An Instrumental Variables Approach to Testing Forecast Efficiency*

Tucker S. McElroy[a] and Xuguang Simon Sheng[b]

[a] *U.S. Census Bureau*

[b] *American University*

Abstract

To achieve forecast efficiency, it is necessary that forecast errors are unpredictable from forecast revisions. Coibion and Gorodnichenko (2015) propose aggregating forecasts and estimating the aggregated time series regression. Bordalo, Gennaioli, Ma, and Shleifer (2020) suggest estimating the average relationship by running separate regressions for each individual and then aggregating. We demonstrate that both estimators can be asymptotically biased in the presence of public noise. To address these biases, we suggest instrumenting forecast revisions with past forecast errors. The Anderson-Rubin likelihood ratio test can be applied to test for forecast efficiency, and remains robust even in the presence of weak instrumental variables. Applications of the tests to the U.S. Survey of Professional Forecasters covering 1968Q4 through 2016Q4 clearly reveal experts' underreaction to news in their macroeconomic expectations.

Keywords: Forecast efficiency, information rigidity, weak instrument

JEL Classification: E31, E37

*We are indebted to Hashem Pesaran for giving us many helpful suggestions, Yueran Ma for sharing the code for cleaning the Survey of Professional Forecasters data, and Badi Baltagi, Carola Binder, Constantin Burgi, John Chao, Oli Coibion, Julio Ortiz, Tara Sinclair, Herbert Zhao and conference participants at the JSM, 25th Federal Forecasters Conference, 29th International Panel Data Conference, and 5th Annual DC-MD-VA Econometrics Workshop for their useful comments on an earlier draft. The views expressed on statistical issues are those of the authors and not those of the U.S. Census Bureau. Corresponding author: Xuguang Simon Sheng, American University, 4400 Massachusetts Avenue NW, Washington, DC 20016, USA. E-mail address: sheng@american.edu.

1 Introduction

Full-information rational expectations (FIRE) require that forecast errors be unpredictable from forecast revisions (assuming quadratic loss). The influential work of Coibion and Gorodnichenko (2015, CG hereafter) suggests testing this relationship by first aggregating the forecasts across individuals and then estimating the aggregated time series regression. In contrast, Bordalo, Gennaioli, Ma, and Shleifer (2020, BGMS hereafter) recommend estimating the average relationship by first running separate regressions for each individual and then aggregating (i.e., taking the mean or median of the coefficients obtained from the first step). While CG (2015) document the underreaction of consensus forecasts to news relative to FIRE, BGMS (2020) find that individual forecasters typically overreact. So there remains a puzzle about whether forecasters overreact or underreact. Given that the expectations formation process significantly affects macroeconomic dynamics and policy decisions, a careful examination of the magnitude of deviation from FIRE cannot be overemphasized.

In this paper we analyze both CG's "average then estimate" approach and the BGMS' "estimate then average" approach. The target parameter in CG is the ratio of information rigidity to the Kalman gain, which we designate as "relative information rigidity." We show that both estimators are asymptotically biased (as estimators of relative information rigidity) in the presence of public noise. This is because when regressing forecast error on forecast revisions (with coefficient given by the relative information rigidity), the regression errors are correlated with forecast revisions due to the presence of public noise in both terms. Private noise also plays a role in the bias, but in the "average then estimate" approach its impact diminishes as the number of agents increases. As shown in Theorem 1 and Corollary 1 below, the bias depends upon the Kalman gain, the representativeness parameter θ (which governs the degree to which beliefs deviate from rational updating), and the persistence of the underlying signal.

Consistent estimation of relative information rigidity will facilitate testing of forecast efficiency. To circumvent the problem of asymptotic bias in the ordinary least squares estimators, we turn to the methodology of instrumental variables. We demonstrate in this paper that instrumental variables, constructed as past forecast errors, are uncorrelated with the regression errors (i.e., being exogenous), yet they may be highly correlated with the independent variables (in the regression of forecast error on forecast revision); see Theorem 2 and Corollary 2. Our simulations find this technique works well if $\theta = 0$, because the instrumental variables have sufficient correlation with the covariates to produce a viable estimator. In contrast, as θ increases from zero (it is bounded above by the relative information rigidity), the correlation of covariate and instrumental variable decreases sharply. Fortunately, Moreira (2003) proposes a likelihood ratio test that remains robust in the presence of weak instrumental variables. In our specific context, where we have only one instrumental variable, the likelihood ratio test of Moreira (2003) simplifies to the test proposed by Anderson and Rubin (1949), and is further extended by Mikusheva and Sun (2022) to a jackknifed version that is robust to heteroscedasticity.

Our work builds on the panel data literature on aggregation and pooling. See, for example, Baltagi, et al. (2000) and Pesaran and Zhou (2018) for general discussions on "to pool or not to pool." Bonham and Cohen

(2001) further argue that, due to heterogeneous individual forecasts and private-information bias, neither aggregation nor pooling is a valid strategy in testing the rational expectations hypothesis using survey data. We contribute to this literature by showing that both aggregation and pooling are appropriate only if we instrument individual (average) forecast revisions by individual (average) past forecast errors in a panel data setting.

Our work contributes to the recent debate on underreaction or overreaction to new information in macro forecasts. CG (2015), Drager and Lamla (2017) and Giacomini, et al. (2020) highlight underreactions in inflation forecasts from professionals, consumers, and market participants. Lahiri and Sheng (2008) note that professional forecasters underweight public information in one-year-ahead GDP and inflation forecasts. Fuhrer (2018) discusses forecaster underreactions and argues for "intrinsic inflation persistence." Ryngaert (2023) explains information frictions by agents' misperceptions about the persistence of the underlying process. Chang and Levinson (2023) identify time-varying inefficiency in the Federal Reserve staff's GDP forecasts.

On the other hand, Binder (2017) reveals that low-frequency data and survey rounding can overestimate information stickiness. Baker, et al. (2020) observe declining information rigidity after natural disasters. BGMS (2020) support the absence of information stickiness at the individual forecaster level. Messina, et al. (2015) find over-responsiveness in Greenbook forecasts. Bürgi and Ortiz (2022) document evidence of individual overreaction to news at the quarterly frequency.

Our contribution involves proposing an instrumental variables approach to reduce asymptotic bias, as well as suggesting a likelihood ratio test –

with correct size and power in the presence of weak instrumental variables – for forecast efficiency tests (where efficiency is based on quadratic loss). This approach indicates that professional forecasters, both at the individual and aggregate levels, tend to underreact to new information, and thus resolves the puzzle of underreaction/overreaction in macro forecasts.

The paper proceeds as follows. Section 2 presents the forecast efficiency test, derives the asymptotic bias of both "average then estimate" and "estimate then average" estimators, proposes an instrumental variables approach, and discusses the likelihood ratio test. Section 3 explores the small sample performance of the likelihood ratio test in Monte Carlo experiments. Section 4 illustrates an empirical study in testing efficiency in professional's forecasts. Section 5 concludes. Additional simulation results and derivations are relegated to the Appendix.

2 Analysis of Regressing Forecast Error on Forecast Revision

2.1 Forecast Efficiency Tests and Their Biases

In CG (2015), a (latent) signal process $\{\pi_t\}$ – which is observed under the obfuscation of private noise – is forecasted h steps ahead by N agents. Specifically, each agent observes a noisy process $\{y_t^i\}$ given by

$$y_t^i = \pi_t + \epsilon_t^i,$$

where $\{\epsilon_t^i\}$ is i.i.d., mean zero, and has a common variance. We make a slight generalization of the CG framework, supposing that the noise process is the straight summation of a public noise ι_t and a private noise η_t^i:

$$\epsilon_t^i = \iota_t + \eta_t^i.$$

We assume that the public noise $\{\iota_t\}$ is i.i.d. with common variance σ_ι^2, and that the private noise $\{\eta_t^i\}$ is i.i.d. and agent-dependent, but with common variance σ_η^2; the two noise processes are independent of one another. Hence the variance of ϵ_t^i is $\sigma_\iota^2 + \sigma_\eta^2$, which is denoted by σ_ϵ^2 for short; so $\sigma_\iota = 0$ corresponds to the CG framework. Note that these latent public and private noises only enter our calculations through their sum ϵ_t^i. The signal is assumed to be an autoregressive process of order one, i.e.,

$$\pi_t = \rho\, \pi_{t-1} + u_t$$

with $-1 < \rho < 1$ and $\{u_t\}$ an i.i.d. Gaussian sequence with mean zero and variance σ_u^2. We denote the (purely rational) h-step ahead forecast by $\pi_{t+h|t}^i$ for agent i, where the notation conveys that a conditional expectation of π_{t+h} is computed based on the information of the ith agent up through time t (i.e., the set $\{y_t^i, y_{t-1}^i, \ldots\}$). The average forecast over all such agents is denoted $\overline{\pi}_{t+h|t}$, i.e., $N^{-1}\sum_{i=1}^N \pi_{t+h|t}^i = \overline{\pi}_{t+h|t}$.

CG (2015) study the regression of the average forecast error $\pi_{t+h} - \overline{\pi}_{t+h|t}$ upon the average forecast revision $\overline{\pi}_{t+h|t} - \overline{\pi}_{t+h|t-1}$ ("average then estimate"); the forecast efficiency test examines the null hypothesis that the regression coefficient is zero. Instead of considering this type of regression based on aggregates, one can use individual agents, constructing

N individual regressions and taking the average of the coefficients ("estimate then average"). Note that the forecast efficiency test examined in this paper is distinct from the weak efficiency test introduced by Nordhaus (1987). Weak efficiency entails that current forecast errors should not be correlated with *past* forecast revisions, as described in Proposition 1 in Nordhaus (1987)'s work. In contrast, our test focuses on the absence of correlations between current forecast errors and *current* forecast revisions.

