingdom. Li Tang, Middlesex University London and University of Essex, l.tang@mdx.ac.uk. Address: The Burroughs, London NW4 4BT. Renbin Zhang: Shandong University, zhang.renbin.ken@gmail.com. Tongbin Zhang: Shanghai University of Finance and Economics. Address: Shanghai University of Finance and Economics, Guoding Road 777, Shanghai, China. We are grateful to Yannis Karavias for advice on econometric testing. We thank the editor, Joscha Beckmann, and two anonymous referees for constructive and very helpful comments.

1 Introduction

Economic agents usually have heterogeneous expectations about macroeconomic outcomes. A large literature analyzes theoretical models with heterogeneous expectations. Heterogeneity in expectations can lead to non-fundamental driven business cycle fluctuations (Lorenzoni, 2009; Angeletos and La'O, 2013; Ilut and Schneider, 2014), extreme long-run wealth distribution (Blume and Easley, 2006), speculative bubbles in asset prices (Scheinkman and Xiong, 2003; Xiong and Yan, 2010), and can influence the effectiveness of monetary policy (Branch and McGough, 2009; Kurz et al., 2018). Empirically, heterogeneous expectations can be described as the disagreement among forecasters in the survey data which is shown to change over the business cycle (Dovern, Fritsche and Slacalek, 2012) and can be used to measure economic uncertainty (Lahiri and Sheng, 2010; Beckmann, 2021).

An important source of heterogeneous expectations is the different information sets possessed by agents, such as in the cases presented in Mankiw, Reis and Wolfers (2003) and Coibion and Gorodnichenko (2012). Another source is to be found in the different subjective forecasting models adopted by agents, e.g., Andrade, Crump, Eusepi and Moench (2016), Angeletos, Huo and Sastry (2020), and Andre, Haaland, Roth and Wohlfart (2021). Recently, Andre, Pizzinelli, Roth, and Wohlfart (2021) studied the subjective models of the macroeconomy of households and experts using a hypothetical vignette approach. They found significant heterogeneity in subjective models within and between households and experts on the effects of several macroeconomic shocks on unemployment and inflation. This heterogeneity is partially explained by agents' selective retrieval of different propagation mechanisms. They find experts tend to recall textbook models, while households with different personal experiences recall different propagation channels of the shocks.

Following the literature which suggests agents' different subjective models as a source of disagreement, this paper empirically studies the heterogeneous expectations of professional forecasters or experts about macroeconomic variables (aggregate output, consumption and investment). Using forecasts data from the Survey of Professional Forecasters (SPF) managed by the Federal Reserve Bank of Philadelphia, we formally test whether professional forecasters utilize textbook stochastic growth models which exhibit long-run equilibrium (or cointegration) relationships to forecast macroeconomic variables.[1] Stochastic growth models (e.g., King, Plosser and Rebelo, 1988; King, Plosser, Stock and Watson, 1991) have been widely used for analyzing economic growth, business cycles, and welfare in recent decades. A salient feature of these models (including their one-sector and multi-sector variants) is a balanced growth path along which aggregate output, consumption, and investment share a common trend. Our analysis sheds light on a crucial assumption underlying these models and improves understanding on the expectation formation about macroeconomic variables.

We firstly show that the median (or mean) survey forecasts of aggregate output are not cointegrated with median (or mean) forecasts of aggregate consumption and investment.[2] This result is robust to different statistical tests, different forecasting horizons (1-, 2-, 3-, and 4-quarter ahead forecasts), using data from output forecasts that is made at the same or different dates from consumption (or investment) forecasts, with or without imposing theoretical restrictions, and using a multivariate analysis.

Using the data of individual-level forecasts, the paper finds significant heterogeneity among professional forecasters in terms of utilizing cointegration relationships. The majority of forecasters do not appear to use the cointegration relation between aggregate output and consumption in forecasting, as their forecasts of aggregate output are not cointegrated with consumption forecasts.[3] Similar results are found using forecasts of aggregate output and investment. The proportion of forecasters whose aggregate output forecasts are not cointegrated with investment forecasts is generally higher than the proportion of forecasters not utilizing the cointegration relation between output and consumption.

[1] The paper studies the expectation formation of macroeconomic variables by professional forecasters, following a large literature on expert forecasts, such as Bordalo, Gennaioli, Ma, and Shleifer (2020) and Coibion and Gorodnichenko (2015). Some influential papers have modeled experts' forecasts as a crucial determinant of households' forecasts. For instance, Carroll (2003) shows that households' inflation expectations are well captured by a model in which households' views are derived from news reports of the views of professional forecasters. In Malmendier and Nagel (2015), households' inflation forecasts are influenced by knowledge gained from personal experience and from experts' forecasts.

[2] The same results were found for the US Federal Reserve's Greenbook forecasts.

[3] Again, this is robust to using different tests, using forecasts over different horizons, forecasts made at the same or different dates, with or without imposing the theory-implied cointegrating vector.

In the tests, we consider a number of econometric issues. For instance, testing many hypotheses separately and simultaneously may lead to false rejections of the null hypothesis (the multiple testing problem). In addition, potential structural breaks may lead to a non-rejection of the null hypothesis that output forecasts are not cointegrated with forecasts of consumption (or investment). After addessing these issues, we found that heterogeneity still exists in utilizing long-run equilibrium relationships for forecasting macroeconomic variables.

The paper is related to a strand of literature that studies economic agents' beliefs about the relationship between different macroeconomic variables. Carvalho and Nechio (2014), Drager, Lamla, and Pfajfar (2016) and Kuchler and Zafar (2019) study the relationship between households' beliefs about unemployment, inflation, and interest rates. By employing a hypothetical vignette approach, Andre, Pizzinelli, Roth and Wohlfart (2019) study households' beliefs about how macroeconomic shocks impact unemployment and inflation. One of their findings is that experts tend to recall textbook models in forecasting. Connecting with this result, we find that some experts appear to utilize long-run equilibrium relationships present in stochastic growth models to forecast macroeconomic variables.[4]

A large literature utilizes survey expectations to test or discipline the modeling of agents' expectation formation process. Greenwood and Shleifer (2014) and Adam, Marcet, and Beutel (2017) reject the rational expectation (RE) hypothesis with the stock return forecast data. Branch (2004) and Coibion and Gorodnichenko (2015) find that full-information RE is hard to square with the inflation forecast updating mechanisms. The evidence provided in this paper contributes to this literature. On the one hand, realized output and consumption (or investment) are shown to be cointegrated. On the other hand, forecasts of output and forecasts of consumption (or investment) are to a large extent not cointegrated. The discrepancy between the cointegration among realized macroeconomic variables and the non-cointegration among forecasts of macroeconomic variables appears difficult to be explained by the RE hypothesis.[5]

[4]The remaining experts may recall other models which do not incorporate cointegration relationships in forecasting.

[5]Kuang, Tang, Zhang, and Zhang (2022) find that survey stock price expectations are not anchored by forecasts of fundamentals, as opposed to a wide range of asset pricing models with various informational assumptions.

The remainder of the paper is structured as follows. Section 2 outlines the survey evidence and tests cointegration between realized data. Section 3 presents evidence from median and mean survey forecasts. Section 4 reports evidence from individual forecast data and discusses potential mechanisms. Section 5 addresses several econometric and other issues. Section 6 concludes.

2 Survey evidence: roadmap

The (one-sector and two-sector) stochastic growth models typically include a balanced growth path along which aggregate output, consumption, and investment share a common trend; Appendix A reviews some general stochastic growth models.[6] Denote lowercase letters y, c, i the logarithm of aggregate output (Y), consumption (C) and investment (I). Aggregate consumption, investment and output contain a unit root. Thus, under the assumption of RE, forecasts of the three variables over an arbitrary forecasting horizon contain a unit root. Denote by $E_{t_1}x_{t_1+k_1}$ the conditional expectations of x made at date t_1 and over horizon k_1. Moreover, we have the following results which summarize the cointegration relationships among forecasts of the three variables under RE:

Proposition 1 *For arbitrary $t_1, t_2, k_1, k_2 > 0$, (1) forecasts of aggregate output $E_{t_1}y_{t_1+k_1}$ are cointegrated with forecasts of aggregate consumption $E_{t_2}c_{t_2+k_2}$ with the cointegrating vector $(1, -1)$, i.e., $E_{t_1}y_{t_1+k_1} - E_{t_2}c_{t_2+k_2}$ are stationary; (2) forecasts of aggregate output $E_{t_1}y_{t_1+k_1}$ are cointegrated with forecasts of investment $E_{t_2}i_{t_2+k_2}$ with the cointegrating vector $(1, -1)$, i.e., $E_{t_1}y_{t_1+k_1} - E_{t_2}i_{t_2+k_2}$ are stationary.*[7]

Three special cases of Proposition 1 are worth mentioning. First, one special case is that forecasts of aggregate output and consumption (or investment) are made at the same date ($t_1 = t_2$) and over the same forecasting horizon ($k_1 = k_2$). Second, the cointegration relation

[6] Note that in some stochastic growth models with imperfect information (e.g., Beaudry and Portier, 2014), agents also have the knowledge of a balanced growth path and their forecasts of the three macroeconomic variables display corresponding cointegration relationships.

[7] We have omitted the proof of the proposition. Lemma 2 and Theorem 4 in Kuang, Tang, Zhang, and Zhang (2022) provide the proof in a more general setting.

holds for forecasts of different variables made at the same date ($t_1 = t_2$) and over different horizons ($k_1 \neq k_2$). Third, the cointegration relation holds for forecasts of different variables made at different dates ($t_1 \neq t_2$).

Based on the proposition, the paper studies if professional forecasters indeed utilize the cointegration relationships in forecasting, and also examines the extent of heterogeneity among the forecasters. Specifically, Table 1 outlines the main survey evidence in this paper.

Table 1: **Main evidence from survey forecasts: roadmap**

Panel A: Analysis using median/mean **forecasts**	
Integration properties of the forecasts	Evidence 1
Cointegration between Y and C with imposing the vector $(1, -1)$	Evidence 2A
Cointegration between Y and I with imposing the vector $(1, -1)$	Evidence 2B
Cointegration between Y and C without imposing any vector	Evidence 3A
Cointegration between Y and I without imposing any vector	Evidence 3B
Multivariate analysis of a common trend shared by Y, C, and I	Evidence 4
Panel B: Analysis using individual-level **forecasts**	
Cointegration between Y and C over the same horizon with imposing the vector $(1, -1)$	Evidence 5A
Cointegration between Y and I over the same horizon with imposing the vector $(1, -1)$	Evidence 5B
Cointegration between Y and C over different horizons with imposing the vector $(1, -1)$	Evidence 6A
Cointegration between Y and I over different horizons with imposing the vector $(1, -1)$	Evidence 6B
Cointegration between Y and C over the same horizon without imposing any vector	Evidence 7A
Cointegration between Y and I over the same horizon without imposing any vector	Evidence 7B
Panel cointegration testing between Y and C	Evidence 8A
Panel cointegration testing between Y and I	Evidence 8B

Before presenting the evidence from the survey forecast data, we test for cointegration between realized output and consumption (or investment) with the cointegrating vector $(1, -1)$ implied by typical stochastic growth models. Data on aggregate output, consumption, and investment for the US are downloaded from the website of the Federal Reserve Bank of St. Louis. The sample covers the period between 1981:Q3 to 2018:Q4, which is the same as that of the survey forecast data. All data are at a quarterly frequency and seasonally adjusted. Real output is re-constructed by subtracting the real Government Consumption Expenditures from the real GDP.

Table 2 reports test results by applying the Dickey-Fuller GLS (DF-GLS) test to log output to consumption (or investment) ratios. Panel A shows that the DF-GLS test rejects the null

Table 2: **Tests of the cointegration with the vector** $(1, -1)$ **using realized data covering 1981:Q3 – 2018:Q4**

	Test stat.	**5% critical**	**10% critical**
Panel A: *Realized (log) output-consumption ratio*			
DF-GLS	-2.738**	-2.060	-1.749
Panel B: *Realized (log) output-investment ratio*			
DF-GLS	-1.757*	-2.040	-1.730

** and *: the test statistics with double asterisks or one asterisk indicate that the corresponding test rejects the null of unit root at the 5% and 10% significance level, respectively.

of unit root at the 5% level, in favor of cointegration between output and consumption. Panel B indicates that the test fails to reject the null of unit root at the 5% significance level, but rejects the null at the 10% level. Therefore, the realized data is broadly consistent with the model-implied cointegration between output and consumption (or investment), respectively.[8]

3 Evidence from median and mean forecast data

3.1 Data

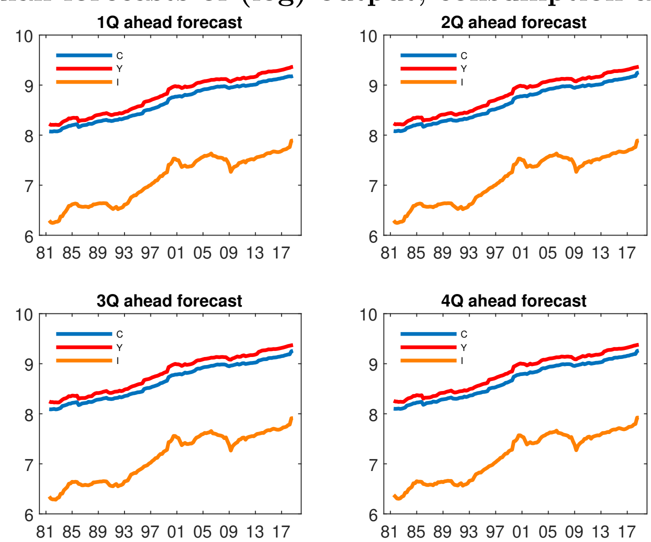
Survey forecast data are obtained from the SPF managed by the Federal Reserve Bank of Philadelphia. Consumption forecasts are measured by forecasts of chain-weighted real personal consumption expenditures (variable RCONSUM in the SPF documentation).[9] Forecasts of real private output are measured by forecasts for the annual level of chain-weighted real GDP (RGDP), subtracting forecasts of the annual level of chain-weighted real federal government consumption and gross investment (RFEDGOV) and local government consumption and gross

[8]In addition, we conducted both Johansen trace and maximum-eigenvalue tests among realized output, consumption, and investment. Test results are in favour of one cointegrating equation at the 5% significance level.

[9]The SPF asks participants to predict real personal consumption expenditures (PCE), which include household expenditures on services and non-durable and durable goods. We cannot separate durable forecasts from non-durable goods and service forecasts. However, in macroeconomic models with durable and non-durable goods, such as Beaudry and Portier (2004) and Monacelli (2009), agents have knowledge of the balanced growth path and their forecasts of total consumption (including durable goods) and output are still cointegrated.

investment (RSLGOV).[10] Forecasts of aggregate investment are measured as the sum of the forecasts of chain-weighted real non-residential fixed investment (RNRESIN) and residential fixed investment (RRESINV). All variables are available at a quarterly frequency from the third quarter (Q3) of 1981 onwards. Four forecasting horizons are available: 1-, 2-, 3-, and 4-quarter ahead.[11] Notably, SPF forecasts of the level of the three variables are provided with time-varying base years. Appendix B explains the rebasing of the forecast data.[12] Figure 1 plots the (normalized) rebased median forecasts of (log) output, consumption, and investment for all available forecasting horizons. Appendix C shows that the Greenbook forecasts from the Federal Reserve display similar results to those of the SPF and that there exists no cointegration between forecasts of output and forecasts of consumption (or investment).

Figure 1: **Median forecasts of (log) output, consumption and investment**



3.2 Integration properties of the forecasts

This section examines the integration properties of the median and mean SPF forecasts of aggregate consumption, output, and investment. Table 3 reports p-values from the Phillips-

[10]Before 1992, forecasts of real GNP are used.

[11]Whelan (1993) discusses that consumption-output and investment-output ratios should be constructed using nominal data, given chain-weighted NIPAs. However, since forecasters in SPF are only asked to predict real variables and no nominal data are available, it is infeasible to construct ratios using nominal data. Therefore, in line with literature, we construct the forecast ratios using real data.

[12]Although forecasters do not directly predict (log) consumption-output and (log) investment-output ratios and the logarithm of an expectation value does not necessarily equal the expectation value of a logarithm (the Jensen's inequality), they are regarded equal when the first-order approximation is implemented.

Perron (PP) and Dickey–Fuller (DF) tests of median SPF forecasts of consumption, output, and investment for all four forecasting horizons.[13]

Table 3: **Integration properties of median SPF forecasts**

P values	1Q ahead	2Q ahead	3Q ahead	4Q ahead
Panel A: I(1) test				
Median consumption forecasts				
PP (Z_t test)	0.9054	0.9041	0.9019	0.9032
Dickey–Fuller	0.9440	0.9425	0.9394	0.9388
Median output forecasts				
PP (Z_t test)	0.7738	0.7873	0.7874	0.7907
Dickey–Fuller	0.8942	0.9007	0.8968	0.8942
Median investment forecasts				
PP (Z_t test)	0.7174	0.7174	0.7091	0.7106
Dickey–Fuller	0.8877	0.8877	0.8835	0.8790
Panel B: I(2) test				
Median consumption forecasts				
PP (Z_t test)	0.000	0.000	0.000	0.000
Dickey–Fuller	0.000	0.000	0.000	0.000
Median output forecasts				
PP (Z_t test)	0.000	0.000	0.000	0.000
Dickey–Fuller	0.000	0.000	0.000	0.000
Median 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 1: 1-, 2-, 3- and 4-quarter ahead median forecasts of aggregate consumption, output, and investment are I(1) but not I(2).

Panel A shows that for all forecasting horizons considered, both tests fail to reject the null that median forecasts of consumption, output, and investment are integrated of order 1, i.e. I(1), at conventional significance levels. Panel B indicates that all median forecasts over different forecasting horizons are not integrated of order 2, i.e. I(2). Tests on mean forecasts

[13] We use PP tests to examine integration properties of forecasts with Newey-West optimal lags imposed and DF tests without any lag included. The critical values of PP and DF tests are computed by Cheung and Lai (1995b).

reach the same conclusion; see Appendix D.

3.3 No cointegration with imposing theoretical restrictions

This section tests the cointegration relation between output forecasts and consumption (or investment) forecasts with the cointegrating vector $(1, -1)$ implied by stochastic growth models. Many papers criticized the power of the standard DF class of unit root tests in the 1980s and 1990s. This paper uses some of the most powerful tests, such as the DF-GLS test. Another way to mitigate this concern is by applying the KPSS test, which tests the null hypothesis of a stationary process against the alternative of a unit root. The robustness of the results is also demonstrated by analyzing the effects of sample size on testing outcomes in Section 3.5 and 4.2.1.

Table 4 reports the results of testing cointegration between median forecasts of output and consumption (or investment) with the cointegrating vector $(1, -1)$ using the PP, DF-GLS, and KPSS tests.[14] The first four columns display test statistics for forecasts made at the same date and over the same forecasting horizon. The last column, labeled as "4Q Y & 1Q C" in Panel A (or "4Q Y & 1Q I" in Panel B), tests cointegration between 4-quarter ahead output forecasts and 1-quarter ahead consumption (or investment) forecasts.

[14] We report test results assuming that there exists no trend component in the consumption-output ratios or investment-output ratios. The cointegration test results are robust when the trend component is added. The critical values of the DF-GLS test have been computed by Cheung and Lai (1995a).

Table 4: **Cointegration tests for median SPF forecasts with the cointegrating vector** $(1, -1)$

Panel A: *no cointegration between forecasts of output and consumption*					
Median	1Q ahead	2Q ahead	3Q ahead	4Q ahead	4Q Y & 1Q C
PP (Z_t test)	-1.586	-1.672	-1.713	-1.720	-1.779
10% critical value	-2.577	-2.577	-2.577	-2.577	-2.577
DF-GLS	-1.378	-1.478	-1.526	-1.508	-1.588
10% critical value	-1.749	-1.749	-1.749	-1.749	-1.749
KPSS	1.885	1.862	1.847	1.861	1.685
5% critical value	0.463	0.463	0.463	0.463	0.463
Panel B: *no cointegration between forecasts of output and investment*					
Median	1Q ahead	2Q ahead	3Q ahead	4Q ahead	4Q Y & 1Q I
PP (Z_t test)	-1.705	-1.694	-1.559	-1.479	-1.719
10% critical value	-2.577	-2.577	-2.577	-2.577	-2.577
DF-GLS	-0.209	-0.182	-0.193	-0.242	-0.155
10% critical value	-1.737	-1.737	-1.743	-1.743	-1.737
KPSS	2.630	2.641	2.773	2.843	2.645
5% critical value	0.463	0.463	0.463	0.463	0.463

Evidence 2A (Panel A): Median aggregate consumption forecasts and output forecasts are not cointegrated with the vector $(1, -1)$;
Evidence 2B (Panel B): Median aggregate investment forecasts and output forecasts are not cointegrated with the vector $(1, -1)$.

In Panel A, the PP and DF-GLS tests fail to reject the null hypothesis that median forecasts of output are not cointegrated with median forecasts of consumption over the same or different forecasting horizons with the vector $(1, -1)$ at the 10% significance level. Meanwhile, the KPSS tests reject the null hypothesis that forecasts of output are cointegrated with forecasts of consumption for all forecasting horizons at the 5% significance level. Thus, all tests yield the same conclusion, which is that forecasts of output are not cointegrated with forecasts of consumption with the vector $(1, -1)$ over the same or different horizons.

Panel B of Table 4 reports the results of testing cointegration between median forecasts of output and investment with the vector $(1, -1)$. All tests produce the same conclusion that forecasts of output are not cointegrated with median forecasts of investment with the vector $(1, -1)$ over the same or different horizons. Similar results are found using mean forecasts; see Appendix E.1.

We tested overidentifying restrictions on the cointegration relation among forecasts using the likelihood-ratio test. We imposed 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. The optimal number of lags is selected by the Akaike information criterion (AIC). Table 5 reports the p values of tests for median forecasts and mean forecasts, respectively. The null hypothesis is rejected at the 5% significance level, implying that the theoretical cointegration relation among the forecasts is not supported by the data.

Table 5: **Likelihood ratio test of over-identifying restrictions**

	1Q ahead	2Q ahead	3Q ahead	4Q ahead
P value (median)	0.00	0.00	0.00	0.00
P value (mean)	0.00	0.00	0.00	0.00

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

3.4 No cointegration without imposing theoretical restrictions

This section tests cointegration between forecasts of macroeconomic variables without imposing the theory-implied cointegrating restriction. Using the Engle–Granger test, Panel A (or B) of Table 6 reports the results of testing the cointegration between output forecasts and consumption (or investment) forecasts over the same or different horizons and the corresponding 10% critical values.[15] Again, the last column is the testing results of using 4-quarter-ahead output forecasts and 1-quarter-ahead consumption (or investment) forecasts.

[15] We report the Engle–Granger test results without incorporating the trend component. Our test results are robust when the trend is included. The Engel-Granger test's critical values have been calculated by MacKinnon (1996).

Table 6: **Cointegration test for median SPF forecasts without imposing the 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.	-2.315	-2.378	-2.424	-2.451	-2.475
10% critical value	-3.073	-3.073	-3.073	-3.073	-3.073
Panel B: *cointegration between forecasts of output and investment*					
Median	1Q ahead	2Q ahead	3Q ahead	4Q ahead	4Q Y &1Q I
Test stats.	-2.257	-2.270	-2.251	-2.250	-2.306
10% critical value	-3.073	-3.073	-3.073	-3.073	-3.073

Evidence 3A (Panel A): Median consumption forecasts are not cointegrated with output forecasts without imposing any cointegrating vector;
Evidence 3B (Panel B): Median investment forecasts are not cointegrated with output forecasts without imposing any cointegrating vector.

The test results uniformly suggest that we cannot reject the null hypothesis that output forecasts are not cointegrated with consumption (or investment) forecasts at the 10% significance level and over the same or different forecasting horizons, as all test statistics are greater than the corresponding 10% critical values. The test results are robust to using mean forecasts; see Appendix E.2.

3.5 Multivariate analysis: testing for common trends among forecasts of output, consumption, and investment

This section tests if forecasts of output, consumption, and investment share a common trend. We tested the number of cointegrating vectors among median forecasts of output, consumption, and investment over a 1- to 4-quarter ahead forecasting horizon, using the Johansen trace and maximum-eigenvalue tests. The Johansen trace test examines the multivariate cointegrating relation with the null of no more than r cointegrating vector(s) (rank of r), against the alternative hypothesis that the number of the cointegrating vector(s) is strictly greater than r. Meanwhile, the Johansen maximum-eigenvalue test evaluates the null of exact r cointegrating vector(s) against the alternative of $r+1$ cointegrating vectors.

Table 7: **Johansen trace and maximum-eigenvalue tests for the number of common trends among median forecasts**

Johansen test				
Trace test: $J^{trace}(r)$, r = rank				
null:	no more than r cointegration vector			
Alternative:	number of the cointegrating vector > r			
Median	r=0	5% critical	r=1	5% critical
1Q ahead	28.5*	29.7	9.6	15.4
2Q ahead	27.6*	29.7	9.7	15.4
3Q ahead	26.2*	29.7	9.2	15.4
4Q ahead	25.2*	29.7	7.6	15.4
Maximum-eigenvalue test: $max(r)$, r = rank				
null:	no more than r cointegration vector			
Alternative:	number of the cointegrating vector = r+1			
Median	r=0	5% critical	r=1	5% critical
1Q ahead	18.9*	21.0	9.5	14.1
2Q ahead	17.9*	21.0	9.7	14.1
3Q ahead	17.2*	21.0	9.1	14.1
4Q ahead	17.5*	21.0	8.4	14.1

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

Evidence 4: Median forecasts of output, consumption, and investment do not share a common trend.

Table 7 reports test results using the median forecasts of output, consumption, and investment.[16] The trace test statistics reported in the upper part of Table 7 suggest that tests for the median forecasts of output, consumption, and investment over the period of 1- to 4- quarter ahead uniformly fail to reject the null of no cointegrating vector against the alternative of the existence of cointegration. Moreover, the maximum-eigenvalue test statistics reported in the lower part of Table 7 also uniformly fail to reject the null of no cointegrating vector against the

[16]For the Johansen trace and maximum-eigenvalue tests, we report the outcomes with 1 plus the number of lags selected by the AIC (robust to other selection criteria, like HQIC and SBIC). This is because the Johansen test uses first difference and, to compensate for the one lag loss, we incorporated one additional lag.

alternative of exactly one cointegrating vector. Thus, the results show that any linear combination of forecasts of output, consumption and investment is not stationary. Similar results are found using the mean forecasts; see Appendix E.3.

A concern is that tests of unit roots or of cointegration may have low power. We have used several most powerful and state-of-the-art tests. Two additional ways are used to address this concern, by analyzing the effect of sample size on testing outcomes. We report the recursive Johansen trace test statistics from testing the common trends shared by forecasts of output, consumption and investment. The initial subsample of median or mean forecasts covers 1981:Q3 to 1986:Q2 (20 quarters). By adding a new observation, the sample gradually expands to the full sample (from 1981:Q3 to 2018:Q4). Figure A.15 in Appendix M plots the test statistics (red lines) and the corresponding 5% critical values (blue lines) of recursive Johansen trace tests 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 the existence of at least one cointegraing vector.[17] In Section 4.2.1 where individual-level forecasts are used, individual forecasters' are split into different groups with different sample sizes, and we find that there is no monotonic relation between the proportions of rejecting the null hypothesis and the sample size.

4 Evidence from individual forecasts

4.1 Data

Individual forecasts from the SPF are used for testing. To obtain reliable testing results, we selected individual forecasters who have answered more than 60 quarters between 1981:Q3 and 2018:Q4; this resulted in 21 forecasters.[18] Individual-level forecasts of real output, consumption,

[17]Similar results are obtained with mean forecasts; see Figure A.16 in Appendix M. Moreover, this appendix shows the recursive trace statistics when testing the null hypothesis of 1 cointegrating vector against more than 1 cointegrating vector.

[18]We found a similar result signifying the existence of significant heterogeneity in utilizing long-run macroeconomic relationships in forecasting when we used a different cutoff value than 60 quarters, such as 70 or 58.

and investment are measured and rebased in the same way as the median forecasts in Section 3.1.

Figure 2: **Number of SPF forecasts reported by each selected forecaster for 1- to 4-quarter ahead forecasting horizons**

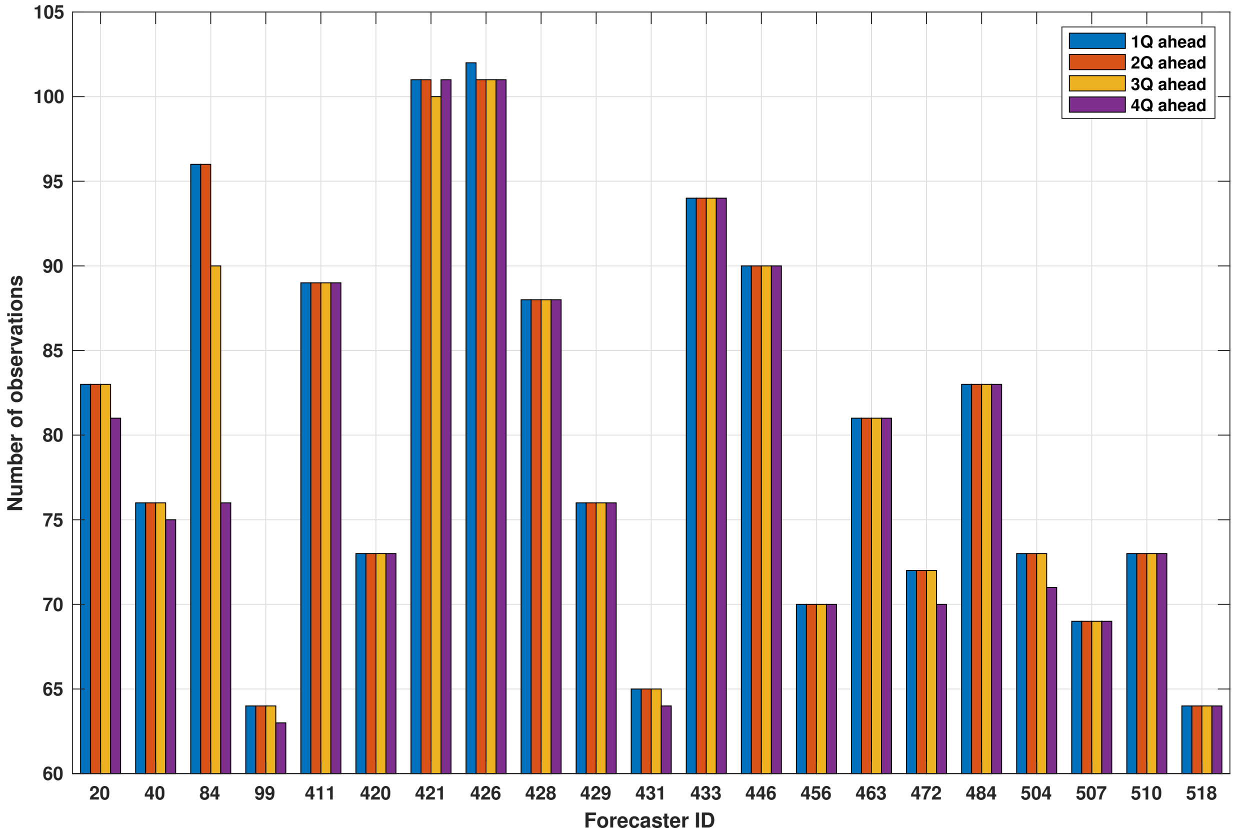


Figure 2 plots the number of valid forecasts of output-consumption ratio (which are also numbers of valid output-investment ratio) for each of the 21 selected individual forecasters, over 1- to 4-quarter ahead. The X-axis denotes each forecaster's ID in the SPF dataset, while the Y-axis denotes the number of forecast observations over different forecasting horizons. In total, we have 84 time series of forecasts for each individual variable (output, consumption, or investment). There is a small number of missing values in the dataset, and the procedure for addressing this is detailed in Appendix F.

4.2 Heterogeneity in utilizing cointegration relationships in forecasting: testing using forecasts made over the same horizon

This section tests cointegration between the individuals' forecasts of output and consumption (or between forecasts of output and investment) over the *same* forecasting horizon, with the theory-implied $(1, -1)$ restriction imposed. Table 8 reports the results that emerged from the PP, DF-GLS, and KPSS tests.