This basic framework is extended and encapsulated by the so-called Diagnostic Kalman filter (DKF) introduced in BGMS (2020), which allows explicitly for "distorted retrieval from memory" and generalizes the purely rational framework of CG. The DKF framework supposes that forecasts are updated using a measure of agent sentiment; Proposition 1 of BGMS provides the updating mechanism for the DKF, defined for any value of $\theta \geq 0$, which "denotes the extent to which beliefs depart from rational updating due to representativeness." Their result states that

$$\pi^{i,\theta}_{t|t} = \pi^{i}_{t|t-1} + (1+\theta)\, G\, (y^i_t - \pi^i_{t|t-1}), \tag{1}$$

where $G = P/(P + \sigma^2_\epsilon)$ and P is the one-step (rational) prediction mean squared error (i.e., $P = \mathrm{Var}[\pi_t - \pi^i_{t|t-1}]$).[1] Actually, (1) is a slight generalization of equation (10) in BGMS, since in that paper there is no public noise. For our purposes we can take (1) as the definition of $\pi^{i,\theta}_{t|t}$, noting that the $\theta = 0$ case of (1) is the classic Kalman filter (where G represents the so-called Kalman gain) studied in CG (but extended to include public

[1] Although it may seem that P, and hence G, should depend on i, it does not, since P is expressible as the solution to a quadratic equation that depends on the parameters ρ, σ^2_u, and σ^2_ϵ, none of which depend on i.

noise). It is shown in BGMS that $\pi^{i,\theta}_{t+h|t} = \rho^h \pi^{i,\theta}_{t|t}$; the authors also define $\pi^{\theta}_{t+h|t}$ by averaging $\pi^{i,\theta}_{t+h|t}$ over the N agents.

BGMS (and CG) consider the regression of forecast error on forecast revision, and in their Proposition 2 present expressions for the coefficient on forecast revision, labeled as β; in CG this coefficient is found to be $\beta = (1-G)/G$. CG refer to $1-G$ as the information rigidity, and note that a value of zero corresponds to $G = 1$. Thus $\beta = (1-G)/G$ is the ratio of information rigidity to the Kalman gain; we call this "relative information rigidity," noting that it is not constrained to be in $[0, 1]$, but can be any non-negative value. The case of $\beta = 0$ corresponds to $G = 1$ (the absence of information rigidity), i.e., forecast efficiency in the CG case of no public noise. We are interested in estimation of β from a regression of forecast errors on forecast revisions, thereby obtaining insight into forecast efficiency; estimates of β that are significantly different from zero indicate inefficiency and the presence of a degree of information rigidity for $\beta > 0$.

The estimation of β in BGMS and CG is based on the assumption that the regression errors are uncorrelated with the forecast revisions. However, the presence of public noise can invalidate this assumption, as our results below indicate. In particular, we investigate both consensus and individual forecasts, either under the rational ($\theta = 0$) or irrational ($\theta > 0$) frameworks, when either or both of public and private noise is present. We find that the forecast efficiency test of $\beta = 0$ based upon ordinary least squares (OLS) can be degraded due to bias in its estimator. Our results explicitly show how the regression errors are connected to forecast revisions, and thereby how bias can arise.

Let $w^i_t = \pi_{t+h} - \pi^{i,\theta}_{t+h|t}$ be the DKF h-step ahead forecast error, which is supposed to be eventually observable (since the signal π_{t+h} becomes known in time period $t+h$). Also, let $x_t^i = \pi_{t+h|t}^{i,\theta} - \pi_{t+h|t-1}^{i,\theta}$ be the DKF forecast revision. The consensus regression considered by CG is

$$\overline{w}_t = \beta\, \overline{x}_t + e_t,$$

where $\overline{w}_t = N^{-1} \sum_{i=1}^{N} w_t^i$ and $\overline{x}_t = N^{-1} \sum_{i=1}^{N} x_t^i$ are the consensus forecast errors and forecast revisions, respectively. There is no constant term in the regression, because both forecast errors and forecast revisions have mean zero. Here, the target of estimation is $\beta = (1-G)/G$, the relative information rigidity, where $G = P/(P+\sigma_\epsilon^2)$. The "average then estimate" approach computes the OLS estimate of β based on the consensus regression:

$$\overline{\beta} = \frac{\sum_t \overline{w}_t \overline{x}_t}{\sum_t (\overline{x}_t)^2}.$$

An extension of these ideas yields the individual regression:

$$w_t^i = \beta\, x_t^i + e_t^i$$

for each $1 \le i \le N$. As noted below equation (1), for each i the prediction mean squared error P does not depend on i, and hence neither does G or β. Then the individual OLS regression coefficient estimates are

$$\widehat{\beta}_i = \frac{\sum_t w_t^i x_t^i}{\sum_t (x_t^i)^2},$$

and in the "estimate then average" approach we construct $\widetilde{\beta} = N^{-1} \sum_{i=1}^{N} \widehat{\beta}_i$.

If the regression errors are uncorrelated with the covariates, then the estimates of relative information rigidity are asymptotically unbiased, but as shown below this situation can fail when public noise is present. As usual, bias is assessed via computing $E[\widehat{\beta}_i]-\beta$ (individual case) or $E[\overline{\beta}]-\beta$ (consensus case).

To state our next result, we introduce some notations. Let $\nu_t^h = \sum_{k=0}^{h-1} \rho^k u_{t+h-k}$, and define the trivariate time series

$$\mathbf{f}_t^i = \begin{bmatrix} \epsilon_t^i \\ \pi_t - \pi_{t|t-1}^i \\ \pi_{t|t-1}^i - \rho y_{t-1}^i \end{bmatrix}, \tag{2}$$

which consists of the public and private noise ϵ_t^i, the past forecast error (for the rational case $\theta = 0$) $\pi_t - \pi_{t|t-1}^i$, and a function $\pi_{t|t-1}^i - \rho y_{t-1}^i$ of the previous period's information. This vector process, together with

$$\boldsymbol{\tau} = \rho^h \, G \begin{bmatrix} 1+\theta \\ 1+\theta \\ \theta/(1-G) \end{bmatrix} \qquad \boldsymbol{\gamma} = -\rho^h \begin{bmatrix} 1+\theta \\ \theta \\ \theta \end{bmatrix}, \tag{3}$$

is featured in our first result, which quantifies the bias arising from correlation between the forecast revisions and the regression error.

Theorem 1. *The regression of individual forecast error on forecast revision in the DKF framework can be expressed as*

$$w_t^i = \frac{1-G}{G} \, x_t^i + \nu_t^h - \rho^h((1+\theta)\epsilon_t^i - \rho\theta\epsilon_{t-1}^i + \theta u_t), \tag{4}$$

i.e., the regression error is $e_t^i = \nu_t^h - \rho^h((1+\theta)\epsilon_t^i - \rho\theta\epsilon_{t-1}^i + \theta u_t)$. *For each* i, $\{\mathbf{f}_t^i\}$ *defined in (2) is a stationary time series with contemporaneous* *covariance matrix*

$$Var[\mathbf{f}_t^i] = diag\{\sigma_\epsilon^2, P, \rho^2(1-G)^2(P+\sigma_\epsilon^2)\},$$

and the forecast revision covariate can be expressed as $x_t^i = \boldsymbol{\tau}'\boldsymbol{f}_t^i$*; the regression error is* $e_t^i = \nu_t^h + \boldsymbol{\gamma}'\boldsymbol{f}_t^i$*. The asymptotic bias of* $\widehat{\beta}_i$ *is*

$$\frac{\boldsymbol{\tau}'\, Var[\mathbf{f}_t^i]\boldsymbol{\gamma}}{\boldsymbol{\tau}'\, Var[\mathbf{f}_t^i]\boldsymbol{\tau}} = -\frac{((1+\theta)^2+\theta^2\rho^2)\sigma_\epsilon^2 + (1+\theta)\theta P}{G((1+\theta)^2+\theta^2\rho^2)(\sigma_\epsilon^2+P)}.$$

Remark 1. *The asymptotic bias for the individual regression can be re-expressed as*

$$-\frac{(1-G)}{G} - \frac{\theta(1+\theta)}{(1+\theta)^2+\theta^2\rho^2},$$

which in the rational case (i.e., $\theta = 0$*) reduces to* $-(1-G)/G = -\beta$*. Hence the OLS estimator has mean tending to zero, indicating that forecast errors are uncorrelated with forecast revisions. Specifically, the asymptotic bias in the individual regression is zero if and only if* $G = 1$*, which corresponds to* $\beta = 0$*. More generally, when* $\theta > 0$ *the asymptotic bias is negative. Note that our result also covers the special case where there is no public noise, in which case* $\sigma_\epsilon^2 = \sigma_\eta^2$*.*

Clearly the bias of $\widetilde{\beta}$ is the same as that of $\widehat{\beta}_i$. On the other hand, in the case of consensus forecasts – obtained by averaging over the agents – we obtain a bias result somewhat different from that stated in Theorem 1. Specifically, let $\overline{P} = \mathrm{Var}[\pi_t - \overline{\pi}_{t|t-1}]$ (which is the solution to a quadratic equation similar to that used to define P, only involving $\sigma_{\overline{\epsilon}}^2 = \sigma_\iota^2 + \sigma_\eta^2/N$ in lieu of σ_ϵ^2) and $\overline{G} = \overline{P}/(\overline{P}+\sigma_{\overline{\epsilon}}^2)$. The "large-agent" form of this is obtained by letting $N \to \infty$, in which case $\sigma_{\overline{\epsilon}}^2 = \sigma_\iota^2$ and the private noise has no impact; further setting $\sigma_\iota = 0$ and $\theta = 0$, we obtain the framework of CG.

Corollary 1. *The regression of consensus forecast error on forecast revision in the DKF framework can be expressed as*

$$\overline{w}_t = \frac{1-G}{G}\,\overline{x}_t + \nu_t^h - \rho^h((1+\theta)\overline{\epsilon}_t - \rho\theta\overline{\epsilon}_{t-1} + \theta u_t), \tag{5}$$

i.e., the regression error is $e_t = \nu_t^h - \rho^h((1+\theta)\overline{\epsilon}_t - \rho\theta\overline{\epsilon}_{t-1} + \theta u_t$. *Also,* $\overline{\mathbf{f}}_t = N^{-1}\sum_{i=1}^N \mathbf{f}_t^i$ *is a stationary time series with contemporaneous covariance matrix*

$$Var[\overline{\mathbf{f}}_t] = diag\{\sigma_{\overline{\epsilon}}^2, \overline{P}, \rho^2(1-G)^2(\overline{P} + \sigma_{\overline{\epsilon}}^2)\},$$

and the forecast revision covariate can be expressed as $\overline{x}_t = \boldsymbol{\tau}'\overline{\mathbf{f}}_t$*; the regression error is* $\nu_t^h + \boldsymbol{\gamma}'\overline{\mathbf{f}}_t$. *The asymptotic bias of* $\overline{\beta}$ *is*

$$\frac{\boldsymbol{\tau}'\,Var[\overline{\mathbf{f}}_t]\boldsymbol{\gamma}}{\boldsymbol{\tau}'\,Var[\overline{\mathbf{f}}_t]\boldsymbol{\tau}} = -\frac{(1+\theta)(1+\theta-\overline{G}) + \theta^2\rho^2(1-G)}{G((1+\theta)^2 + \theta^2\rho^2)}.$$

Remark 2. *The asymptotic bias for the consensus regression can be re-expressed as*

$$-\frac{(1-G)}{G} - \frac{(1+\theta)(1+\theta-\overline{G}/G)}{(1+\theta)^2 + \theta^2\rho^2},$$

which in the rational case (with $\theta = 0$*) is* $-(1-G)/G - (1-\overline{G}/G)$*. If* $N = 1$ *we have* $\overline{G} = G$ *(since* $\sigma_{\overline{\epsilon}}^2 = \sigma_\epsilon^2$ *in this case) and the result is the same as that of Theorem 1. In contrast, the large-agent case (letting* $N \to \infty$*) yields* $\sigma_{\overline{\epsilon}}^2 = \sigma_\iota^2$ *so that only public noise is featured in* $\overline{G}$*; hence in the CG case with* $\sigma_\iota = 0$*, we find* $\overline{G} = 1$ *and the asymptotic bias is zero. More generally, when* $\theta > 0$ *the bias is negative.*

In summary, the "estimate then average" approach yields a bias described in Theorem 1 and Remark 1, whereas the "average then estimate" approach yields a bias given in Corollary 1 and Remark 2.

2.2 A Likelihood Ratio Test for Forecast Efficiency

In view of these asymptotic biases for the relative information rigidity, it is of interest to use an alternative unbiased estimator. A natural idea is to use instrumental variables. We seek new variables z_t^i that are uncorrelated with the regression error e_t^i of equation (4), and yet have non-trivial correlation with the covariates; also, these instrumental variables must be computable (available to us) at time t. Let $v_t^i = \pi_t - \pi_{t|t-1}^{i,\theta}$ be the one-step ahead forecast error from the DKF, which will be known to us at time $t+h$ (when we investigate the regression of forecast error on forecast revision); then we propose the instrumental variables

$$z_t^i = \begin{cases} v_t^i & \text{if } \theta = 0 \\ v_{t-1}^i & \text{if } \theta > 0 \end{cases} . \tag{6}$$

Note that the rational ($\theta = 0$) and irrational ($\theta > 0$) cases require different treatment.

Theorem 2. *In the DKF framework, the instrumental variables z_t^i defined in (6) are asymptotically uncorrelated with the regression error e_t^i of equation (4), and provide a consistent estimate of relative information rigidity $\beta = (1 - G)/G$.*

It follows from the preceding result that the bias in the "estimate then average" approach can be fixed by using the instrumental variables defined in Theorem 2, followed by averaging the resulting parameter estimates. In contrast, if we "average then estimate" then different instrumental variables are needed; it turns out that we can use the average $\overline{z}_t = N^{-1} \sum_{i=1}^{N} z_t^i$, as the following result demonstrates.

Corollary 2. *In the DKF framework, the consensus instrumental variables* $\overline{z}_t = N^{-1} \sum_{i=1}^{N} z_t^i$ *are asymptotically uncorrelated with the consensus regression error* e_t *of equation (5), and provide a consistent estimate of relative information rigidity* $\beta = (1 - G)/G$.

To summarize, in equation (6), we define the instrumental variable through lagged forecast errors. If $\theta = 0$, we can use lag 1 or higher; but if $\theta > 0$, we must use lag 2 or higher. Since we don't know a priori the value of θ, in practice we recommend taking the instrumental variable to be the two-period lagged forecast errors.

A practical concern with the use of instrumental variables is their strength. According to Stock, et al. (2002), the first-stage F-statistic (for regression of forecast revision on the instrumental variables) should be larger than 10 in order for inference to be reliable. Moreira (2003) further discusses the issue of weak instrumental variables, and proposes a likelihood ratio test for the null hypothesis (i.e., $\beta = 0$) that is robust (i.e., still correctly sized) to the presence of weak instrumental variables. In our context, because there is only one instrumental variable, the likelihood ratio test of Moreira (2003) reduces to the test of Anderson and Rubin (1949), which is asymptotically χ^2 with one degree of freedom. This

Anderson-Rubin statistic has the formula

$$\frac{\sum_t \overline{z}_t \overline{w}_t}{\sqrt{\sum_t \overline{z}_t^2 \, \widehat{\Omega}}} \tag{7}$$

in the consensus case, where $\widehat{\Omega} = (\sum_t \overline{w}_t^2 - (\sum_t \overline{z}_t \overline{w}_t)^2 / \sum_t \overline{z}_t^2)/(n-3)$.[2] To compute the Anderson-Rubin statistic for the individual case, simply substitute z_t^i and w_t^i for $\overline{z}_t$ and $\overline{w}_t$ in (7) and the formula for $\widehat{\Omega}$.

The Anderson-Rubin statistic can be rendered more robust to heteroscedasticity by implementing a jackknifed version, as proposed in Mikusheva and Sun (2022). Setting $P = \overline{z}_t \overline{z}_t' / (\overline{z}_t' \overline{z}_t)$, the Mikusheva-Sun test statistic is given by equation (2) of their paper, which in our context with a single instrumental variable takes the form

$$\frac{\sum_{t \neq s} P_{ts} \overline{z}_t \overline{z}_s}{\sqrt{2 \sum_{t \neq s} P_{ts}^2 \overline{z}_t^2 \overline{z}_s^2}}. \tag{8}$$

We choose to use in the denominator the basic variance estimate of Section 4.1 of Mikusheva and Sun (2022), as this estimator is guaranteed to be positive. Straightforward modifications of (8) yield the individual case.

3 Simulations

We investigate the performance of the Anderson-Rubin statistic (7) based on both consensus and individual forecasts, through simulations. Since our simulations involve homoscedastic errors, we do not here study the Mikusheva-Sun statistic (8), although this statistic is applied in our data

[2] Although the sample size is n, the common set of observations for forecast errors and the instrumental variables has size $n-2$, and we subtract 1 for the number of instrumental variables to obtain $n-3$ in the formula for $\widehat{\Omega}$.

analysis in Section 4.

We explore the Gaussian process defined in Section 2 with the following parameter settings: N is either 10 or 50; $\sigma_u = 1$, $\sigma_\epsilon^2 = \sigma_u^2 \cdot \text{NSR}$, and $\sigma_\iota = \sigma_\eta = \sigma_\epsilon/\sqrt{2}$, where the noise-to-signal ratio (NSR) is either 0 (non-existent), .8 (low), or 1.6 (high). The null hypothesis of $\beta = 0$ occurs if and only if $\sigma_\epsilon = 0$, in which case the signal is observed without error (by all agents); this corresponds to NSR $= 0$. The persistence is either low ($\rho = .85$) or high ($\rho = .95$), and we consider three values of θ: 0, .5, or .9, corresponding to completely rational, moderate irrationality, and substantial irrationality relative to FIRE. We also consider a fourth setting, where θ is drawn randomly for each agent, independently from a uniform distribution on $[.3, 1.5]$. We consider samples of size $T = 500$ (Tables 1 and 2) and 1,000 (Tables B.1 and B.2 of Appendix B); size and power results are based on 10,000 Monte Carlo replications.

In the DKF framework, the weight on $\pi^i_{t|t-1}$ in (1) is $1 - (1 + \theta)G$; BGMS assumes in their Proposition 2 that this quantity is positive − hence we require that $\theta < (1-G)/G = \beta$, i.e., rationality is upper bounded by relative information rigidity. We also enforce this requirement in our simulations; note that this condition is violated for $\theta = .9$ if $\rho = .85$ (for either the low or high NSR), so these cases are excluded from our simulations. In the fourth setting, we increase NSR to 4.0 to guarantee that $\theta < \beta$ in all cases.

We are interested in the size and power of the Anderson-Rubin statistic used to test the null hypothesis ($\beta = 0$) of forecast efficiency against the alternative hypothesis that $\beta \neq 0$. In each simulation study, we compute the first-stage F-statistic for the consensus instrumental variables, obtained by the regression of consensus forecast revision on the lagged consensus forecast error. We also compute the consensus Anderson-Rubin statistic, obtaining a single statistic using the aggregated variables.