Table 8: **Cointegration tests using individual forecasts over same forecasting horizon**

No. of forecasters:	**21** forecasters and **4** horizons (1 – 4Q-ahead)	
With $(1, -1)$ restriction	**No. of no cointegration detected out of 84 forecasts**	**Proportion of no cointegration detected**
Panel A: *cointegration between forecasts of output and consumption*		
PP Z_t test (10% crit. value)	63	75.0%
DF-GLS (10% crit. value)	66	78.6%
KPSS (5% crit. value)	56	66.7%
Panel B: *cointegration between forecasts of output and investment*		
PP Z_t test (10% crit. value)	77	91.7%
DF-GLS (10% crit. value)	76	90.5%
KPSS (5% crit. value)	62	73.8%

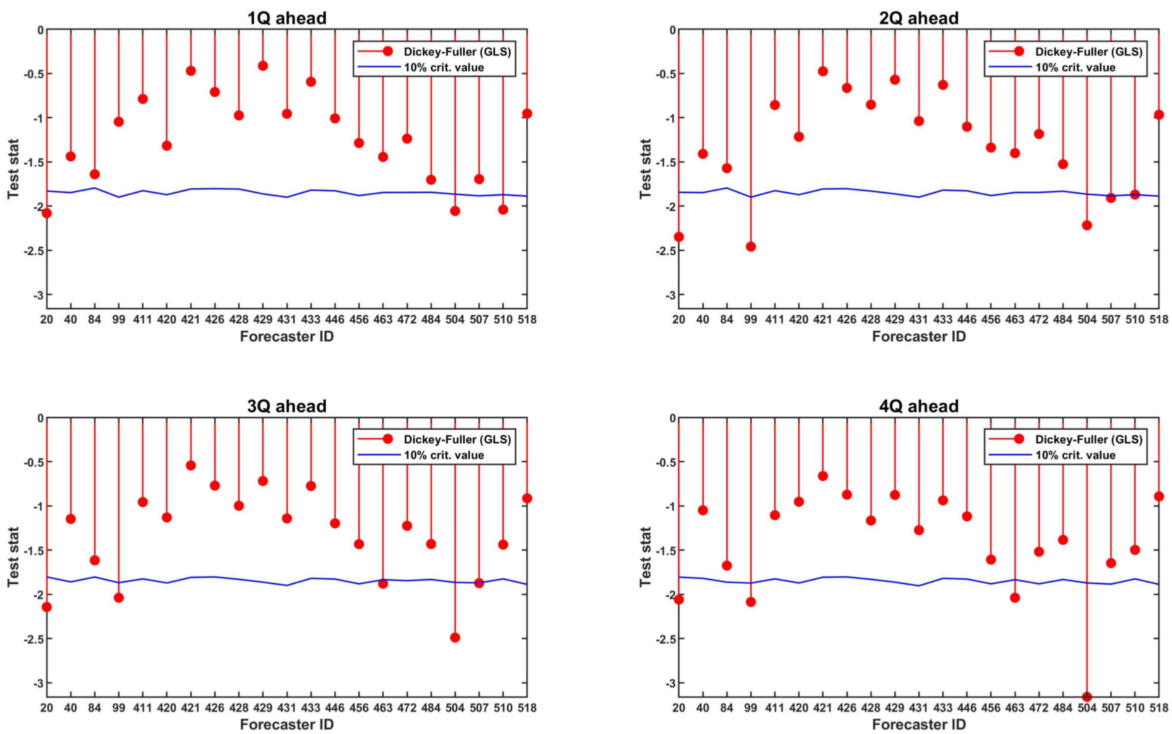
Evidence 5A (Panel A): In 67% – 79% of the 84 cases (21 forecasters and 4 forecasting horizons), forecasts of output and consumption are not cointegrated with vector $(1, -1)$;
Evidence 5B (Panel B): In 74% – 92% of the 84 cases (depending on the test), forecasts of output and investment are not cointegrated with vector $(1, -1)$.

The testing results suggest a significant disagreement among individual forecasters in terms of utilizing the cointegration relation between output and consumption (or investment) in forecasting. Panel A shows that, using the PP test, in 75% of the 84 cases, output forecasts are not cointegrated with consumption forecasts, while they are cointegrated in the remaining 25% of cases.[19] The DF-GLS test suggests that for about 21% of the 84 cases, forecasts of output and consumption are cointegrated; the corresponding number is about 1/3 based on the KPSS test.

[19] Recall that there are 84 time series of forecasts for each variable (21 forecasters times four forecasting horizons).

Figure 3 plots the DF-GLS test statistics and the associated critical values from Panel A of Table 8. The test statistics are pinned down by the circle at the end of each red stem, while the corresponding critical value is located on the blue line.[20] For the majority of forecasters, the test statistics are above the blue line of critical values, suggesting no cointegration between output forecasts and consumption forecasts. The DF-GLS test indicates that three forecasters (with IDs of 20, 99, and 504) form cointegrated output and consumption forecasts for at least 3 forecasting horizons. In addition, the forecaster with ID 463 (or ID 507) forms cointegrated output and consumption forecasts over 3- and 4-quarter ahead (or over 2- and 3-quarter ahead).

Figure 3: **DF-GLS test statistics vs. critical values using individual-level output and consumption forecasts**



Panel B of Table 8 tests the cointegration between output forecasts and investment forecasts over the same forecasting horizon and with the $(1, -1)$ restriction. In 92% of the 84 cases, the PP tests suggest that output forecasts are not cointegrated with investment forecasts at the 10% significance level, while the two forecasts are cointegrated for the remaining 8% of the cases. The DF-GLS test fails to reject the null of no cointegration between forecasts of output and

[20] For the DF-GLS test, since we report the test statistics and critical values with the number of lags that minimizes the modified AIC criterion, the critical value for different tests differs. See Ng and Perron (2000).

investment for 91% of the cases, while the KPSS test rejects the null of cointegration between forecasts of output and investment for around 74% of the cases (and fails to reject the null for 26% of the cases).[21]

Overall, the majority of forecasters do not utilize the cointegration relation between output and consumption (or investment) with the cointegrating vector $(1, -1)$ in forecasting. Moreover, for all three tests, the proportion of forecasters who do not utilize the cointegration relation between output and investment in forecasting is higher than the proportion of forecasters who do not utilize the cointegration relation between output and consumption.[22]

Using individual-level forecasts, we tested overidentifying restrictions by imposing two cointegration relations implied by stochastic growth models when estimating a VECM. The two cointegration vectors are $(1, -1, 0)$ and $(1, 0, -1)$ for forecasts of output, consumption, and investment. The optimal number of lags was selected by the AIC criterion. Table A.13 reports outcomes of the likelihood-ratio test. The test rejected the cointegrating restrictions at the 5% significance level for the majority of cases, while the null hypothesis was not rejected for a handful of cases. Thus, there still exists heterogeneity in the utilization of the long-run relationships in forecasting macroeconomic variables.

We also recursively applied Johansen trace test to examine if a common trend exists in forecasts of consumption, output, and investment for several forecasters IDs. Figure A.17 in Appendix M displays the test statistics against the corresponding 5% critical values. The null hypothesis of a common trend was persistently rejected for some IDs (ID 20 and 510) as the sample became longer, and was not rejected for some other IDs (ID 421 and 429).

4.2.1 Test results by groups with different sample size and characteristics

We examined how the proportion of rejecting the null hypothesis of no cointegration depends on the sample size. Forecasters were divided into three groups according to their sample sizes (top, middle, and bottom one-third). Table **A.20** in Appendix N reports the proportions of no

[21] Figure A.4 plots the DF-GLS test statistics against the corresponding critical values associated with Panel B of Table 8.

[22] PP and KPSS test statistics are illustrated by plots in Appendix H.

cointegration between output forecasts and forecasts of consumption (or investment) in each group. There is no monotonic relation between the proportions and the sample size.

Moreover, we examined how the testing results depend on the forecasters' characteristics. The SPF provides limited information regarding the participants' characteristics. It contains information on the types of institutions to which the forecasters belong, which are either financial service providers or non-financial service providers. One may wonder if the test results are robust across different groups or if there exists heterogeneity across the two groups.

Table A.21 reports the proportion of forecasters within each group whose forecasts of consumption (or investment) are not cointegrated with their own output forecasts with the cointegration vector (1, -1). Within each group (financial or non-financial service providers), heterogeneity still exists in utilizing the cointegration relation between output and consumption (or investment) in forecasting. Across the two groups, it appears that a higher proportion of forecasters from financial service institutions make use of the cointegration relations between macroeconomic variables in their forecasts.[23]

4.3 Heterogeneity in utilizing cointegration relationships in forecasting: testing using forecasts made over different horizons

Using individual forecasts data, this section tests the cointegration between forecasts of output and consumption (or investment) over *different* forecasting horizons and with imposing the theory-implied $(1, -1)$ cointegrating vector. Panel A (or B) of Table 9 reports the test results produced by the PP, DF-GLS and KPSS tests, using 4-quarter ahead output forecasts and the 1-quarter ahead consumption (or investment) forecasts of 21 forecasters.

Panel A reports that there is no cointegration between 4Q-ahead output forecasts and 1Q-ahead consumption forecasts in 13 out of 21 cases based on the PP test in 17 out of 21 cases based on the DF-GLS test, and in 16 out of 21 cases based on the KPSS test. Thus, all three tests suggest that in over 62% of the cases, 4-quarter ahead forecasts of output and 1-quarter

[23] A potential reason for this is that financial institutions may have more demanding requirements on practitioners' economic and quantitative skills and therefore the experts employed in this type of institutions are more likely to recall stochastic growth models in forecasting.

Table 9: **Testing using individual forecasts over different forecasting horizons**

With $(1, -1)$ restriction	**No. of no cointegration detected out of 21**	**Proportion of no cointegration detected**
Panel A: *cointegration between 4Q ahead Y forecasts and 1Q ahead C forecasts*		
PP Z_t test (10% crit. value)	13	61.9%
DF-GLS (10% crit. value)	17	81.0%
KPSS (5% crit. value)	15	71.4%
Panel B: *cointegration between 4Q ahead Y forecasts and 1Q ahead I forecasts*		
PP Z_t test (10% crit. value)	20	95.2%
DF-GLS (10% crit. value)	19	90.5%
KPSS (5% crit. value)	16	76.2%

Evidence 6A (Panel A): In 62% – 81% of 21 cases (depending on the test used), 4-quarter ahead output forecasts and 1-quarter ahead consumption forecasts are not cointegrated, when the $(1, -1)$ cointegrating vector is imposed.
Evidence 6B (Panel B): In 77% – 95% of 21 cases, 4-quarter ahead output forecasts and 1-quarter ahead investment forecasts are not cointegrated, when the $(1, -1)$ cointegrating vector is imposed.

ahead forecasts of consumption are not cointegrated. Panel B shows that, compared with the results in Table 8, a larger proportion of no cointegration is detected between 4-quarter ahead output forecasts and 1-quarter ahead investment forecasts by all three tests. Overall, the testing results suggest that heterogeneity exists in terms of utilizing cointegration relations in making forecasts among individual forecasters and that the majority of forecasters do not make use of the long-run relationships.

4.4 Individual level cointegration tests without imposing theoretical restrictions

Table 10: **Cointegration tests without imposing** $(1, -1)$ **restriction, using individual-level forecasts over the same forecasting horizon**

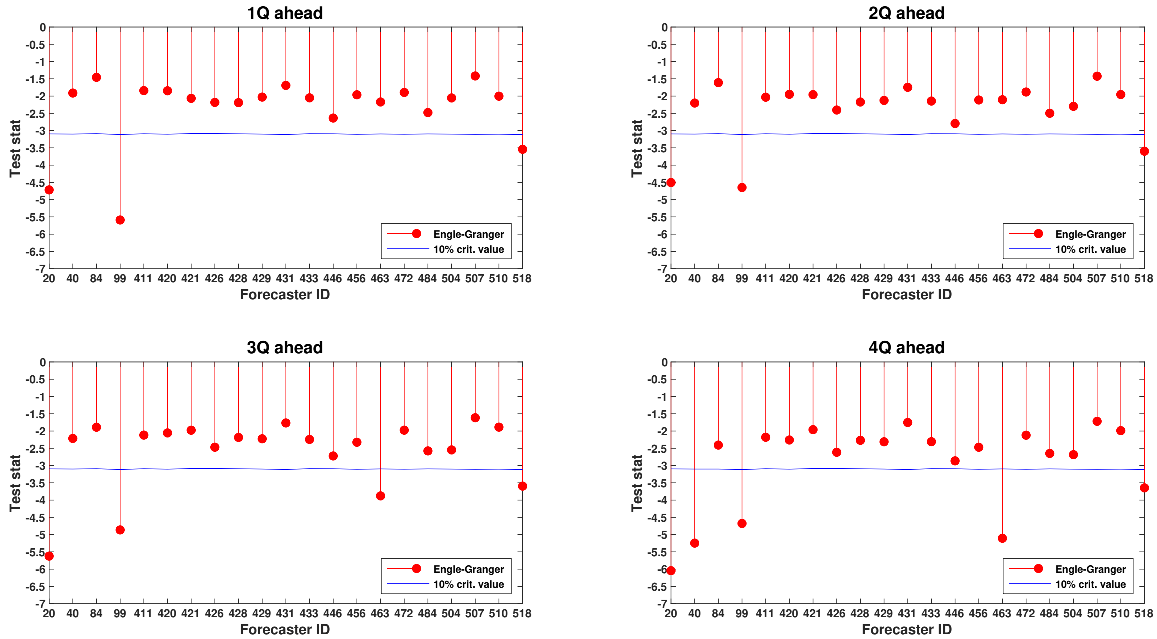
Total individual forecasters:	**21**, with **4** forecasts each (1-, 2-, 3- & 4Q ahead)	
Engle–Granger test (10% crit. value) **over the same horizons**	**No.** of no cointegration detected out of 84 forecasts	**Proportion** of no cointegration detected
Forecasts of Y and C	69	82.1%
Forecasts of Y and I	82	97.6%

Evidence 7A (Panel A): In 82% of 84 cases, the forecasts of Y and C are not cointegrated, when no cointegrating vector is imposed.
Evidence 7B (Panel B): In 98% of 84 cases, the forecasts of Y and I are not cointegrated, when no cointegrating vector is imposed.

Table 10 provides results of the Engle–Granger test of cointegration between forecasts of output and consumption (or investment), without imposing the theoretical cointegration restrictions. In 61 out of 84 (or 82 out of 84) cases, the test suggests no cointegration between individual output forecasts and consumption (or investment) forecasts at the 10% significance level.

Figure 4 presents the test statistics and critical values, when the test is applied to output and consumption forecasts. For the majority of forecasters, forecasts of output and consumption are not cointegrated. They are cointegrated for three forecasters (those labelled as ID 20, 99, and 518) over all four horizons. Moreover, the forecasts of output and consumption made by the forecaster with ID 463 are cointegrated over 3- and 4-quarter ahead, respectively. The 4-quarter ahead forecast of output and consumption made by the forecaster designated as ID 40 are cointegrated. Figure A.9 in Appendix I displays the test statistics and critical values, when the test is applied to output and investment forecasts.

Figure 4: **Engle–Granger test statistics vs. critical values for testing cointegration between output and consumption forecasts and without imposing the $(1, -1)$ restriction**



4.5 Discussion on potential mechanisms

The empirical analysis reveals heterogeneity in utilizing the cointegration relationships present in stochastic growth models in the forecasting of macroeconomic variables. Some experts (or professional forecasters) make use of these relationships, while the majority of the experts do not. Why do the forecasts made by most forecasters not make use of cointegration relationships?

A special survey, SPF Panelists' Forecasting Methods, conducted by the Federal Reserve Bank of Philadelphia in 2009, explicitly asked participants for their forecasting methods.[24] Most respondents – 21 out of 25 – responded that they adopt models to produce their forecasts. Among them, the vast majority (20 out of 21) add subjective adjustments (i.e., based on their judgment) on top of the forecasts produced by the models. This result may not be surprising, as add-factoring is a common practice in the forecasting community, where forecasters incoporate their personal judgements on top of the statistical models. Moreover, the vast majority (20 out of 21) of forecasters acknowledge that a model is a prominent part in producing forecasts

[24]The exact questions and answers can be seen from p.10 of a summary of survey available at https://www.philadelphiafed.org/-/media/frbp/assets/surveys-and-data/survey-of-professional-forecasters/spf-special-survey-on-forecast-methods.pdf?la=en&hash=DA9492A3DE5E3BF70D40F807B7278C83.

over up to two years ahead. The forecasting horizons (1-, 2-, 3- and 4-quarter ahead) examined in our empirical evidence fall within the 2–year ahead horizon, indicating that most of these forecasts are produced by a method which consists of a forecasting model alongside the inclusion of forecasters' judgements.

There are two potential causes for the documented evidence that most forecasters' forecasts of consumption (or investment) are not cointegrated with forecasts of output. First, the component of subjective judgments that these SPF forecasters add to their forecasts of macroeconomic variables do not display a cointegration relation. Various studies have examined the implications of employing subjective judgements in macroeconomics and policy making. For instance, Svensson and Tetlow (2005) empirically demonstrate the impact of judgment on forecasting by the Federal Reserve Board.[25] The second possible cause of no cointegration between these macroeconomic forecasts is that, the forecasting models (the components before adding adjustments) of most SPF forecasters do not impose long-run equilibrium relationships.

5 Econometric issues and discussions

5.1 Addressing the multiple testing problem

In the previous section, we studied the cointegration relation between individual-level forecasts of output and consumption (or investment) by performing multiple unit root (or cointegration) tests simultaneously. A concern is that the testing outcomes might be subject to the multiple testing problem. This issue was addressed in two ways. First, we used Anderson's sharpened False Discovery Rate (FDR) q-values, which is a corrected version of p-values and has greater power than many other methods. Second, we performed a panel cointegration test that considers cross-sectional dependence, which utilizes a larger sample size and has higher power.

Appendix J reports test outcomes using Anderson's sharpened q-values for both forecasts of output-consumption ratios and forecasts of output-investment ratios. Compared with the

[25] Svensson (2003, 2005) formally illustrates that economic performance can be improved if monetary policymakers explicitly incorporate subjective adjustments to the forecasts of key variables. Bullard, Evans, and Honkapohja (2007) show the inclusion of the add-factoring judgmental adjustment in forecasting can lead to self-fulfilling fluctuations.

results in Table 8, the proportions of rejecting the null hypothesis of no cointegration become lower. However, the testing results suggest that heterogeneity still exists in terms of utilizing cointegration relations in making forecasts among individual forecasters, and that the majority of forecasters do not make use of long-run relationships.

Since these professional forecasters are exposed to common aggregate shocks, their forecasts may be highly correlated. Appendix J.1 tests and confirms the cross-sectional dependence for forecasts of output-to-consumption ratios (or output-to-investment ratios) for 1- to 4-quarter forecasting horizons. First-generation panel unit root tests such as the Fisher-type panel tests generally ignore such instances of dependence and suffer from size distortions. To address this issue, we adopted a version of second-generation panel unit root tests, the cross-sectionally augmented Dickey–Fuller (CADF) test proposed by Pesaran (2003). It tests the unit root in heterogeneous panels with the null hypothesis that all panels are non-stationary, against the alternative hypothesis that at least one panel is stationary. The CADF test eliminates cross-sectional dependence by augmenting the standard Dickey–Fuller (DF) or the Augmented Dickey–Fuller (ADF) regression with cross-sectional average lagged levels and the first differences of the individual data series.[26] We examine if 1- to 4-quarter ahead forecasts of output are cointegrated with forecasts of consumption (or investment) with the $(1, -1)$ restriction.

Panel A and B of Table 11 report the p-values of the panel unit root tests on forecasts of output-consumption ratios and output-investment ratios over 1- to 4-quarter ahead forecasting horizons, respectively. In Panel A, the CADF panel unit root tests uniformly reject the null that all panels are non-stationary at any conventional significance level, indicating that at least one forecaster imposes the $(1, -1)$ cointegration relation when forecasting output and consumption. Similar results are obtained in Panel B where the CADF tests uniformly reject the null, suggesting that at least one forecaster imposes the cointegrated relation between output and investment in forecasting.

[26] All panel unit root tests are performed with Newey-West optimal lags. Test results are robust to different numbers of lags incorporated.

Table 11: **P-values of Pesaran panel cointegration test with the cointegrating vector** $(1, -1)$

CADF panel unit root test	**Number of panels**: 21 in each test			
H_0: All panels are non-stationary; H_1: At least one panel is stationary				
	1Q ahead	2Q ahead	3Q ahead	4Q ahead
Panel A: *cointegration between forecasts of output and consumption*				
P-values	0.000	0.000	0.000	0.000
Panel B: *cointegration between forecasts of output and investment*				
P-values	0.001	0.001	0.001	0.007

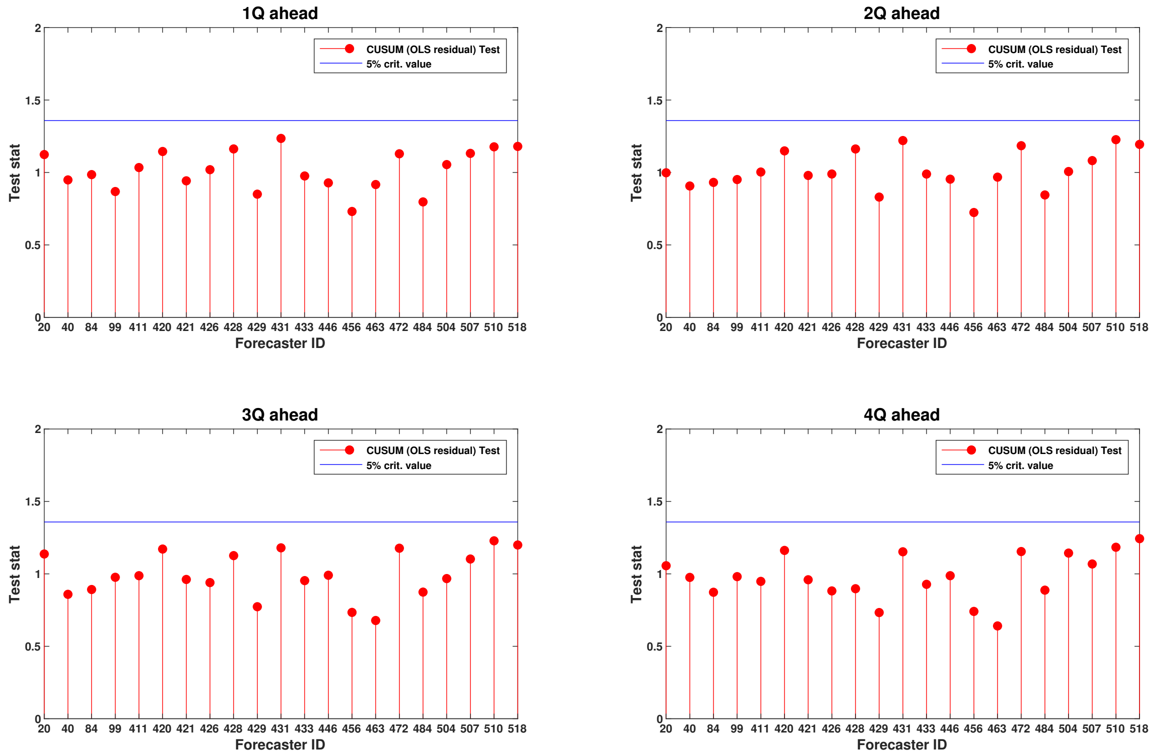
Evidence 8A (Panel A): The CADF panel cointegration tests suggest that at least one forecaster's output forecasts are cointegrated with consumption forecasts over 1-, 2-, 3- and 4-quarter ahead with the $(1, -1)$ restriction.
Evidence 8B (Panel B): The CADF panel cointegration tests suggest that at least one forecaster's output forecasts are cointegrated with investment forecasts over 1-, 2-, 3- and 4-quarter ahead with the $(1, -1)$ restriction.

5.2 Structural breaks

The results so far have not considered potential structural breaks in the sample. A concern is that a structural break may lead to the non-rejection of no cointegration between the forecasts. To address this concern, this section employs the Recursive Cusum test (Krämer and Ploberger, 1992; Brown, Durbin, and Evans, 1975) and the Gregory and Hansen (1996) cointegration test.

The Recursive Cusum test investigates parameter stability with the null hypothesis being no structural break. The test statistics are based on whether the time series abruptly changes in a way not predicted by the model across rolling samples. Figure 5 displays the individual-level Recursive Cusum test statistics and the corresponding 5% critical value, assuming OLS residuals for forecasts of output-consumption ratios. The test statistics (red dots) are all below the corresponding critical values (blue lines). This implies that for all individual forecasts, the Recursive Cusum tests indicate that no structural break is found in the estimated coefficients from the augmented DF regression. Similar results are obtained with recursive residuals. Appendix L shows that the same results are obtained for individual-level forecasts of output-investment ratios and for median and mean forecasts.

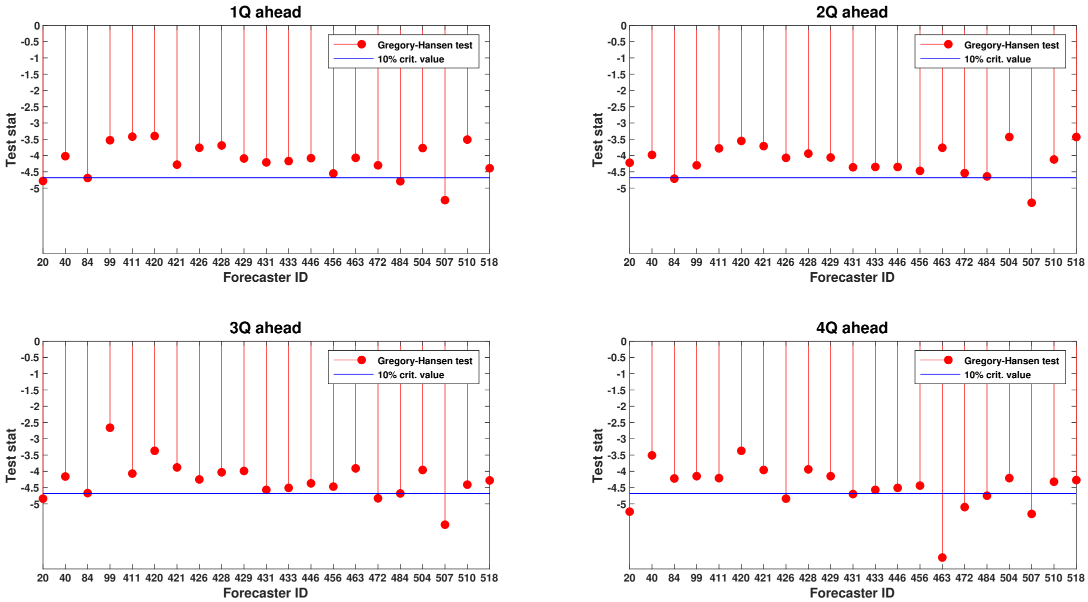
Figure 5: **Illustration of individual-level cumulative sum test (OLS residuals) for coefficient stability, forecasts of output-consumption ratios**



Now turn to the Gregory-Hansen test. The null hypothesis of the test assumes the existence of a structural break and no cointegration at the break point. Rejecting the null hypothesis implies the existence of a cointegration relation with a structural break. Figure 6 plots the Gregory–Hansen test statistics and the corresponding critical values for individual-level forecasts of output-to-consumption ratios. Each red dot stands for the test statistics for an individual forecaster, while the blue horizontal line corresponds to the 10% critical value. The null hypothesis is not rejected for the vast majority of cases. This implies that there is no cointegration between the two forecasts, taking into account structural breaks. In a small proportion of cases, the null hypothesis is rejected, implying the existence of cointegration between the two forecasts with a structural break. Overall, the results suggest heterogeneity in utilizing long-run equilibrium relationships in forecasting among the forecasters.[27]

[27]Appendix L shows that the same results are obtained when using forecasts of output and investment ratios. Moreover, it shows that the Gregory–Hansen test indicates no cointegration in median and mean forecasts.

Figure 6: **Gregory–Hansen cointegration test statistics with forecasts of output-consumption ratios**



5.3 Testing for forecasters' beliefs about mean reversion

One issue of interest is whether the forecasters believe in mean reversion to the long-run equilibrium implied by the full-information RE stochastic growth models. This section tests this question using the data of individual-level forecasts and following Armona, Fuster, and Zafar (2019). They conducted a survey experiment in which participants were provided information regarding recent local house price growth rates. They found that participants were more likely to believe in momentum in house price growth rather than mean reversion in house price growth.

First, note that for the majority of forecasters, forecasts of output are not cointegrated with forecasts of consumption (or investment). This implies that they do not believe in mean reversion to the long-run equilibrium. Second, for the forecasters whose output forecasts are cointegrated with forecasts of consumption (or investment), we tested if they believe in mean reversion to the long-run equilibrium in the following way. We regressed the revisions to forecasts of output to consumption (or investment) ratios on the forecast errors of the ratios. The SPF provides participants information about the latest data in the survey questionnaire and elicits their macroeconomic expectations. Specifically, the regression is as follows:

$$(E_t^i V_{t+1} - E_{t-1}^i V_t) = \alpha^i + \beta^i (V_t - E_{t-1}^i V_t) + u_t^i, \tag{1}$$

where i is an individual, t is time (quarter), and E_t^i is the expectation operator for individual i based on the information available at t. V represents output to consumption (or investment) ratios. α^i is an intercept term, and the slope coefficient β^i represents how strongly, and in what direction, the individual i revises their expectation in response to the forecast errors of the ratios. Table 12 reports the estimation results. The regressions yield positive regression coefficients β^i, suggesting that the forecasters are extrapolating (or believe in momentum) instead of believing in mean reversion.

Table 12: **Regression results, extrapolating v.s. mean reversion**

	Estimated coefficient (β^i)	P value
Panel A: output to consumption ratios		
ID 20	1.09***	0.00
	(0.10)	
ID 504	0.92***	0.00
	(0.11)	
ID 510	0.82***	0.00
	(0.09)	
Panel B: output to investment ratios		
ID 504	0.91***	0.00
	(0.09)	
ID 510	0.78***	0.00
	(0.09)	

Note: ***means significant at 1% level.

5.4 Cointegration among other macroeconomic variables

We examined the integration and cointegration properties of other macroeconomic variables, such as inflation, unemployment, and nominal interest rates. We briefly summarize the results and relegate the detailed testing results in Appendix O.

First, the mean and median forecasts of inflation and unemployment are I(0). Appendix O also shows that the Johansen trace and maximum-eigenvalue tests confirm the existence of multiple cointegrating vectors between inflation forecasts and unemployment forecasts. All individual forecasts of inflation are also I(0). Most individual unemployment forecasts are I(0) and a small proportion of unemployment forecasts are I(1). Thus, it is not meaningful to analyze the cointegration between inflation forecasts and unemployment forecasts.

Second, if the real interest rate is stationary, the nominal interest rate and the inflation rate are cointegrated with the (1, -1) vector, according to the Fisher equation. King, Plosser, Stock, and Watson (1991) provide some evidence for the co-integration between realized nominal interest rates and inflation. Median and mean forecasts of nominal interest rate are I(1). As mean, median, and individual inflation forecasts are I(0), there exists no coingration vector that yields cointegration between the non-stationary forecasts of nominal interest rate and the stationary inflation forecasts. The recursive trace test is utilized to illustrate the non-existence of cointegrating vectors between the forecasts of the inflation rate and the nominal interest rate in Appendix O.

5.5 Fitting forecasting models

Macroeconomists may be interested in how to model the expectation formation process of forecasters who use or do not use long-run relationships in forecasting. We explore one way to approximate the expectation formation of the two types of forecasters by fitting parsimonious recursive forecasting models (constant gain learning algorithms) to the survey forecast data, following Branch and Evans (2006). Specifically, the expectation formation process of the forecasters who utilize cointegration relationships is approximated by recursive estimation and forecasting using a cointegrated Vector Autoregressive (VAR) model which imposes cointegration among output, consumption, and investment. The expectation formation process of the forecasters who do not utilize cointegration relationships is approximated by recursive estimation and forecasting using a simple univariate autoregressive process. Appendix P.2 provides the details of fitting the models to survey forecasts data.

5.6 Forecast accuracy

Another interesting issue concerns the accuracy of forecasts made by the forecasters who utilize (or do not utilize) long-run relationships. To evaluate this, individual forecasters from the SPF are divided into two groups: those who utilize a long-run relationship in forecasting and those who do not. Forecasts made by each individual from each group are evaluated against the corresponding realized values. Table A.28 reports the accuracy of forecasts, measured by root-mean-square errors (RMSEs) over 1-, 2-, 3-, and 4-quarter horizons. The forecasters who do not use long-run cointegration relationships are generally more accurate, as they have slightly smaller average root-mean-square errors. However, these differences are unimportant economically and it remains an open question that whether utilizing the long-run relationships improves the accuracy of forecasting.

6 Conclusion

This paper studies the expectation formation of macroeconomic variables (aggregate output, consumption, and investment) by professional forecasters. We focused on analyzing whether professional forecasters utilize long-run cointegration relationships among the aformentioned macroeconomic variables postulated by stochastic growth models to make forecasts. The median (or mean) survey forecasts of aggregate output are not cointegrated with the median (or mean) forecasts of aggregate consumption and investment. Significant heterogeneity exists among forecasters in terms of utilizing long-run relationships. The majority of the forecasters do not appear to utilize these relationships in forecasting. The results are robust to using different tests, using forecasts over different horizons, forecasts made at the same or different dates, with or without imposing the theory-implied cointegrating vector, and considering the multiple testing problem and structural breaks.