In contrast, for each individual case we obtain an Anderson-Rubin statistic and p-value. These p-values can be combined according to Simes (1986), as discussed in Sheng and Yang (2011), viz. the sorted p-values $p_{(i)}$ are compared to $i\alpha/N$ (for $1 \leq i \leq N$) and some given Type I error rate α, and each null hypothesis that β is zero is rejected if at least one $p_{(i)}$ exceeds the threshold. Equivalently, there is failure to reject if $\min_{1\leq i\leq N} p_{(i)}N/i > \alpha$, and the composite p-value can be taken to be the largest α such that one still fails to reject. Hence we report $\min_{1\leq i\leq N} p_{(i)}N/i$ as the composite p-value.

Table 1: Test for forecast efficiency with $N = 10$ and $T = 500$

ρ	NSR	θ	CONS IV F-stat	CONS IV AR-stat	IND IV AR-stats
.85	0	0	.0521	.0496	.0496
	.8	0	1	.5901	.5926
		.5	.0777	.0600	.0439
	1.6	0	1	.9679	.9819
		.5	.6626	.5218	.5612
		.9	.0618	.0702	.0629
	4.0	random	.0895	.0935	.4802
.95	0	0	.0512	.0470	.0470
	.8	0	1	.6906	.7299
		.5	.0626	.0535	.0427
	1.6	0	1	.9889	.9976
		.5	.6104	.4963	.5567
		.9	.0501	.0510	.0456
	4.0	random	.5123	.8674	.9975

Note: rejection rates are provided by process $\rho = .85, .95$, NSR $= 0, .8, 1.6, 4.0$, and $\theta = 0, .5, .9$, for the Consensus IV first-stage F-statistic (column 4), the Consensus IV Anderson-Rubin (AR) statistic (column 5), and the combined Individual IV Anderson-Rubin (AR) statistic (column 6). Results are based on 10,000 simulations of length 500, with 10 agents.

On the basis of these test statistics and p-values, we determine the proportion of rejections of the null hypothesis, reported in Tables 1 and 2 (as well as Tables B.1 and B.2 of Appendix B). Note that the NSR = 0 case corresponds to these null hypotheses, so the corresponding rows yield information on the size of tests. Both the Anderson-Rubin statistics and the first-stage F-statistics have very good empirical size, which is quite close to the nominal size at 5%.

Table 2: Test for forecast efficiency $N = 50$ and $T = 500$

ρ	NSR	θ	CONS IV F-stat	CONS IV AR-stat	IND IV AR-stats
.85	0	0	.0486	.0497	.0497
	.8	0	1	.5889	.5591
		.5	.0834	.0593	.0368
	1.6	0	1	.9675	.9795
		.5	.6864	.5114	.5457
		.9	.0664	.0702	.0569
	4.0	random	.6433	.9487	.9998
.95	0	0	.0458	.0520	.0520
	.8	0	1	.6853	.7031
		.5	.0565	.0540	.0348
	1.6	0	1	.9873	.9971
		.5	.6275	.5091	.5827
		.9	.0489	.0514	.0390
	4.0	random	.5204	.8939	1

Note: rejection rates are provided by process $\rho = .85, .95$, NSR $= 0, .8, 1.6, 4.0$, and $\theta = 0, .5, .9$, for the Consensus IV first-stage F-statistic (column 4), the Consensus IV Anderson-Rubin (AR) statistic (column 5), and the combined Individual IV Anderson-Rubin (AR) statistic (column 6). Results are based on 10,000 simulations of length 500, with 50 agents.

Other values of NSR yield power. The power decreases with θ, the representativeness parameter governing the extent to which beliefs depart from rational updating. When $\theta = 0$, both the Anderson-Rubin statistics and the first-stage F-statistics have satisfactory power; however, the power deteriorates as θ increases. When $\theta = 0.9$, the power of both statistics is very low. As expected, the power increases significantly with T going from 500 (Tables 1 and 2) to 1,000 (Tables B.1 and B.2). The power gain is marginal as N increases from 10 (Tables 1 and B.1) to 50 (Tables 2 and B.2).

We are also interested in the impact of missing data – which is quite common in forecasting datasets – upon the power of tests. In this simulation, we focus on the particular setting that $\rho = 0.85$ and NSR $= 0.8$, with $\theta = 0$, sample size $T = 500$, and $N = 10$ agents. We introduce a proportion of missing values (denoted as m) into both the forecast errors and the forecast revisions (so these variables can have different NA patterns, obtained at random for each simulation) using Bernoulli random variables, varying the proportion m in $\{0, 0.2, 0.5\}$. Then missing values are imputed using a simplistic approach: for each agent's time series, all NA values are replaced by the sample mean over the available values. As shown in Table 3, the power of the first-stage F-statistic and the Anderson-Rubin statistic based on consensus forecasts remains almost the same, while the power of the Anderson-Rubin statistic based on individual forecasts moderately decreases with increasing m.

Table 3: Test for forecast efficiency: The impact of missing values

m	CONS IV F-stat	CONS IV AR-stat	IND IV AR-stats
0	1	.5779	.5833
0.2	.9999	.5827	.5695
0.5	.9998	.5826	.4635

Note: rejection rates are provided for process $\rho = .85$, NSR $= .8$ with $\theta = 0$, for the Consensus IV first-stage F-statistic (column 2), the Consensus IV Anderson-Rubin (AR) statistic (column 3), and the combined Individual IV Anderson-Rubin (AR) statistic (column 4), by the proportion m of both the dependent and independent variables are missing at random. Results are based on 10,000 simulations of length 500, with 10 agents.

4 Efficiency in Macroeconomic Expectations

This section provides an analysis of the individual and consensus forecasts from the Survey of Professional Forecasters (SPF) currently run by the Federal Reserve Bank of Philadelphia covering 1968.Q4 through 2016.Q4.[3] To facilitate the comparison to BGMS (2020), we use an annual forecast horizon. For GDP and inflation, we transform the level of these variables into implied growth rates from quarter $t-1$ to quarter $t+3$. For variables such as unemployment rate and interest rates, we study the level in quarter $t+3$. We compute consensus forecasts as the mean of individual forecasts.

The original data set has 333 agents and 193 quarters. About 90% of the data is missing, and the time points that are missing for each agent differ greatly. It is infeasible to merely omit times for which any agents have an NA, as this would remove all of the data; therefore we use the following techniques to refine the data. First, we split the data into two distinct time spans: (1) 1968.Q4 through 1990.Q4 and (2) 1992.Q1 through 2016.Q4. The American Statistical Association and the National Bureau of Economic Research initiated the survey in 1968Q4. Due to a rapidly declining participation rate in the late 1980s, the Federal Reserve Bank of Philadelphia took over the survey in 1990 with a new infusion of forecasters. Our decision for sample split is driven by this structural change in the survey. Also for this first span (called panel A), we restrict to the six variables: nominal GDP, real GDP, GDP price index, industrial production, housing starts, and unemployment. For the second span

[3] Federal Reserve Bank of Philadelphia, Survey of Professional Forecasters: 1968.Q4 - 2016.Q4, `https://www.philadelphiafed.org/surveys-and-data/real-time-data-research/individual-forecasts`

(called panel B) we retain all 15 pairs of variables. In addition to the six variables above, we also consider forecasts for consumer price index, real consumption, real nonresidential investment, real residential investment, real federal government consumption, real state and local government consumption, three-month Treasury rate, ten-year Treasury rate, and AAA corporate bond rate.

Further, within each of these two data subsets we consider those agents who have responded at least 20% of the time. This gives $N = 38$ agents and $T = 89$ quarters for the first subset and $N = 36$ agents and $T = 100$ quarters for the second subset. The remaining missing values are imputed for each variable and agent by using the temporal average (described in Section 3).

Individual forecast errors are calculated as actual values minus individual forecasts. We use the first-released actual values in real time from the Real Time Data Set for Macroeconomists provided by the Federal Reserve Bank of Philadelphia; see Croushore and Stark (2001).[4] Individual forecast revisions are computed as individual i's forecasts made in quarter t minus her forecasts made in quarter $t-1$. We obtain consensus forecast errors and forecast revisions as the average of the corresponding individual forecast errors and revisions. We take the instrumental variables to be the two-period lagged forecast errors.

Table 4 presents the results based on consensus forecasts. We apply the Anderson-Rubin test statistic; but since we have observed some heteroscedasticity in the regression residuals for certain cases, we also apply

[4]Federal Reserve Bank of Philadelphia, Real-Time Data Set for Macroeconomists: 1968.Q4 - 2016.Q4, `https://www.philadelphiafed.org/surveys-and-data/real-time-data-research/real-time-data-set-for-macroeconomists`

Table 4: Test for forecast efficiency in consensus forecast

(1) Variable	(2) AR-stat	(3) MS-stat	(4) F-stat	(5) IV	(6) OLS
Panel A: 1968.Q4-1990.Q4					
NGDP	6.47	7.04	17.83	2.15	−0.11
RGDP	15.08	4.56	16.62	3.56	0.61
PGDP	60.06	62.16	79.32	3.93	2.02
INDP	3.25	0.54	6.89	NA	0.59
HOUS	42.14	135.46	8.90	NA	0.59
UNEM	11.38	29.04	65.09	1.93	0.37
Panel B: 1992.Q1-2016.Q4					
NGDP	11.91	7.84	118.61	1.27	0.72
RGDP	10.39	7.26	83.00	1.34	0.58
PGDP	51.78	330.25	22.33	4.65	0.65
CPI	3.03	0.01	75.94	NA	0.22
RCON	43.11	27.31	54.79	2.85	0.45
INDP	17.72	9.27	168.57	2.30	0.78
RNRE	24.76	26.11	110.89	2.26	1.43
RRES	76.76	367.73	92.32	3.84	1.51
RFGC	46.63	205.5	21.46	5.85	0.18
RGSL	37.35	169.49	15.21	6.04	0.87
HOUS	137.46	325.61	5.75	NA	0.35
UNEM	70.97	24.38	233.55	1.85	1.26
tb3m	74.71	122.93	142.59	2.07	0.92
tn10y	7.38	17.14	69.36	0.93	−0.18
AAA	11.42	38.02	51.58	1.32	−0.28

Note: Column 2 gives the consensus Anderson-Rubin statistic. Column 3 gives the consensus Mukisheva-Sun statistic. Column 4 gives the first-stage F-statistic for the regression of consensus forecast revision on the consensus IV. Column 5 provides the consensus IV estimate of β ("average then estimate"). Column 6 provides the consensus OLS estimate of β. Data source: Philadelphia Fed Survey of Professional Forecasters (SPF) forecasts for nominal GDP (NGDP), real GDP (RGDP), GDP price index (PGDP), industrial production (INDP), housing start (HOUS), unemployment (UNEM), consumer price index (CPI), real consumption (RCON), real nonresidential investment (RNRE), real residential investment (RRES), real federal government consumption (RFGC), real state and local government consumption (RGSL), three-month Treasury rate (tb3m), ten-year Treasury rate (tn10y) and AAA corporate bond rate (AAA).

the Mikusheva-Sun test statistic. For all cases except industrial production in panel A and consumer price index in panel B, the Anderson-Rubin test statistic (column 2) exceeds 3.84, the critical value of the chi-squared distribution with 1 degree of freedom at the 5% significance level. Also for these variables, the Mukisheva-Sun test statistic (column 3) exceeds the standard normal critical value. This suggests consensus forecast inefficiency when using public information. According to the first-stage F statistic reported in column 4, consensus instrumental variables (i.e., the two-period lagged consensus forecast errors) are strong for all cases, except for three instances (industrial production and housing start in panel A, and housing start in panel B).

For the cases with strong instrumental variables, we report the two-stage least square coefficient estimates in column 5, revealing that they are positive and statistically significant. This points to the underreaction of consensus forecasts to news relative to FIRE. The level of underreaction is considerably more pronounced than what the potentially biased OLS estimates in column 6 suggest. Our two-stage least square estimates yield an implied Kalman gain of 0.31, resulting in an average degree of information rigidity of 0.69. It is worth noting that the implied degree of information rigidity is higher than that found in CG (2015), reflecting the presence of common noise and, therefore, a downward bias in their estimate of information rigidity.[5]

Table 5 presents results based on individual forecasts. In column 2, the combined p-values from the Anderson-Rubin test statistics are all below 1%, indicating significant forecast inefficiency when using public information at the individual forecaster level. The results in column 3, the p-values from the Mukisheva-Sun test statistics, are the same. As shown in column 4, the range of first-stage F-statistics, from the regression of individual forecast revision on individual instrumental variable (i.e., the

[5] Using the SPF inflation forecasts, CG (2015)'s estimate yields an implied Kalman gain of 0.42 and therefore a degree of information rigidity of 0.58.

two-period lagged individual forecast error) for each agent, indicates a large proportion of weak instruments. It is worth noting that the p-values of the Anderson-Rubin and Mukisheva-Sun test statistics remain valid even in the presence of weak instrumental variables.

In column 5, we present the two-stage least square coefficient estimates for cases with strong instrumental variables, revealing that they are both positive and statistically significant. This indicates that individual forecasts exhibit an underreaction to news relative to FIRE. These estimates, based on individual forecasts, imply an average degree of information rigidity of 0.63, which is lower than that based on consensus forecasts in the previous table (i.e. 0.69). Nonetheless, this finding provides pervasive evidence consistent with the presence of information rigidities at the individual forecaster level, in contrast to the overreaction finding reported in the BGMS (2020) study.

This result differs from the inference drawn from the OLS estimates in column 6. According to the OLS estimates, it appears that individual forecasters tend to overreact to public information. In fact, a majority of the estimates (14 out of 21 cases) exhibit a negative sign.

5 Conclusion

This paper delves into the issue of regressing forecast errors on forecast revisions, both at the individual and aggregate levels, which is a specific aspect of forecast efficiency. Coibion and Gorodnichenko (2015) propose an approach that involves aggregating forecasts and estimating the aggregated time series regression – a method known as "average then estimate."

Table 5: Test for forecast efficiency in individual forecasts

(1) Variable	(2) AR p-value	(3) MS p-value	(4) F-stat Range	(5) IV	(6) OLS
Panel A: 1968.Q4-1990.Q4					
NGDP	$<.01$	$<.01$	$[0.05, 61.61]$	0.88	-0.24
RGDP	$<.01$	$<.01$	$[0.00, 32.96]$	1.44	-0.04
PGDP	$<.01$	$<.01$	$[0.03, 48.30]$	3.00	0.22
INDP	$<.01$	$<.01$	$[0.00, 24.70]$	0.94	-0.26
HOUS	$<.01$	$<.01$	$[0.00, 8.54]$	NA	-0.11
UNEM	$<.01$	$<.01$	$[0.08, 92.24]$	1.33	-0.03
Panel B: 1992.Q1-2016.Q4					
NGDP	$<.01$	$<.01$	$[0.00, 213.34]$	1.28	-0.01
RGDP	$<.01$	$<.01$	$[0.02, 128.56]$	1.33	0.02
PGDP	$<.01$	$<.01$	$[0.06, 55.61]$	2.38	-0.31
CPI	$<.01$	$<.01$	$[0.27, 73.27]$	1.07	-0.25
RCON	$<.01$	$<.01$	$[0.33, 100.50]$	2.23	-0.09
INDP	$<.01$	$<.01$	$[0.14, 189.30]$	2.20	0.01
RNRE	$<.01$	$<.01$	$[0.09, 66.52]$	2.02	0.21
RRES	$<.01$	$<.01$	$[0.01, 44.85]$	3.42	0.17
RFGC	$<.01$	$<.01$	$[0.20, 26.19]$	3.42	-0.40
RGSL	$<.01$	$<.01$	$[0.00, 55.21]$	3.94	-0.29
HOUS	$<.01$	$<.01$	$[0.00, 29.68]$	3.13	-0.24
UNEM	$<.01$	$<.01$	$[0.00, 243.83]$	1.75	0.58
tb3m	$<.01$	$<.01$	$[2.30, 105.28]$	1.84	0.40
tn10y	$<.01$	$<.01$	$[1.14, 71.90]$	0.88	-0.23
AAA	$<.01$	$<.01$	$[0.00, 44.10]$	1.06	-0.32

Note: Column 2 combines the p-values for the individual Anderson-Rubin statistics. Column 3 combines the p-values for the individual Mukisheva-Sun statistics. Column 4 gives the range of first-stage F-statistics by agent, for the regression of individual forecast revision on the two-period lagged individual forecast error. Column 5 provides the average of the individual IV estimators for those cases where the first-stage F-statistic is above 10. NA values correspond to cases of a weak IV, where the first-stage F-statistic is below 10. Column 6 is the average of the individual OLS estimates of β. Data source: Philadelphia Fed Survey of Professional Forecasters (SPF) forecasts for nominal GDP (NGDP), real GDP (RGDP), GDP price index (PGDP), industrial production (INDP), housing start (HOUS), unemployment (UNEM), consumer price index (CPI), real consumption (RCON), real nonresidential investment (RNRE), real residential investment (RRES), real federal government consumption (RFGC), real state and local government consumption (RGSL), three-month Treasury rate (tb3m), ten-year Treasury rate (tn10y) and AAA corporate bond rate (AAA).

On the other hand, Bordalo, Gennaioli, Ma, and Shleifer (2020) suggest a different approach, which involves estimating the average relationship by running separate regressions for each individual and then aggregating the results, known as the "estimate then average" approach. We demonstrate that in the presence of public noise, both of these estimators exhibit asymptotic bias. Both the "average then estimate" and "estimate then average" approaches have bias that is dependent on the Kalman gain, the representativeness parameter θ (which governs the degree to which beliefs deviate from rational updating), and the persistence of the underlying signal. When $\theta = 0$, in the large-agent case the asymptotic bias of the "average then estimate" approach does not depend upon private noise, and therefore is unbiased if there is no public noise.

To address these biases, we propose instrumenting forecast revisions with past forecast errors. The choice of lags required for instrumentation varies with the value of θ. When $\theta = 0$, the instrumental variable is the one-period lagged forecast error. However, if $\theta > 0$, we need to use two lags or higher. Given that we typically don't have prior knowledge of the exact value of θ, in practical applications, we recommend employing the two-period lagged forecast errors as the instrumental variable.

We propose utilizing the Anderson-Rubin test for assessing forecast efficiency. This test has been demonstrated to be robust in the presence of weak instrumental variables, as established by Moreira (2003). In Monte Carlo simulations, we observe that the Anderson-Rubin statistics exhibit excellent empirical size, closely aligning with the nominal size of 5%. Furthermore, the test demonstrates satisfactory power when θ is either 0 or takes very small values.

When testing forecast efficiency using data from the U.S. Survey of Professional Forecasters dataset, two distinct approaches yield contrasting results. The "average then estimate" approach provides positive but potentially biased slope coefficient estimates. This suggests that consensus forecasts tend to underreact to new information compared to the full-information rational expectations. On the other hand, the “estimate then average” approach, using individual forecasts, yields a majority of negative estimates. This indicates overreaction to news at the individual forecaster level.