The findings have implications for policymakers.[28] First, not utilizing long-run equilibrium relationships in forecasting or estimations by the policymakers may lead to mis-measurements in the output gap, which is an important input for monetary and fiscal policy. For instance, fiscal policymakers in the EU target a certain level of structural fiscal balance, which is crucially determined by estimates of the output gap. Kuang and Mitra (2020) examine the interaction between policymakes' mis-measurements of the output gap, fiscal policy decisions, and the prolonged recession in the aftermath of the 2007–08 financial crisis in European countries. Kuang, Mitra and Tang (2021) study monetary policy design when policymakers do not make use of knowledge of the balanced growth path and use various detrending methods to learn about the output gap. They find that the methods that intrinsically produce larger and more volatile output gap estimates are more prone to self-reinforcing deflation spirals and large welfare losses, and that the optimal response to the output gap estimates from these methods is smaller.

Second, households and firms who do not make use of long-run equilibrium relationships in forecasting may hold overly optimistic or pessimistic beliefs about macroeconomic variables. This may lead to inefficient fluctuations in economic decisions in areas such as consumption and investment. As an example, Kuang and Mitra (2016) develop a quantitative business cycle model where agents do not have knowledge of the balanced growth path and learn about the long-run growth rates of economic variables. The model replicates pro-cyclical forecasts of long-run growth rates of output and productivity in the data and suggests that optimism and pessimism about long-run growth rates is crucial to understanding business cycle fluctuations.

References

Adam, K., A. Marcet, and J. Beutel, "Stock price booms and expected capital gains,"

[28] Like the majority of experts, the long-run relationships are not utilized by policymakers either; Appendix C shows the evidence from the US Federal Reserves' Greenbook forecasts. One caveat is that households' expectations may deviate substantially from those of experts, as in Coibion, Gorodnichenko and Kumar (2018),and Andre, Pizzinelli, Roth, and Wohlfart (2021). Additionally, forecasts of firms often deviate significantly from those of households as well (Candia, Coibion and Gorodnichenko 2021; Link, Peichl, Roth and Wohlfart 2021). It is worthwhile to test households' or firms' expectations on the long-run relationships. For our purpose, unfortunately we do not have an appropriate dataset of survey forecasts for households or firms. Nevertheless, some influential papers modeled experts' forecasts as a crucial determinant of households' forecasts, such as Carroll (2003) and Malmendier and Nagel (2015).

American Economic Review, 2017, *107* (8), 2352–2408.

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.

Andrade, P., R. Crump, S. Eusepi, and E. Moench, "Fundamental disagreement," *Journalof Monetary Economics*, 2016, *83*, 106–128.

Andre, P., C. Pizzinelli, C. Roth, and J. Wohlfart, "Subjective models of the macroeconomy: Evidence from experts and representative samples," *Review of Economic Studies*, 2021, *Forthcoming.*

_ , I. Haaland, C. Roth, and J. Wohlfart, "Inflation narratives," 2021. Available at SSRN.

Angeletos, G. and J. La'O, "Sentiments," *Econometrica*, 2013, *81* (2), 739–779.

_ , Z. Huo, and K. Sastry, "Imperfect macroeconomic expectations: Evidence and theory," *NBER Macroeconomics Annual 2020*, 2020, *35*, 1–86.

Armona, L., A. Fuster, and B. Zafar, "Home price expectations and behaviour: Evidence from a randomized information experiment," *Review of Economic Studies*, 2019, *84* (4), 1371–1410.

Beaudry, P. and F. Portier, "An exploration into Pigou's theory of cycles," *Journal of Monetary Economics*, 2004, *51.*

**_ and _ **, "News-driven business cycles: Insights and challenges," *Journal of Economic Literature*, 2014, *52* (4), 993–1074.

Beckmann, J., "Measurement and effects of euro/dollar exchange rate uncertainty," *Journal of Economic Behavior and Organization*, 2021, *183* (C), 773–790.

Blume, L. and D. Easley, "If you're so smart, why aren't you rich? Belief selection in complete and incomplete markets," *Econometrica*, 2006, *74* (4), 929–966.

Bordalo, P., N. Gennaioli, Y. Ma, and A. Shleifer, "Overreaction in macroeconomic expectations," *American Economic Review*, 2020, *110* (9), 2748–2782.

Branch, W., "The theory of rationally heterogeneous expectations: Evidence from survey data on inflation expectations," *Economic Journal*, 2004, *114* (497), 592–621.

_ **and B. McGough**, "A New Keynesian model with heterogeneous expectations," *Journal of Economic Dynamics and Control*, 2009, *33* (5), 1036–1051.

_ **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.

Bullard, J., G. Evans, and S. Honkapohja, "Monetary policy, judgment, and near-rational exuberance," *American Economic Review*, 2008, *98* (3), 1163–77.

Candia, B., O. Coibion, and Y. Gorodnichenko, "The inflation expectations of US firms: Evidence from a new survey," 2021. NBER Working Paper, W28836.

Capistran, C. and A. Timmermann, "Disagreement and biases in inflation expectations," *Journal of Money, Credit and Banking*, 2009, *41*, 365–396.

Carroll, C., "Macroeconomic expectations of households and professional forecasters," *Quarterly Journal of Economics*, 2003, *118* (1), 269–298.

Carvalho, C. and F. Nechio, "Do people understand monetary policy?," *Journal of Monetary Economics*, 2014, *66* (C), 108–123.

Cheung, Y. and K. Lai, "Lag order and critical values of a modified Dickey-Fuller test," *Oxford Bulletin of Economics and Statistics*, 1995, *57*, 411–419.

— **and —**, "Lag order and critical values of the augmented Dickey-Fuller test," *Journal of Business and Economic Statistics*, 1995, *13* (3), 277–280.

Coibion, O. and Y. Gorodnichenko, "What can survey forecasts tell us about information rigidities," *Journal of Political Economy*, 2012, *120* (1), 116 – 159.

— **and —**, "Information rigidity and the expectations formation process: A simple framework and new facts," *American Economic Review*, 2015, *105* (8), 2644–2678.

—, —, and S. Kumar, "How do firms form their expectations? New survey evidence," *American Economic Review*, 2018, *108* (9), 2671–2713.

Dovern, J., "A multivariate analysis of forecast disagreement: Confronting models of disagreement with survey data," *European Economic Review*, 2015, *80*, 16–35.

—, U. Fritsche, and J. Slacalek, "Disagreement among forecasters in G7 countries," *Review of Economics and Statistics*, 2012, *94* (4), 1081–1096.

Drager, L., M. Lamla, and D. Pfajfar, "Are survey expectations theory-consistent? The role of central bank communication and news," *European Economic Review*, 2016, *85* (C), 84–111.

Eusepi, S. and B. Preston, "Expectations, learning, and business cycle fluctuations," *American 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.

Giglio, S., M. Maggiori, J. Stroebel, and P. Utkus, "Five facts about beliefs and portfolios," 2019. NBER Working Papers, No. 25744.

Greenwood, R. and A. Shleifer, "Expectations of returns and expected returns," *Review of Financial Studies*, 2014, *27* (3), 714–746.

Gregory, A. and B. Hansen, "Residual-based tests for cointegration in models with regime shifts," *Journal of Econometrics*, 1996, *70* (1), 99–126.

Herbst, E. and F. Winkler, "A factor structure of disagreement." Working Paper.

Hommes, C., "Heterogeneous agent models in economics and finance," *Chapter 23 in Handbook of Computational Economics*, 2006, *2*, 1109–1186.

Ilut, C. and M. Schneider, "Ambiguous business cycles," *American Economic Review*, 2014, *104* (8), 2368–2399.

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.

_ , _ , and _ , "Production, growth and business cycles: II. New directions," *Journal of Monetary Economics*, 1988, *21* (3), 309–341.

_ , _ , J. Stock, and M. Watson, "Stochastic trends and economic fluctuations," *American Economic Review*, 1991, *81* (4), 819–840.

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

_ and _ , "Potential output pessimism and austerity in the European Union," *University of Birmingham mimeo*, 2020.

_ , _ , and L. Tang, "Output gap estimation and monetary policy under imperfect knowledge," *Working paper*, 2021.

_ , L. Tang, R. Zhang, and T. Zhang, "Survey forecasts, sentiment and stock market volatility," *University of Birmingham mimeo*, 2022.

Kuchler, T. and B. Zafar, "Personal experiences and expectations about aggregate outcomes," *Journal of Finance*, 2019, *74* (5), 2491–2542.

Kurz, M., M. Motolese, M. Piccillo, and H. Wu, "Monetary policy with diverse private expectations," 2018. Working Paper.

Lahiri, K. and X. Sheng, "Measuring forecast uncertainty by disagreement: The missing link.," *Journal of Applied Econometrics*, 2010, *25* (4), 514–538.

Link, S., A. Peichl, C. Roth, and J. Wohlfart, "Information frictions among firms and households," 2021. Available at SSRN.

Lorenzoni, G., "A theory of demand shocks," *American Economic Review*, 2009, *99* (5), 2050–2084.

MacKinnon, J., "Numerical distribution functions for unit root and cointegration tests," *Journal of Applied Econometrics*, 1996, *11*, 601–618.

Malmendier, U. and S. Nagel, "Depression babies: do macroeconomic experiences affect risk taking?," *The Quarterly Journal of Economics*, 2011, *126* (1), 373–416.

_ **and** _ , "Learning from Inflation Experiences," *Quarterly Journal of Economics*, 2015, *113* (1), 53–97.

Mankiw, G., R. Reis, and J. Wolfers, "Disagreement about inflation expectations," *NBER Macroeconomics Annual*, 2003, *18.*

Monacelli, T., "New Keynesian models, durable goods, and collateral constraints," *Journal of Monetary Economics*, 2009, *56*, 242–254.

Ng, F. and R. Perron, "Lag length selection and the construction of unit root tests with good size and power," *Econometrica*, 2000, *69*, 1519–1554.

Patton, A. and A. Timmermann, "Why do forecasters disagree? Lessons from the term structure of cross-sectional dispersion," *Journal of Monetary Economics*, 2010, *57*, 803–820.

_ **and** _ , "Predictability of output growth and inflation: a multi-horizon survey approach," *Journal of Business and Economic Statistics*, 2011, *29* (3), 397–410.

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

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

—, "Testing weak cross-sectional dependence in large panels," *Econometric Reviews*, 2015, *34*, 1089–1117.

Pfajfar, D. and E. Santoro, "Heterogeneity, learning and information stickiness in inflation expectations," *Journal of Economic Behavior and Organization*, 2010, *75* (3), 426–444.

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

Scheinkman, J. and W. Xiong, "Overconfidence and speculative bubbles," *Journal of Political Economy*, 2003, *111* (6).

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.

Svensson, L., "What is wrong with Taylor rules? Using judgement in monetary policy through targeting rules," *Journal of Economic Literature*, 2003, *41*, 426–77.

—, "Monetary policy with judgement: Forecast targeting," *International Journal of Central Banking*, 2005, *1* (1), 1–54.

— **and R. Tetlow**, "Optimal policy projections," *International Journal of Central Banking*, 2005, *1* (3), 177–207.

Xiong, W. and H. Yan, "Heterogeneous expectations and bond markets," *Review of Financial Studies*, 2010, *23* (4), 1433–1466.

Appendix (Not for publication)

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)

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

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).[29] Thus,

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

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).[30] **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 $\ln(\mu_t^z/\mu^z) = D_{11}\epsilon_t^1$. This becomes a version of the one-sector stochastic growth, e.g., like King, Plosser and Rebelo (1988).

[29]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.

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

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 $X_t^Y = (X_t^z)^{\frac{1-\alpha}{1-\alpha\xi}} (X_t^a)^{\frac{\alpha}{1-\alpha\xi}}$. 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

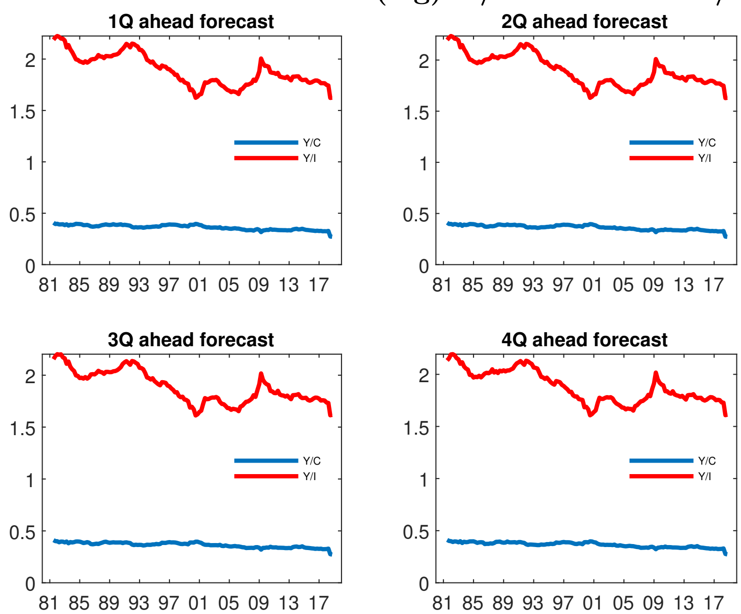
Since the Survey of Professional Forecasters (SPF) began, there have been a number of changes of the base year in the national income and product accounts (NIPA). The forecasts for levels of consumption, investment and output use the base year that was in effect when the forecasters received the survey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

Table A.1: **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

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 vintages 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**

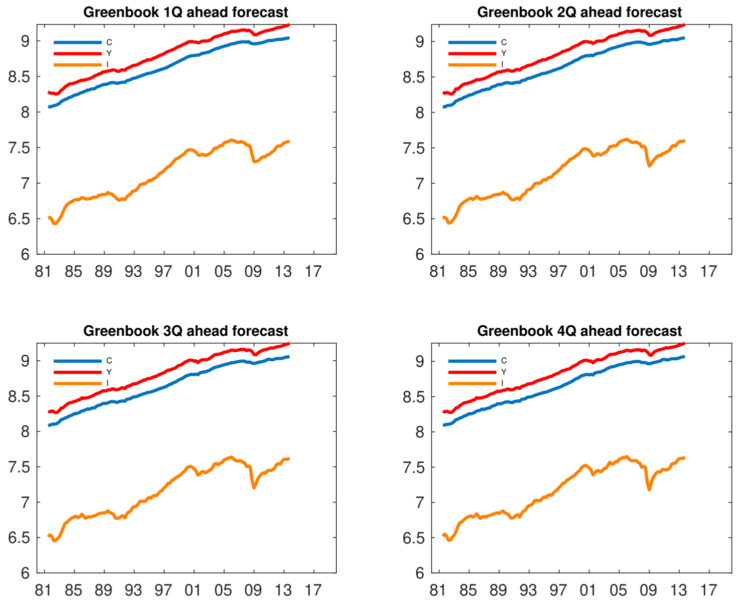


C Testing using the Greenbook forecasts

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.

Real consumption level forecast is obtained by multiplying the consumption growth forecast

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



(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.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).

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

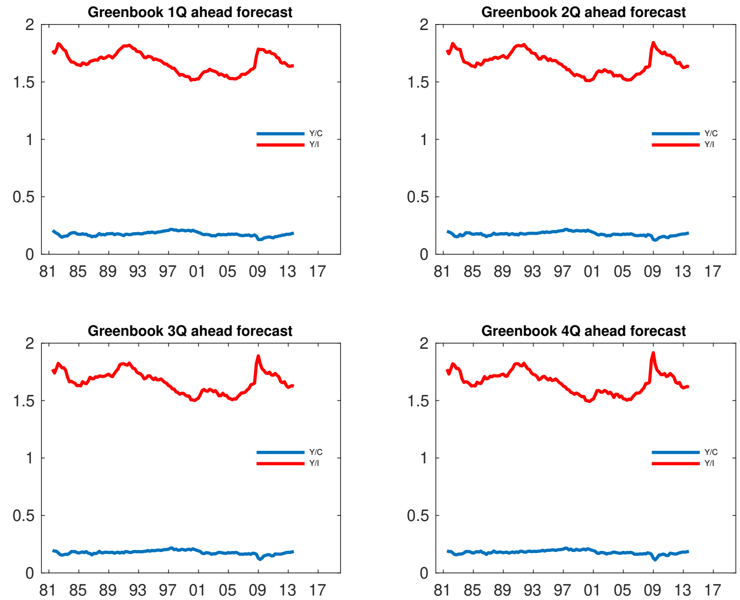


Table A.2: **Integration properties of Greenbook forecasts**

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.[31] 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.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

Table A.4 reports the Engle-Granger cointegration test outcomes when no cointegration restriction is imposed.[32] Again, the Engle-Granger test indicates that the forecasts of output are not cointegrated with the forecasts of consumption (or investment) at 10% significance level, consistent with the testing results from SPF forecasts.

[31] 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.

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

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.

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.

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.

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

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 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

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.

rest of the test statistics suggest that mean forecasts of output, consumption, and investment do not share a common trend, similar to median forecasts.

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 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.

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.

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

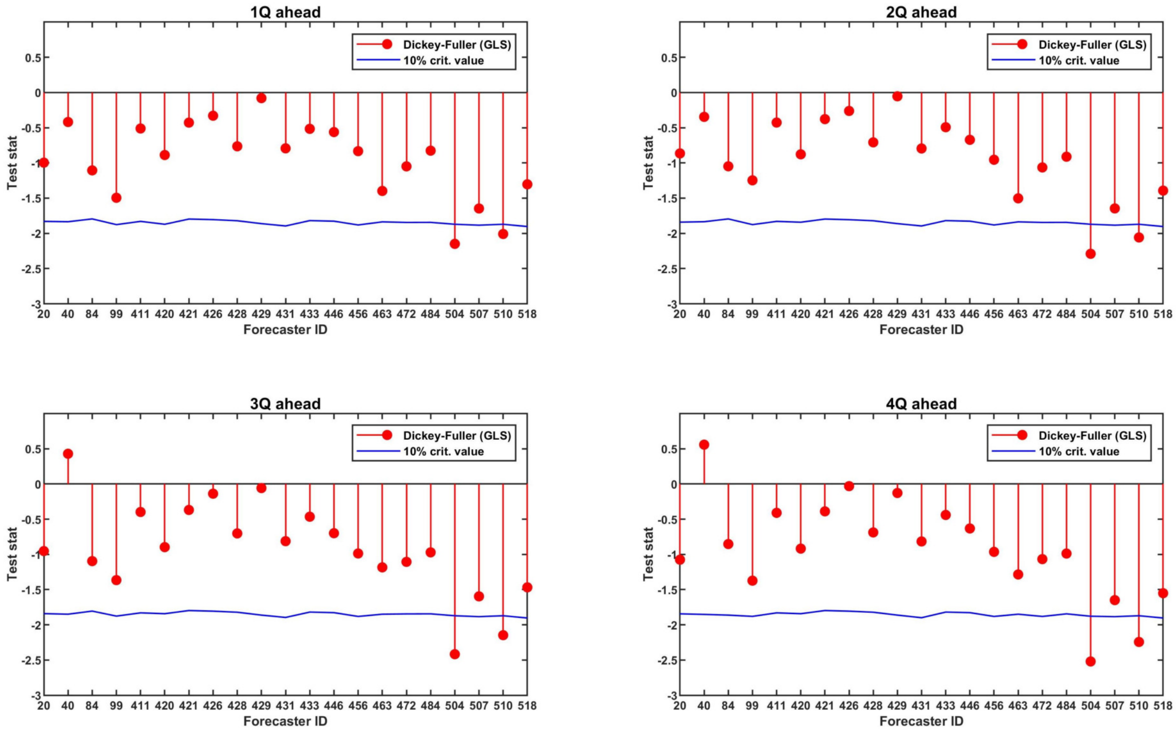


Table A.13: **Likelihood ratio test of over-identifying restrictions (individual forecasts)**

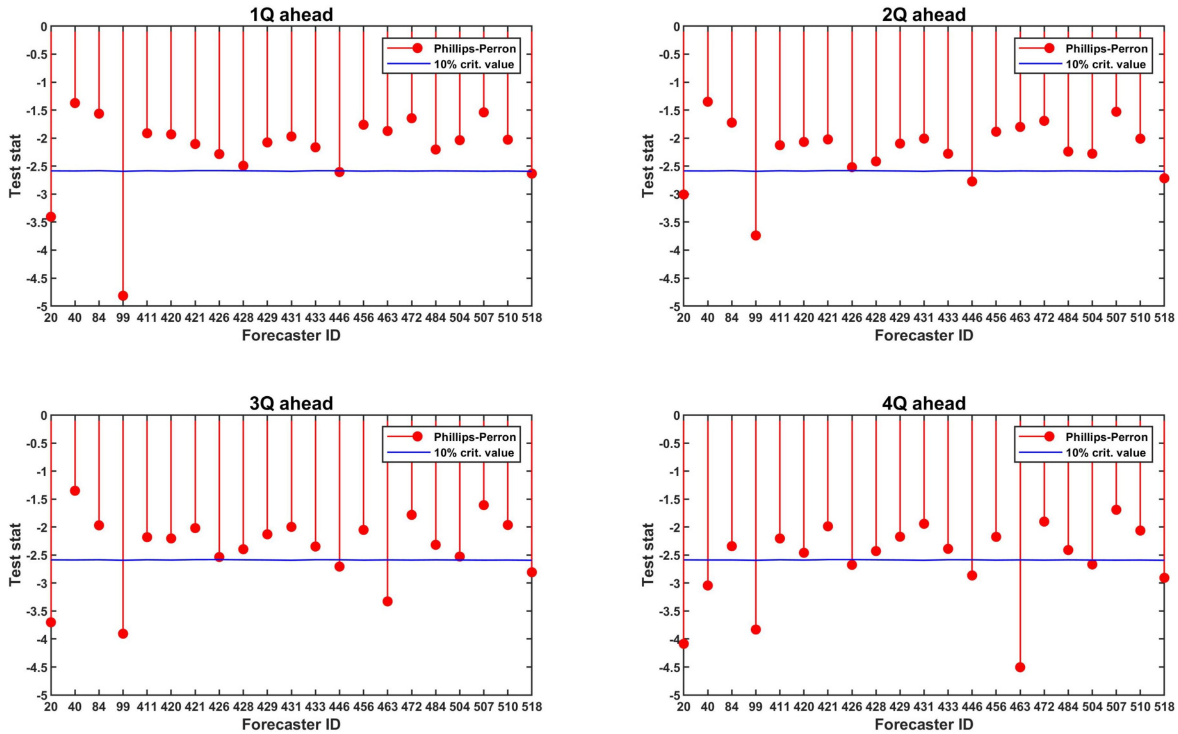
Forecasting horizon	Number of cointegration	Proportion of cointegration
1Q	3	14.3%
2Q	4	19.0%
3Q	4	19.0%
4Q	6	23.8%

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 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

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

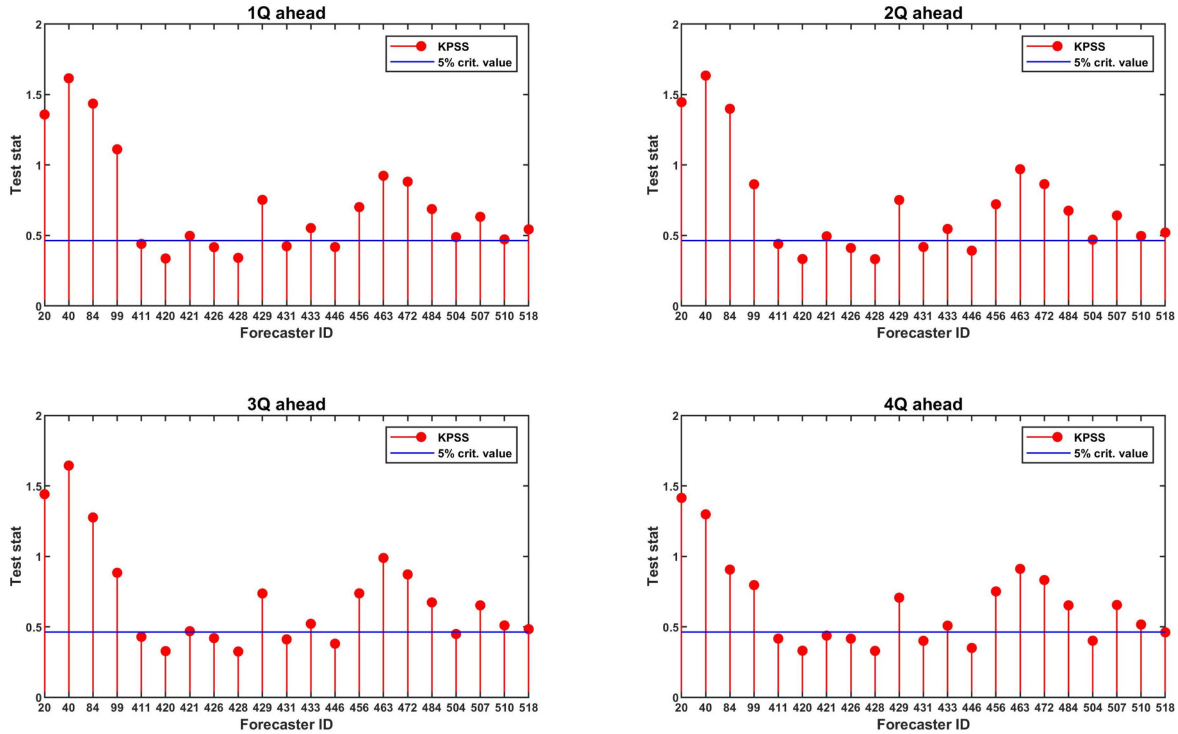


locates on the blue horizontal line.

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 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.

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



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**

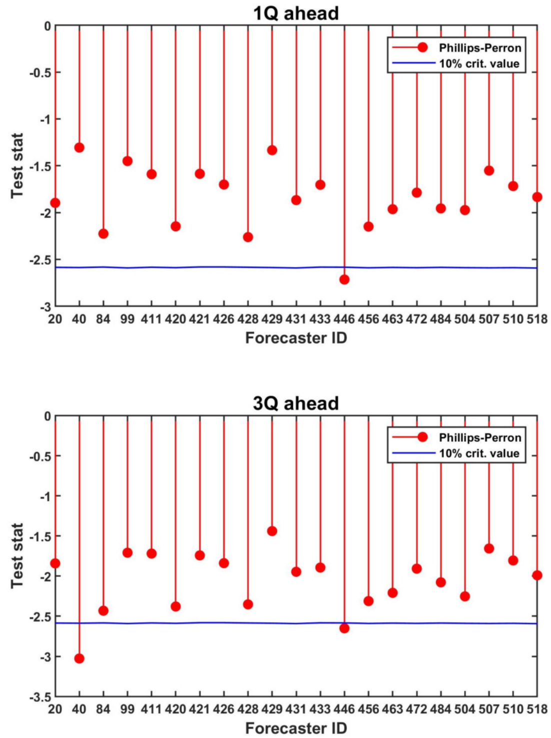
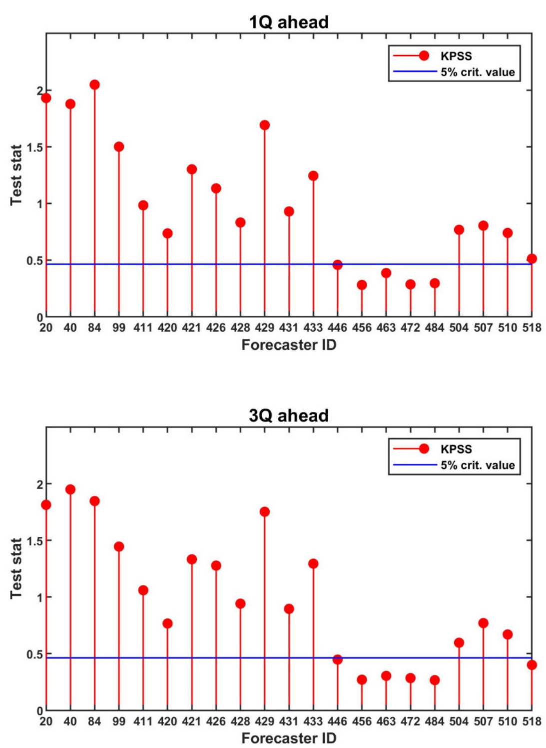


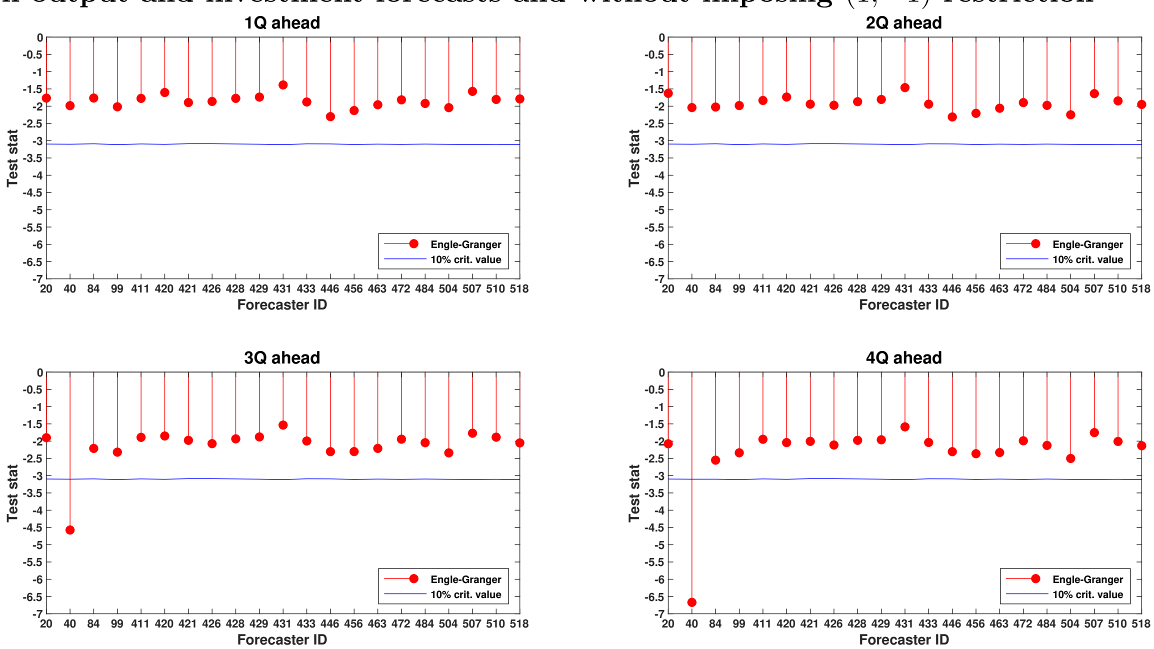
Figure A.8: **Illustration of individual level KPSS test outcomes of forecasts of output-investment ratio**