In line with these findings, the Anderson-Rubin test and its jackknifed version (which is robust to heteroscedasticity, as proposed in Mikusheva and Sun (2022)) strongly reject the null hypothesis that forecasters efficiently utilize information for almost all consensus forecasts and individual forecasts. For consensus forecasts, our two-stage least square coefficient estimates are both positive and statistically significant, signifying an underreaction to news. It is noteworthy that this level of underreaction is more pronounced than what the potentially biased OLS estimates suggest.

What’s even more intriguing and unexpected is the result that individual forecasters also underreact to news. The two-stage least square coefficient estimates are positive and statistically significant in this case. This finding sharply contradicts the inference drawn from the OLS estimates, which implies that individual forecasters tend to overreact to public information.

In conclusion, professional forecasters both at the individual and aggregate levels tend to underreact to new information. This resolves the puzzle in the literature regarding the underreaction of consensus forecasts and the overreaction of individual forecasts.

References

Anderson, T.W. and H. Rubin (1949). Estimation of the parameters of a single equation in a complete system of stochastic equations. *The Annals of Mathematical Statistics* 20(1), 46-63.

Baker, S.R., T.S. McElroy, and X.S. Sheng (2020). Expectation formation following large unexpected shocks. *Review of Economics and Statistics* 102 (2), 287-303.

Baltagi, B.H., J.M. Griffin, and W. Xiong (2000). To pool or not to pool: Homogeneous versus heterogeneous estimations applied to cigarette demand. *Review of Economics and Statistics* 82, 117-126.

Binder, C. (2017). Consumer forecast revisions: Is information really so sticky? *Economics Letters* 161, 112-115.

Bonham, C.S. and R.H. Cohen (2001). To aggregate, pool, or neither. *Journal of Business and Economic Statistics* 19(3), 278-291.

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

Bürgi, C. and J.L. Ortiz (2022). Overreaction through anchoring. *CESifo Working Paper* No. 10193.

Chang, A.C. and T.J. Levinson (2023). Raiders of the lost high-frequency forecasts: New data and evidence on the efficiency of the Fed's forecasting. *Journal of Applied Econometrics* 38(1), 88-104.

Coibion, O. and Y. Gorodnichenko (2015). Information rigidity and the expectations formation process: A simple framework and new facts. *American Economic Review* 105(8), 2644-2678.

Croushore, D. and T. Stark (2001). A real-time data set for macroeconomists. *Journal of Econometrics* 105, 111-130.

Drager, L. and M.J. Lamla (2017). Imperfect information and inflation expectations: Evidence from microdata. *Oxford Bulletin of Economics and Statistics* 79, 933-968.

Fuhrer, J.C. (2018). Intrinsic expectations persistence: Evidence from professional and household survey expectations. *FRB of Boston Working Paper* No. 18-9.

Giacomini, R., V. Skreta, and J. Turen (2020). Heterogeneity, inattention and Bayesian updates. *American Economic Journal: Macroeconomics* 12, 282-309.

Lahiri, K. and X. Sheng (2008). Evolution of forecast disagreement in a Bayesian learning model. *Journal of Econometrics* 144, 325-340.

Messina, J.D., T.M. Sinclair, and H.O. Stekler (2015). What can we learn from revisions to the Greenbook forecasts? *Journal of Macroeconomics* 45, 54-62.

Mikusheva, A., and L. Sun (2022). Inference with many weak instruments. *Review of Economic Studies* 89(5), 2663-2686.

Moreira, M.J. (2003). A conditional likelihood ratio test for structural models. *Econometrica* 71(4), 1027-1048.

Nordhaus, W.D. (1987). Forecasting efficiency, concepts, and applications. *Review of Economics and Statistics* 69, 667-674.

Pesaran, M.H. and Q. Zhou (2018). To pool or not to pool: Revisited. *Oxford Bulletin of Economics and Statistics* 80(2), 185-217.

Ryngaert, J. (2023). What do (and don't) forecasters know about U.S. inflation? Forthcoming in *Journal of Money, Credit and Banking.*

Sheng, X., and Yang, J. (2013). Truncated product methods for panel unit root tests. *Oxford Bulletin of Economics and Statistics* 75, 624-636.

Simes, R.J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika* 73(3), 751-754.

Stock, J.H., J.H. Wright and M. Yogo (2002). A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business & Economic Statistics* 20(4), 518-529.

Appendix A Proofs

A.1 Proof of Theorem 1

First, from the autoregressive definition of $\{\pi_t\}$ we obtain

$$\pi_{t+h} = \rho^h \pi_t + \nu_t^h.$$

Next, (1) together with $\pi^i_{t|t} = \pi^i_{t|t-1} + G(y^i_t - \pi^i_{t|t-1})$ (the updating formula for the rational Kalman Filter) implies that

$$\pi^{i,\theta}_{t|t} = (1+\theta)\pi^i_{t|t} - \theta\pi^i_{t|t-1}.$$

Hence the forecast error is

$$w^i_t = \rho^h(\pi_t - (1+\theta)\,\pi^i_{t|t} + \theta\pi^i_{t|t-1}) + \nu^h_t,$$

using $\pi^{i,\theta}_{t+h|t} = \rho^h \pi^{i,\theta}_{t|t}$. For the forecast revision, we can use the above calculations to obtain

$$\begin{aligned}
\pi^{i,\theta}_{t+h|t} &= \rho^h \left((1+\theta)((1-G)\pi^i_{t|t-1} + Gy^i_t) - \theta\pi^i_{t|t-1}\right) \\
&= \rho^h \left(((1+\theta)(1-G) - \theta)\pi^i_{t|t-1} + (1+\theta)Gy^i_t\right).
\end{aligned}$$

Thus

$$\begin{aligned}
\pi^{i,\theta}_{t+h|t-1} &= \rho^{h+1}\pi^{i,\theta}_{t-1|t-1} \\
&= \rho^{h+1}\left(\pi^i_{t-1|t-2} + (1+\theta)G(y^i_{t-1} - \pi^i_{t-1|t-2})\right) \\
&= \rho^{h+1}\left((1-(1+\theta)G)\pi^i_{t-1|t-2} + (1+\theta)Gy^i_{t-1}\right),
\end{aligned}$$

from which it follows – using $(1-G)\pi^i_{t-1|t-2} = \pi^i_{t-1|t-1} - Gy^i_{t-1}$ from the rational Kalman Filter – that

$$
\begin{aligned}
(1-G)\pi^{i,\theta}_{t+h|t-1} &= \rho^{h+1}\left((1-(1+\theta)G)(\pi^i_{t-1|t-1} - Gy^i_{t-1}) + (1-G)(1+\theta)Gy^i_{t-1}\right) \\
&= \rho^{h+1}\left((1-(1+\theta)G)\pi^i_{t-1|t-1} + \theta G y^i_{t-1}\right) \\
&= \rho^{h}\left((1-(1+\theta)G)\pi^i_{t|t-1} + \rho\theta G y^i_{t-1}\right).
\end{aligned}
$$

The $h = 0$ case of this is

$$
\pi^{i,\theta}_{t|t-1} = \frac{(1-(1+\theta)G)\pi^i_{t|t-1} + \rho\theta G y^i_{t-1}}{1-G}. \tag{A.1}
$$

From this result we obtain

$$
\begin{aligned}
\pi^{i,\theta}_{t+h|t} - (1-G)\pi^{i,\theta}_{t+h|t-1} &= \rho^h((1+\theta)Gy^i_t - \rho\theta G y^i_{t-1}) \\
&= \rho^h((1+\theta)G\epsilon^i_t + G\pi_t + \theta G u_t - \rho\theta G\epsilon^i_{t-1}),
\end{aligned}
$$

using

$$
\rho y^i_{t-1} = \pi_t - u_t + \rho\epsilon^i_{t-1}. \tag{A.2}
$$

Now multiplying the forecast error by G, we obtain

$$
\begin{aligned}
Gw_t^i &= \rho^h G(\pi_t - (1+\theta)\pi_{t|t}^i + \theta\pi_{t|t-1}^i) + G\nu_t^h \\
&= \pi_{t+h|t}^{i,\theta} - (1-G)\pi_{t+h|t-1}^{i,\theta} + G\nu_t^h \\
&- \rho^h G((1+\theta)\pi_{t|t}^i - \theta\pi_{t|t-1}^i) \\
&- \rho^h G((1+\theta)G\epsilon_t^i + \theta u_t - \rho\theta\epsilon_{t-1}^i) \\
&= (1-G)(\pi_{t+h|t}^{i,\theta} - \pi_{t+h|t-1}^{i,\theta}) + G\nu_t^h \\
&- \rho^h G((1+\theta)G\epsilon_t^i - \rho\theta\epsilon_{t-1}^i + \theta u_t).
\end{aligned}
$$

Finally, dividing by G yields (4).

Turning to $\mathbf{f}_t^i$, the stationarity is immediate; also ϵ_t^i is uncorrelated with the $\{\pi_t\}$ process as well as past values $y_{t-1}^i, y_{t-2}^i, \ldots$, and hence is uncorrelated with the third component, which is a linear combination of such random variables. Moreover, the second component is a prediction error, and therefore is uncorrelated with past data, and hence is uncorrelated with the third component. This third component can be rewritten as

$$
\pi_{t|t-1}^i - \rho y_{t-1}^i = \rho(\pi_{t-1|t-1}^i - y_{t-1}^i) = \rho(1-G)(\pi_{t-1|t-2}^i - \pi_{t-1} - \epsilon_{t-1}^i), \quad \text{(A.3)}
$$

which has variance $\rho^2 G^2(P + \sigma_\epsilon^2)$. Next, using

$$
u_t - \rho\epsilon_{t-1}^i = \pi_t - \rho y_{t-1}^i = [0,\ 1,\ 1]\mathbf{f}_t^i
$$

(from (A.2)), the regression error can be written

$$e_t^i = \nu_t^h - \rho^h((1+\theta)\epsilon_t^i + \theta(\pi_t - \rho y_{t-1}^i)) = \nu_t^h + \gamma' \mathbf{f}_t^i.$$

Next, the covariate can be re-expressed as

$$\begin{aligned}
x_t^i &= \rho^h G(\pi_t - \pi_{t|t-1}^{i,\theta} + (1+\theta)\epsilon_t^i - \rho\theta\epsilon_{t-1}^i + \theta u_t) \\
&= \rho^h G\Big((1+\theta)(\pi_t - \pi_{t|t-1}^i) - \theta\pi_t + \theta/(1-G)(\pi_{t|t-1}^i - \rho G y_{t-1}^i) \\
&+ (1+\theta)\epsilon_t^i - \rho\theta\epsilon_{t-1}^i + \theta u_t\Big) \\
&= \rho^h G\Big((1+\theta)(\pi_t - \pi_{t|t-1}^i) + \theta/(1-G)(\pi_{t|t-1}^i - \rho y_{t-1}^i) + (1+\theta)\epsilon_t^i\Big) \\
&= \tau' \mathbf{f}_t^i,
\end{aligned}$$

where the second equality uses (A.1). Asymptotically, the bias in an OLS regresssion is given by the covariance of x_t^i with e_t^i, divided by the variance of x_t^i. Since ν_t^h is uncorrelated with $\mathbf{f}_t^i$, the covariance is $\boldsymbol{\tau}'\mathrm{Var}[\mathbf{f}_t^i]\boldsymbol{\gamma}$; the variance of x_t^i is $\boldsymbol{\tau}'\mathrm{Var}[\mathbf{f}_t^i]\boldsymbol{\tau}$. Simplifying yields the stated expression. □

A.2 Proof of Corollary 1

We can apply averaging over agents to the regression equations used in the proof of Theorem 1, obtaining (5) directly from (4). The consensus noise is

$$\bar{\epsilon}_t = N^{-1}\sum_{i=1}^{N}\epsilon_t^i = \iota_t + N^{-1}\sum_{i=1}^{N}\eta_t^i,$$

which has variance $\sigma_{\bar{\epsilon}}^2 = \sigma_\iota^2 + \sigma_\eta^2/N$. The other calculations follows directly from the proof of Theorem 1. □

A.3 Proof of Theorem 2

Let us define the bivariate time series $\mathbf{g}_t^i = [\epsilon_t^i, \pi_t - \pi_{t|t-1}^i]'$, which is just the first two components of $\mathbf{f}_t^i$, and hence is a stationary process. Using (A.3), we find that

$$\mathbf{f}_t^i = \begin{bmatrix} \mathbf{g}_t^i \\ -\rho(1-G)[1,\,1]\mathbf{g}_{t-1}^i \end{bmatrix}.$$

Thus we can rewrite our covariates and regression errors e_t^i in terms of $\mathbf{g}_t^i$:

$$\begin{aligned} x_t^i &= \rho^h G\left((1+\theta)[1,\,1]\mathbf{g}_t^i - \rho\theta[1,\,1]\mathbf{g}_{t-1}^i\right) \\ e_t^i &= \nu_t^h - \rho^h\left([(1+\theta),\,\theta]\mathbf{g}_t^i - \rho\theta(1-G)[1,\,1]\mathbf{g}_{t-1}^i\right). \end{aligned}$$

Next we compute the autocovariance function $\Gamma(k) = \mathrm{Cov}[\mathbf{g}_t^i, \mathbf{g}_{t-k}^i]$. From our calculations of the variance of $\mathbf{f}_t^i$ we know that $\Gamma(0) = \mathrm{diag}\{\sigma_\epsilon^2, P\}$. Now consider $k = 1$. Write

$$\begin{aligned} \pi_t - \pi_{t|t-1}^i &= \rho(\pi_t - \pi_{t-1|t-1}^i) + u_t \\ &= \rho\left(\pi_{t-1} - (1-G)\pi_{t-1|t-2}^i - Gy_{t-1}^i\right) + u_t \\ &= \rho\left((1-G)(\pi_{t-1} - \pi_{t-1|t-2}^i) - G\epsilon_{t-1}^i\right) + u_t, \end{aligned}$$

so that we obtain

$$\mathrm{Cov}(\pi_t - \pi_{t|t-1}^i, \pi_{t-1} - \pi_{t-1|t-2}^i) = \rho(1-G)P.$$

Moreover, using $\epsilon_{t-1}^i = y_{t-1}^i - \pi_{t-1} = (y_{t-1}^i - \pi_{t-1|t-1}^i) + (\pi_{t-1|t-1}^i - \pi_{t-1})$ and the orthogonality of forecast errors to linear functions of past data, we obtain

$$\mathrm{Cov}(\pi_t - \pi^i_{t|t-1}, \epsilon^i_{t-1}) = \mathrm{Cov}(\pi_t - \pi^i_{t|t-1}, \pi^i_{t-1|t-1} - \pi_{t-1}) = -\rho M,$$

where $M = \mathrm{Var}[\pi_t - \pi^i_{t|t}]$. From prior calculations we know $\pi^i_{t|t} - \pi_t = (1-G)(\pi^i_{t|t-1} - \pi_t) + G\epsilon^i_t$, so that $M = (1-G)^2 P + G^2\sigma^2_\epsilon = P(1-G)$. By similar arguments for $k > 1$, we find that

$$\Gamma(k) = \begin{bmatrix} 0 & 0 \\ -\rho^k(1-G)^k P & \rho(1-G)^k P \end{bmatrix}$$

for $k \geq 1$. Next, we can relate the quantities v^i_t to the bivariate $\mathbf{g}^i_t$ as follows:

$$\begin{aligned} v^i_t &= \pi_t - \pi^i_{t|t-1} + (\pi^i_{t|t-1} - \rho y^i_{t-1})\theta G/(1-G) \\ &= [0,\, 1]\mathbf{g}^i_t - \theta\rho G[1,\, 1]\mathbf{g}^i_{t-1}. \end{aligned}$$

Then we can compute $\mathrm{Cov}[e^i_t, v^i_{t-j}]$ for $j = 0, 1$:

$$\begin{aligned} \mathrm{Cov}[e^i_t, v^i_t] &= -\rho^h \Big([1+\theta, \theta]\Gamma(0)[0,\, 1]' - \rho\theta(1-G)[1,\, 1]\Gamma(-1)[0,\, 1]' \\ &\quad -\rho\theta G[1+\theta,\, \theta]\Gamma(1)[1,\, 1]' + \rho^2\theta^2 G(1-G)[1,\, 1]\Gamma(0)[1,\, 1]'\Big) \\ &= -\rho^h P(\theta + \rho^2\theta^2(1-G)) \\ \mathrm{Cov}[e^i_t, v^i_{t-1}] &= -\rho^h \Big([1+\theta, \theta]\Gamma(1)[0,\, 1]' - \rho\theta(1-G)[1,\, 1]\Gamma(0)[0,\, 1]' \\ &\quad -\rho\theta G[1+\theta,\, \theta]\Gamma(2)[1,\, 1]' + \rho^2\theta^2 G(1-G)[1,\, 1]\Gamma(1)[1,\, 1]'\Big) \\ &= -\rho^h P(\theta\rho P(1-G) - \rho\theta(1-G)P) = 0. \end{aligned}$$

When $\theta = 0$ the instrumental variable z_t^i defined in (6) is v_t^i, and the covariance with e_t^i is zero; but if $\theta > 0$ we instead set the instrumental variable to be any multiple of v_{t-1}^i (we choose this multiple to be one). This shows that the instrumental variable is uncorrelated with the regression error. For the covariate, similar calculations yield

$$
\begin{aligned}
\mathrm{Cov}[x_t^i, v_t^i] &= \rho^h G \big((1+\theta)[1, 1]\Gamma(0)[0, 1]' - \rho\theta[1, 1]\Gamma(-1)[0, 1]' \\
&\quad -\rho\theta G(1+\theta)[1, 1]\Gamma(1)[1, 1]' + \rho^2\theta^2 G[1, 1]\Gamma(0)[1, 1]'\big) \\
&= \rho^h GP(1+\theta+\rho^2\theta^2) \\
\mathrm{Cov}[x_t^i, v_{t-1}^i] &= \rho^h G \big((1+\theta)[1, 1]\Gamma(1)[0, 1]' - \rho\theta[1, 1]\Gamma(0)[0, 1]' \\
&\quad -\rho\theta G(1+\theta)[1, 1]\Gamma(2)[1, 1]' + \rho^2\theta^2 G[1, 1]\Gamma(1)[1, 1]'\big) \\
&= \rho^{h+1} GP((1+\theta)(1-G)-\theta).
\end{aligned}
$$

Hence the covariance of the instrumental variable with the covariate is

$$
\mathrm{Cov}[x_t^i, z_t^i] = \begin{cases} \rho^h GP & \text{if } \theta = 0 \\ \rho^{h+1} GP((1+\theta)(1-G)-\theta) & \text{if } \theta > 0. \end{cases}
$$

Therefore the instrumental variable yields a consistent estimator of the regression parameter. $\square$

A.4 Proof of Corollary 2

By averaging over the agents, we obtain

$$
\bar{\mathbf{f}}_t = N^{-1}\sum_{i=1}^{N} \mathbf{f}_t^i = \begin{bmatrix} \bar{\mathbf{g}}_t \\ -\rho(1-G)[1, 1]\bar{\mathbf{g}}_{t-1} \end{bmatrix}, \qquad \bar{g}_t = \begin{bmatrix} \bar{\epsilon}_t \\ \pi_t - \bar{\pi}_{t|t-1} \end{bmatrix}.
$$