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**

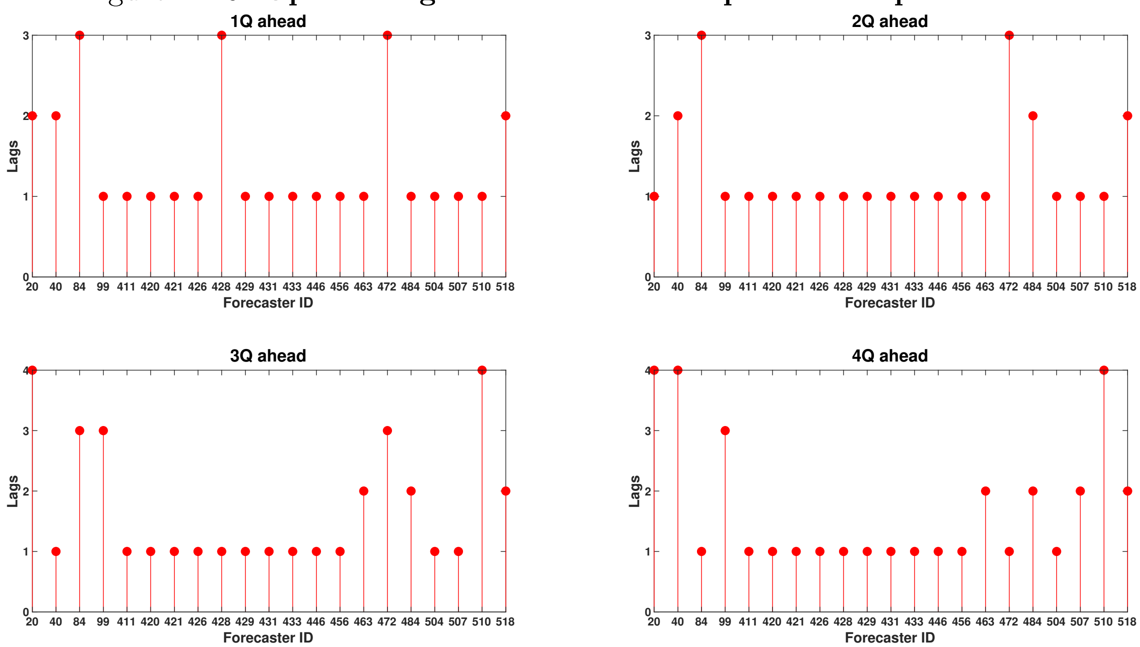


Figure A.11: **Optimal lags for forecasts of output-investment ratios**

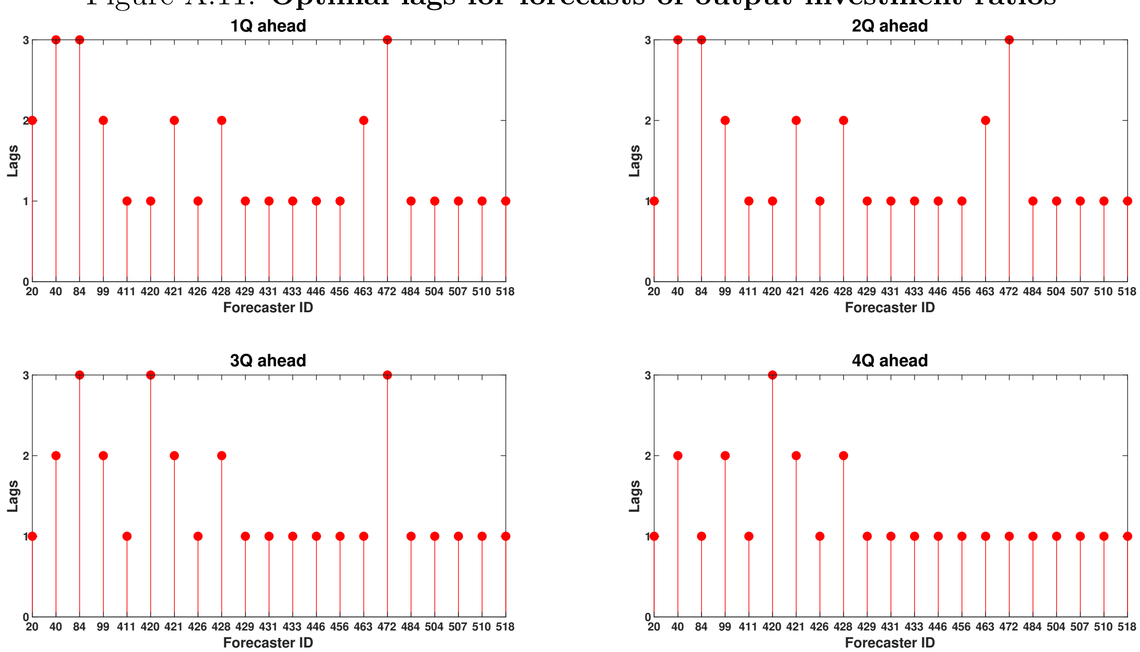
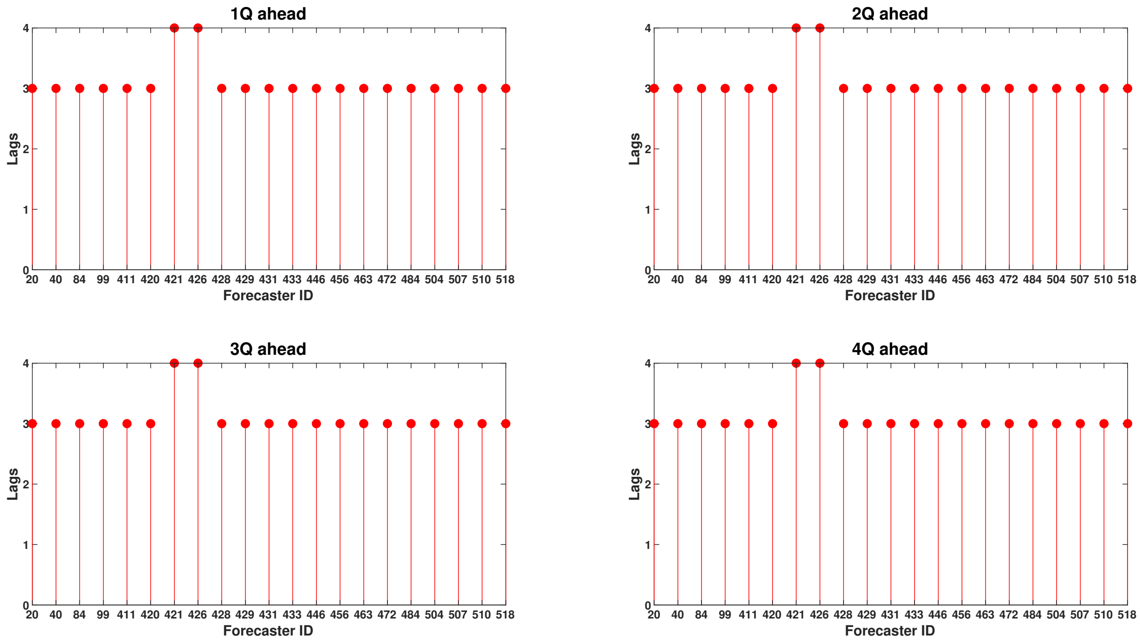


Figure A.12 plots the Newey-West optimal lags for individual-level forecasts of output-consumption ratio. The Newey-West Optimal lags for forecasts of output-investment ratio are identical to Figure A.12.

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**

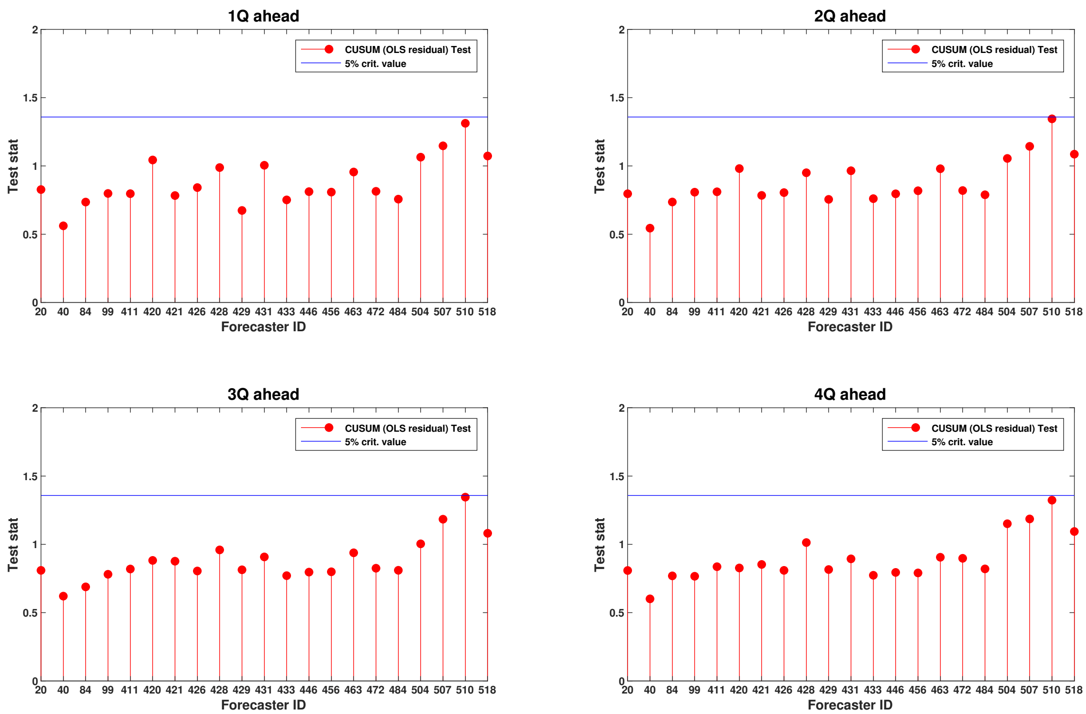


Table A.18 reports Gregory-Hansen test outcomes for median forecasts. Panel A and Panel B analyze forecasts of output-consumption ratios and output-investment ratios, respectively. As ADF and Z_t test statistics are uniformly above the corresponding 10% (and thus 5%) critical value, it implies that the 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.

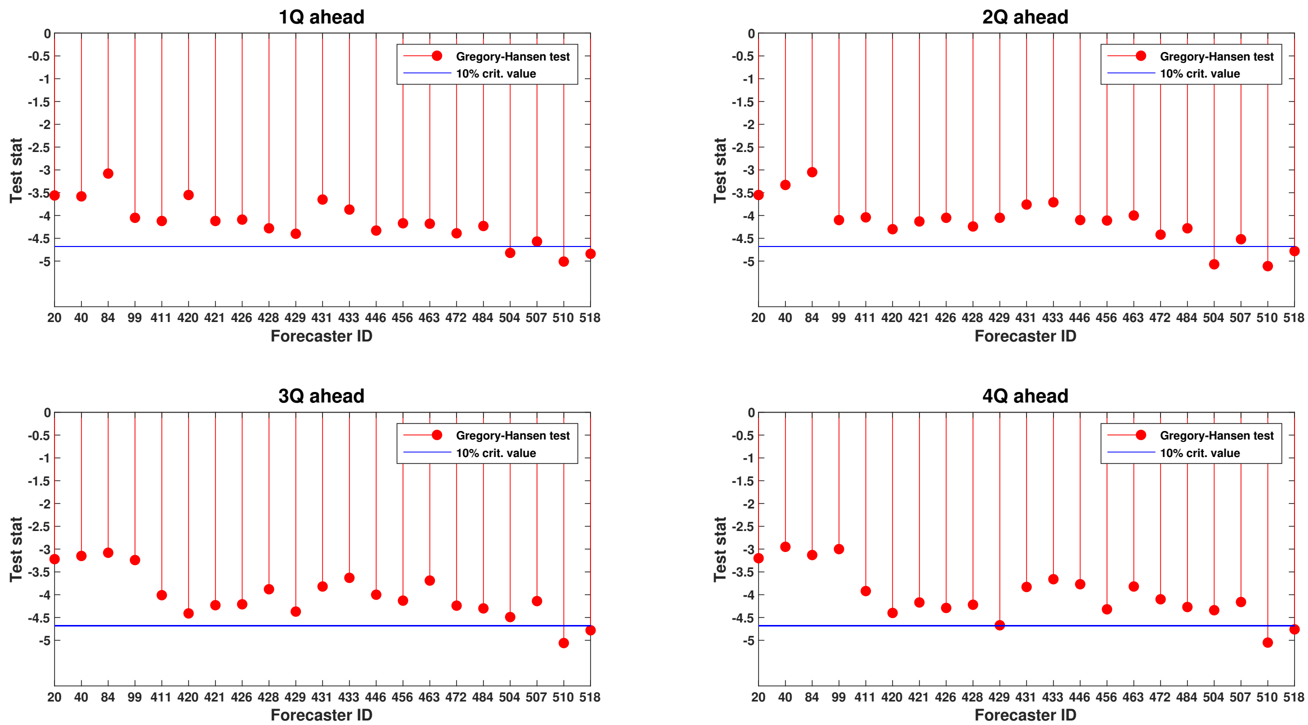
Table A.19: **Gregory-Hansen cointegration test with structural break (ADF stats) with mean output-consumption ratio forecasts**

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

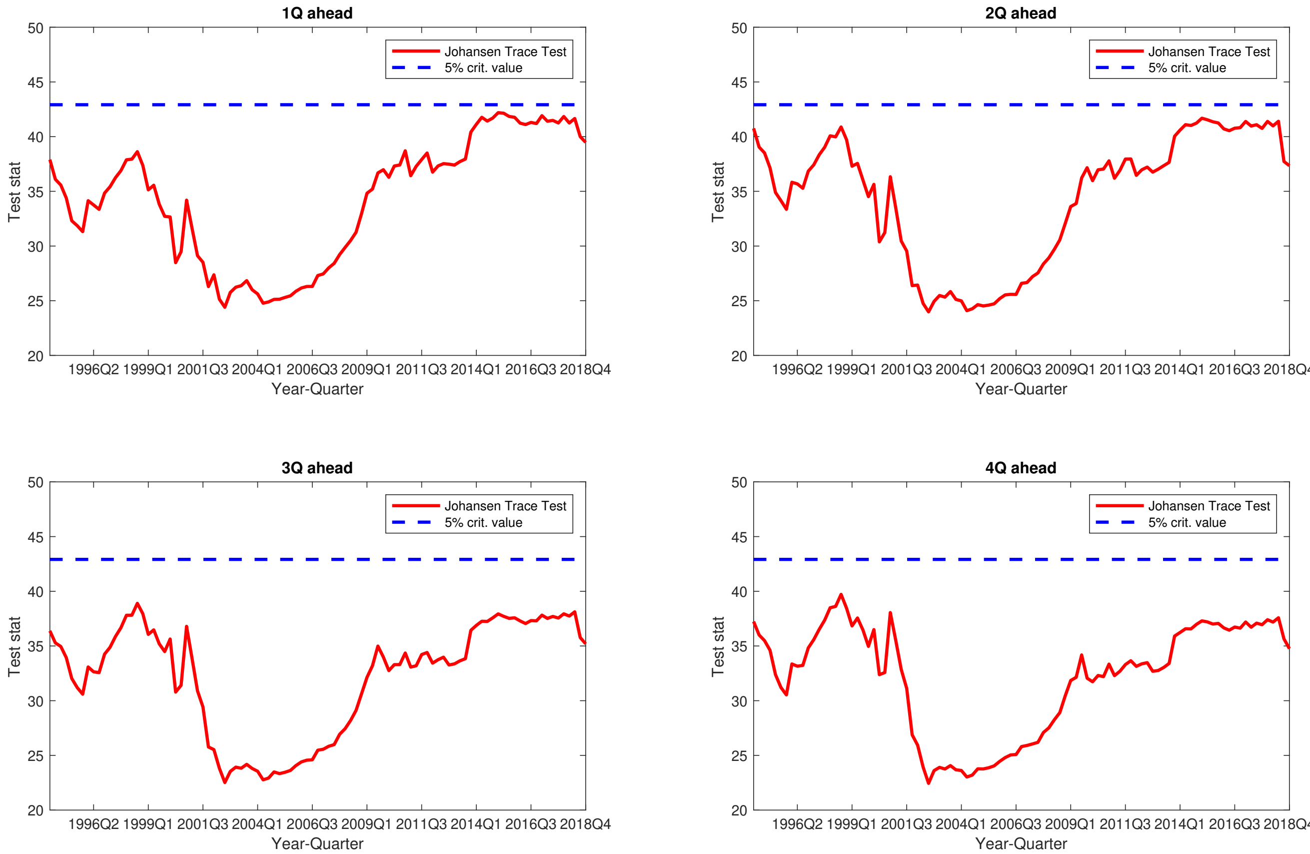


Figure A.16: Recursive Johansen trace test (rank = 0) for common trend, mean forecasts of output, consumption and investment