Let $\overline{\Gamma}(k) = \mathrm{Cov}[\overline{g}_t, \overline{g}_{t-k}]$ be the autocovariance function. With $k \geq 0$,

$$\mathrm{Cov}[\overline{\epsilon}_t, \overline{\epsilon}_{t-k}] = \mathrm{Cov}[\iota_t + \overline{\eta}_t, \iota_{t-k} + \overline{\eta}_{t-k}] = 1_{\{k=0\}}(\sigma_\iota^2 + \sigma_\eta^2/N),$$

where $\overline{\eta}_t = N^{-1}\sum_{i=1}^N \eta_t^i$. Also it follows that $\overline{\epsilon}_t$ is uncorrelated with $\pi_{t-k} - \overline{\pi}_{t-k|t-k-1}$, the second component of $\overline{g}_{t-k}$. Next, by recursively repeating the argument used in the proof of Theorem 2, we find

$$\begin{aligned}
\pi_t - \pi^i_{t|t-1} &= \rho^k(1-G)^k(\pi_{t-k} - \pi^i_{t-k|t-k-1}) \\
&\quad - G\sum_{\ell=1}^{k}\rho^\ell(1-G)^\ell\epsilon^i_{t-\ell} + \sum_{\ell=0}^{k-1}\rho^\ell(1-G)^\ell u_{t-\ell} \\
\pi_t - \overline{\pi}_{t|t-1} &= \rho^k(1-G)^k(\pi_{t-k} - \overline{\pi}_{t-k|t-k-1}) \\
&\quad - G\sum_{\ell=1}^{k}\rho^\ell(1-G)^\ell\overline{\epsilon}_{t-\ell} + \sum_{\ell=0}^{k-1}\rho^\ell(1-G)^\ell u_{t-\ell}.
\end{aligned}$$

From this it follows that

$$\begin{aligned}
\mathrm{Cov}[\pi_t - \overline{\pi}_{t|t-1}, \overline{\epsilon}_{t-k}] &= -1_{\{k>0\}}\, G\rho^k(1-G)^{k-1}(\sigma_\iota^2 + \sigma_\eta^2/N) \\
\mathrm{Cov}[\pi_t - \overline{\pi}_{t|t-1}, \pi_{t-k} - \overline{\pi}_{t-k|t-k-1}] &= \rho^k(1-G)^k\overline{P},
\end{aligned}$$

and hence

$$\overline{\Gamma}(0) = \begin{bmatrix} \sigma_\iota^2 + \sigma_\eta^2/N & 0 \\ 0 & \overline{P} \end{bmatrix} \quad \overline{\Gamma}(k) = \begin{bmatrix} 0 & 0 \\ -\rho^k(1-G)^{k-1}(1-\overline{G})\overline{P}G/\overline{G} & \rho^k(1-G)^k\overline{P} \end{bmatrix}.$$

Also we obtain $\overline{\Gamma}(k+1) = \rho(1-G)\overline{\Gamma}(k)$ if $k \geq 1$. Next, let

$$\overline{v}_t = N^{-1}\sum_{i=1}^{N} v_t^i = \pi_t - \overline{\pi}^\theta_{t|t-1} = [0,1]\overline{g}_t - \rho\theta G[1,1]\overline{g}_{t-1},$$

and observe that the regression errors are

$$e_t = N^{-1}\sum_{i=1}^{N} e_t^i = \nu_t^h - \rho^h \left([(1+\theta), \theta]\overline{g}_t - \rho\theta(1-G)[1,1]\overline{g}_{t-1} \right).$$

Then we can compute for $j \geq 0$:

$$\begin{aligned}\mathrm{Cov}[e_t, \overline{v}_{t-j}] = -\rho^h &\left([1+\theta, \theta]\overline{\Gamma}(j)[0,1]' - \rho\theta(1-G)[1,1]\overline{\Gamma}(j-1)[0,1]' \right.\\ &\left. -\rho\theta G[1+\theta,\theta]\overline{\Gamma}(j+1)[1,1]' + \rho^2\theta^2(1-G)G[1,1]\overline{\Gamma}(j)[1,1]' \right).\end{aligned}$$

For $j \geq 1$ this covariance is

$$\begin{aligned}&- \rho^h \left(\rho(1-G)[1+\theta,\theta]\overline{\Gamma}(j-1)[0,1]' - \rho\theta(1-G)[1,1]\overline{\Gamma}(j-1)[0,1]' \right.\\ &\left. -\rho^2\theta G(1-G)[1+\theta,\theta]\overline{\Gamma}(j)[1,1]' + \rho^2\theta^2(1-G)G[1,1]\overline{\Gamma}(j)[1,1]' \right)\\ &= -\rho^h \left(\rho(1-G)([1+\theta,\theta] - [\theta,\theta])\overline{\Gamma}(j-1)[0,1]' \right.\\ &\left. +\rho^2\theta G(1-G)([\theta,\theta] - [1+\theta,\theta])\overline{\Gamma}(j)[1,1]' \right)\\ &= -\rho^h \left(\rho(1-G)[1,0]\overline{\Gamma}(j-1)[0,1]' - \rho^2\theta G(1-G)[1,0]\overline{\Gamma}(j)[1,1]' \right) = 0.\end{aligned}$$

Also for $j = 0$ we obtain

$$\begin{aligned}&- \rho^h \left([1+\theta,\theta]\overline{\Gamma}(0)[0,1]' - \rho\theta(1-G)[1,1]\overline{\Gamma}(-1)[0,1]' \right.\\ &\left. -\rho\theta G[1+\theta,\theta]\overline{\Gamma}(1)[1,1]' + \rho^2\theta^2(1-G)G[1,1]\overline{\Gamma}(0)[1,1]' \right)\\ &= -\theta\rho^h(\overline{P} + \sigma_\epsilon^2) \left(\overline{G} - \rho^2(1-G+\theta G)(\overline{G} - G) + \rho^2\theta(1-G)G \right),\end{aligned}$$

which is zero if and only if $\theta = 0$. Hence $\overline{z}_t$ is uncorrelated with the regression error e_t, whether $\theta = 0$ or not. For the consensus covariate $\overline{x}_t$, its covariance with $\overline{v}_{t-j}$ is

$$\begin{aligned}\mathrm{Cov}[\overline{x}_t, \overline{v}_{t-j}] = \rho^h G \Big(&(1+\theta)[1,1]\overline{\Gamma}(j)[0,1]' - \rho\theta(1+\theta)G[1,1]\overline{\Gamma}(j+1)[1,1]' \\ &-\rho\theta[1,1]\overline{\Gamma}(j-1)[0,1]' + \rho^2\theta^2 G[1,1]\overline{\Gamma}(j)[1,1]' \Big).\end{aligned}$$

When $j = 0$ this becomes

$$\rho^h G(\overline{P} + \sigma^2_{\overline{\epsilon}}) \Big((1+\theta)\overline{G} - \rho^2\theta(1 + (1+\theta)G)(\overline{G} - G) + \rho^2\theta^2 G \Big),$$

which is non-zero. For $j \geq 1$, we obtain

$$\rho^{h+j} G(1-G)^{j-1}\overline{P}((1-G)(1+\theta) - \theta)(1 - \rho^2\theta G(1 - G/\overline{G})),$$

which is also non-zero. Hence the $\overline{z}_t$ has non-trivial correlation with the covariates, and therefore as an instrumental variable will yield a consistent estimate of the regression parameter. $\square$

Appendix B Additional Simulation Results

Table B.1: Test for forecast efficiency $N = 10$ and $T = 1,000$

ρ	NSR	θ	CONS IV F-stat	CONS IV AR-stat	IND IV AR-stats
.85	0	0	.0454	.0486	.0486
	.8	0	1	.8591	.8759
		.5	.1150	.0724	.0552
	1.6	0	1	.9991	.9997
		.5	.9275	.7928	.8425
		.9	.0809	.0923	.0877
	4.0	random	.8656	.9957	1
.95	0	0	.0528	.0511	.0511
	.8	0	1	.9224	.9475
		.5	.0720	.0590	.0464
	1.6	0	1	1	1
		.5	.8918	.7766	.8495
		.9	.0514	.0484	.0399
	4.0	random	.7626	.9799	1

Note: rejection rates are provided by process $\rho = .85, .95$, NSR $= 0, .8, 1.6, 4.0$, and $\theta = 0, .5, .9$, for the Consensus IV first-stage F-statistic (column 4), the Consensus IV Anderson-Rubin (AR) statistic (column 5), and the combined Individual IV Anderson-Rubin (AR) statistic (column 6). Results are based on 10,000 simulations of length $1,000$, with 10 agents.

Table B.2: Test for forecast efficiency $N = 50$ and $T = 1,000$

ρ	NSR	θ	CONS IV F-stat	CONS IV AR-stat	IND IV AR-stats
.85	0	0	.0517	.0495	.0495
	.8	0	1	.8489	.8513
		.5	.1241	.0720	.0460
	1.6	0	1	.9997	.9999
		.5	.9343	.8091	.8580
		.9	.0907	.0932	.0802
	4.0	random	.9022	.9989	1
.95	0	0	.0448	.0504	.0504
	.8	0	1	.9288	.9491
		.5	.0761	.0622	.0405
	1.6	0	1	1	1
		.5	.9041	.7793	.8639
		.9	.0499	.0518	.0361
	4.0	random	.8121	.9910	1

Note: rejection rates are provided by process $\rho = .85, .95$, NSR $= 0, .8, 1.6, 4.0$, and $\theta = 0, .5, .9$, for the Consensus IV first-stage F-statistic (column 4), the Consensus IV Anderson-Rubin (AR) statistic (column 5), and the combined Individual IV Anderson-Rubin (AR) statistic (column 6). Results are based on 10,000 simulations of length $1,000$, with 50 agents.