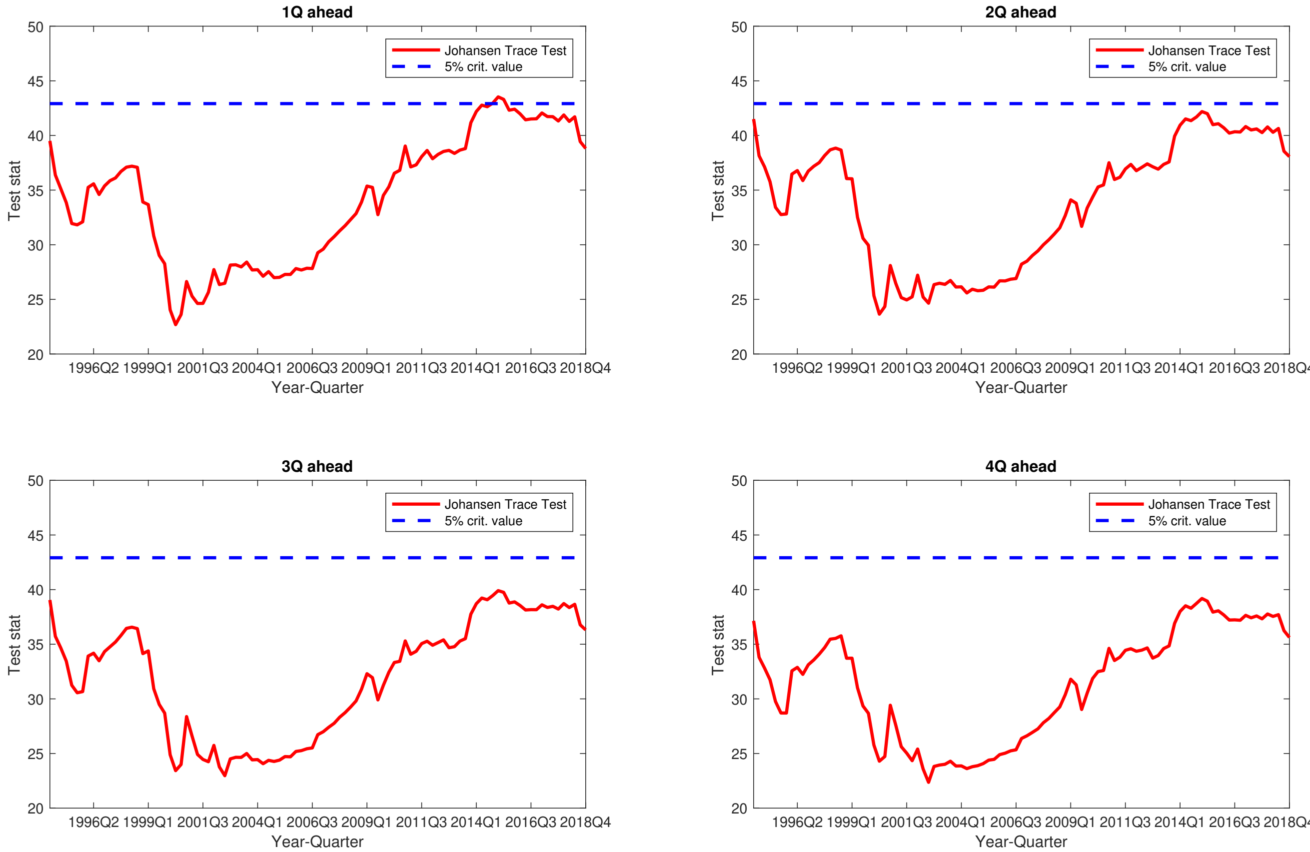
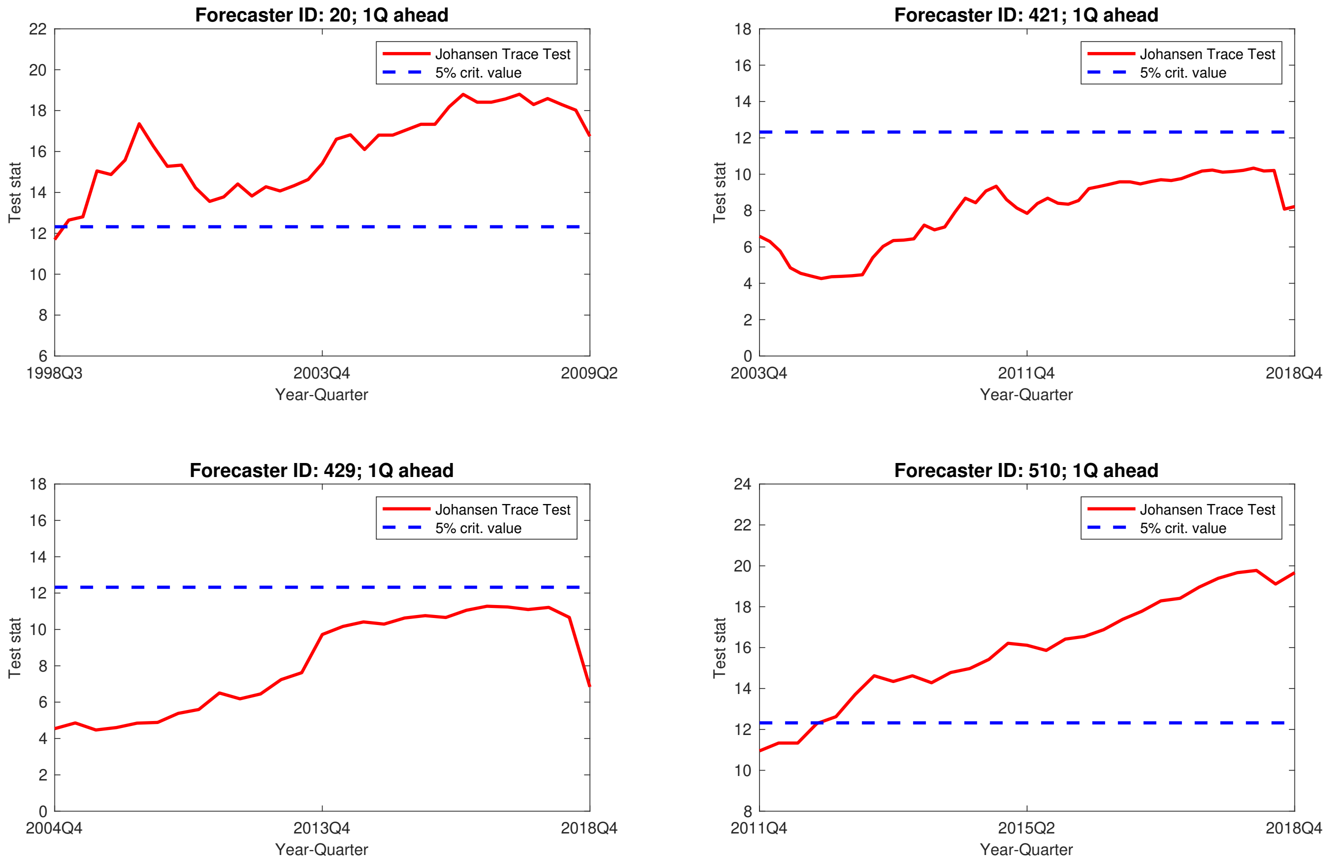


Figure A.17 illustrates the recursive trace test statistics for several forecaster IDs against the corresponding 5% critical values. For ID 20 and 510, as the sample becomes longer, the null hypothesis is rejected. For ID 421 and 429, the null hypothesis cannot be rejected for the whole rolling sample.

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

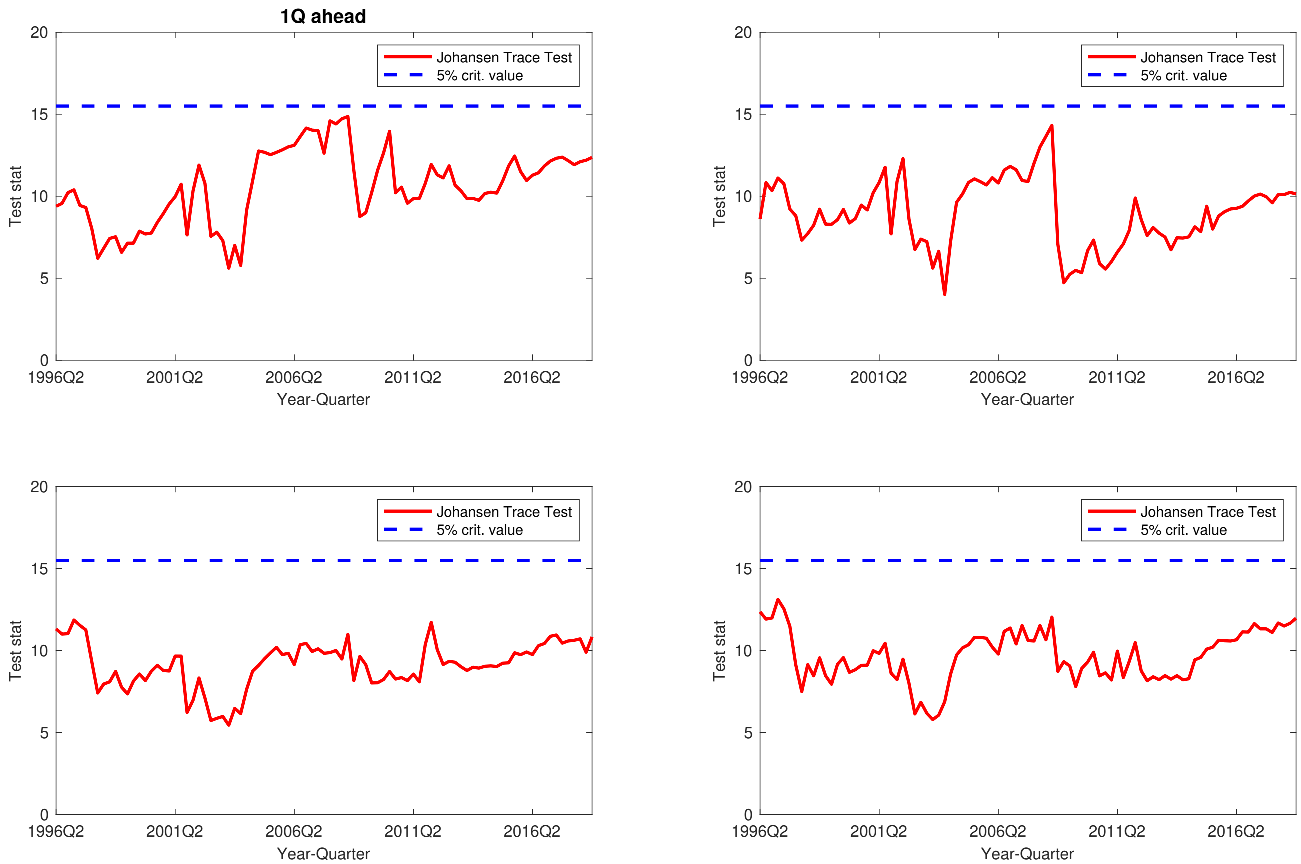
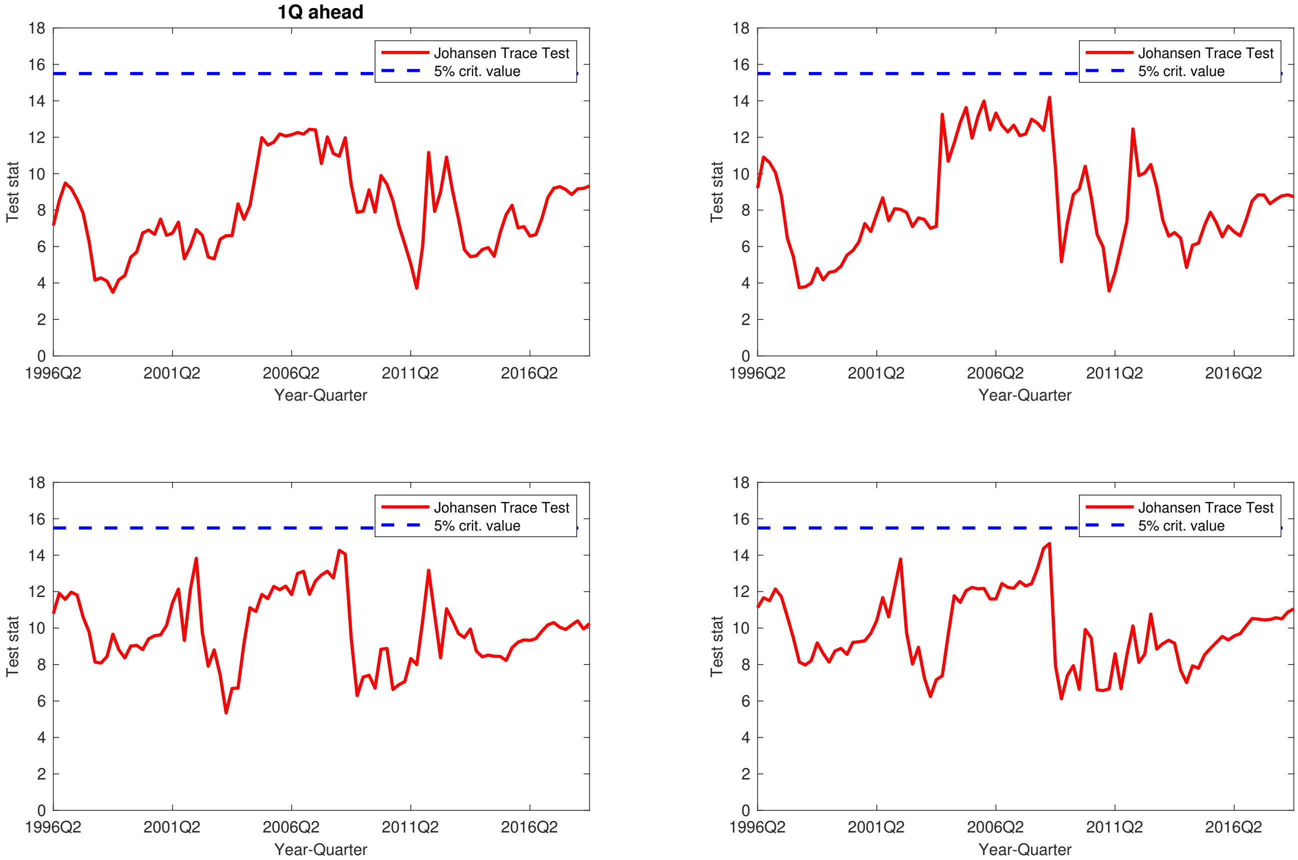


Figure A.19: Recursive Johansen trace test (rank = 0) for common trend, mean forecast of nominal interest and inflation



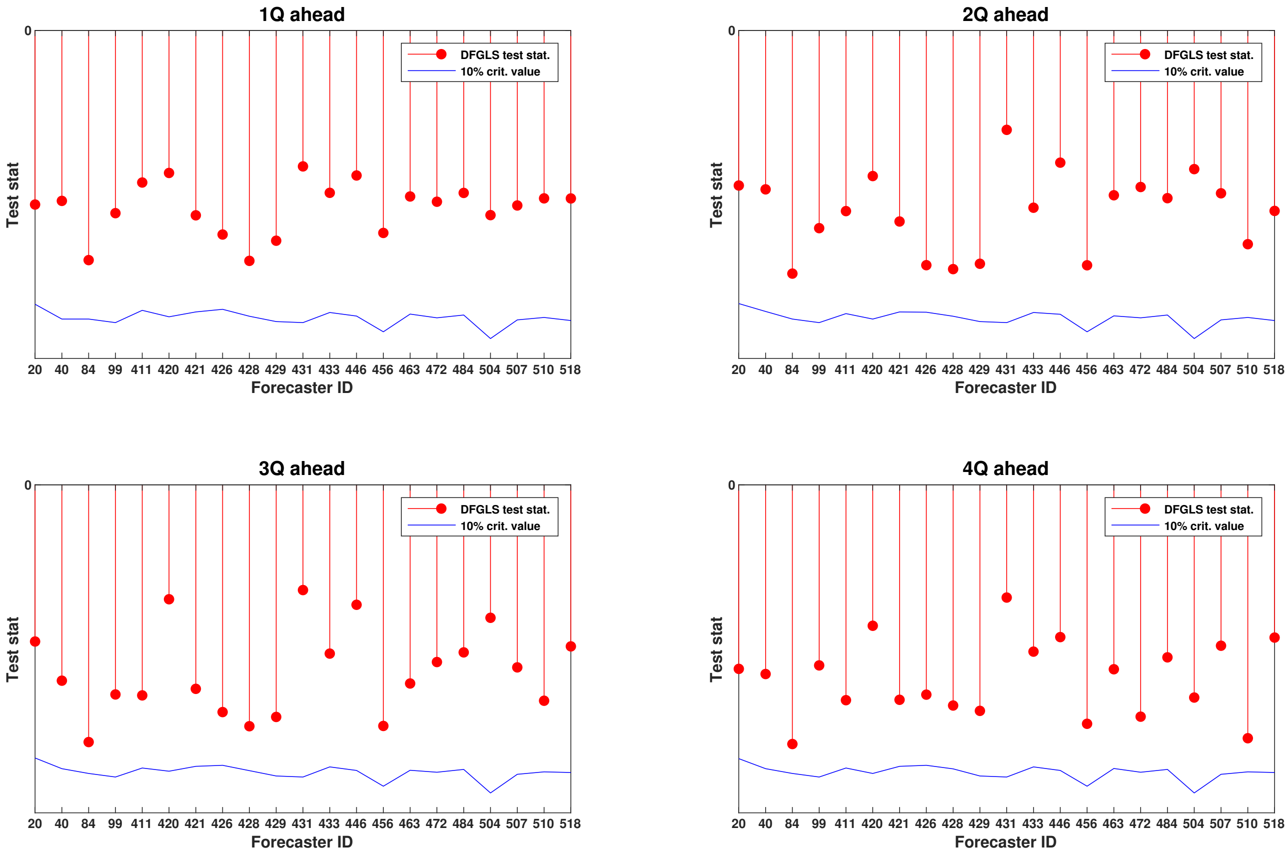
Next, we test the integration property of individual-level forecasts. Table A.27 reports the numbers and proportion of individual-level forecasts that are I(1). We find that the all individual forecasts of nominal interest rate are I(1). Again, this implies that for each forecaster, there exists no cointegrating 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.[33] 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

[33]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. [34]

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

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

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.

[34] 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.