

Model Averaging and Persistent Disagreement

In-Koo Cho and Kenneth Kasa

The authors consider the following scenario: Two agents construct models of an endogenous price process. One agent thinks the data are stationary, the other thinks the data are nonstationary. A policymaker combines forecasts from the two models using a recursive Bayesian model averaging procedure. The actual (but unknown) price process depends on the policymaker's forecasts. The authors find that if the policymaker has complete faith in the stationary model, then beliefs and outcomes converge to the stationary rational expectations equilibrium. However, even a grain of doubt about stationarity will cause beliefs to settle on the nonstationary model, where prices experience large self-confirming deviations away from the stationary equilibrium. The authors show that it would take centuries of data before agents were able to detect their model misspecifications. (JEL C63, D84)

Federal Reserve Bank of St. Louis *Review*, Third Quarter 2017, 99(3), pp. 279-94.
<https://doi.org/10.20955/r.2017.279-294>

Parameter uncertainty, while difficult, is at least a relatively well defined problem. Selecting the right model from among a variety of non-nested alternatives is another matter entirely. While there is some formal literature on this problem, I think it is safe to say that central bankers neither know nor care much about this literature. I leave it as an open question whether they are missing much. My approach to this problem while on the Federal Reserve Board was relatively simple: Use a wide variety of models and don't ever trust any one of them too much. ...My usual procedure was to simulate a policy on as many of these models as possible, throw out the outliers, and average the rest to get a point estimate of a dynamic multiplier path.

—Blinder (1998, p. 12)

1 INTRODUCTION

The Federal Reserve is in the forecasting business. Monetary policy affects the economy with a lag, so policy must be based on forecasts. Unlike the weather, where the basic equations of fluid dynamics are well-known, no one knows what the economy's underlying equations

In-Koo Cho is a professor of economics at the University of Illinois, Urbana-Champaign, and Hanyang University and a research fellow at the Federal Reserve Bank of St. Louis. Kenneth Kasa is a professor of economics at Simon Fraser University. The authors thank two anonymous reviewers from the Federal Reserve Bank of St. Louis for several useful comments and suggestions and the National Science Foundation for financial support. The authors are grateful for the hospitality provided by the Federal Reserve Bank of St. Louis.

© 2017, Federal Reserve Bank of St. Louis. The views expressed in this article are those of the author(s) and do not necessarily reflect the views of the Federal Reserve System, the Board of Governors, or the regional Federal Reserve Banks. Articles may be reprinted, reproduced, published, distributed, displayed, and transmitted in their entirety if copyright notice, author name(s), and full citation are included. Abstracts, synopses, and other derivative works may be made only with prior written permission of the Federal Reserve Bank of St. Louis.

are. In fact, there is widespread disagreement about what the best approximation of these equations might be. By itself, disagreement is not a problem. Even weather forecasters face choices about the best approximation of the highly nonlinear equations of fluid dynamics. A commonly employed strategy for dealing with competing forecasts, both in economics and weather forecasting, is to *average* them, as the above quote from Blinder (1998) illustrates. The case for averaging rests on both solid decision-theoretic foundations and a wealth of practical experience.¹

In Cho and Kasa (2016) we argue that model averaging confronts dangers that are not present in weather forecasting. When a weather forecaster calls for rain, his forecast does not influence the likelihood of rain. In contrast, the forecasts of macroeconomic policymakers *do* influence the likelihood of future outcomes, since policy is based on forecasts and future outcomes depend on current policy. It is precisely this feedback, or endogeneity, that creates dangers for model averaging.

Although feedback might seem to pose insurmountable hurdles to those schooled in the natural sciences, economists have in fact devised sophisticated methods for dealing with it. In particular, Lucas (1972) showed that feedback produces a fixed-point problem in the mapping between beliefs and outcomes. Solutions of this fixed-point problem are called a rational expectations equilibrium, and models that violate this fixed-point condition are said to be vulnerable to the “Lucas critique.”

Unfortunately, the solution proposed by Lucas (1972) is not that useful to macroeconomic policymakers, who must decide how to weigh competing forecasts. In a rational expectations equilibrium, model averaging becomes a moot issue, since everyone is presumed to have common knowledge of the correct underlying model. As a result, there is no disagreement to average out.

So in place of Lucas (1972), Cho and Kasa (2016) pursue an approach advocated by Hansen (2014). Hansen suggests that economists populate their models with agents who behave just like econometricians who construct simple and useful provisional models and revise them as needed in response to statistical evidence. In a sense, this is not a new idea. Adaptive learning has long been used as a defense of the rational expectations hypothesis, and there are results that support this defense.² However, these results do not confront the issues of model uncertainty and competing forecasts. They assume agents have somehow settled on a *single* model and must merely learn its coefficients. Hansen (2014) instead emphasizes that the real problem confronting econometricians is not how to estimate a given model, but rather how to test and discriminate among competing models, all of which are recognized to be imperfect approximations of the underlying data-generating process.

As noted in the opening quote by Blinder (1998), this is a substantially more difficult problem. In fact, without a priori knowledge of the model, the results of Cho and Kasa (2016) cast doubt on the ability of agents to learn their way to a stationary rational expectations equilibrium. In particular, we show that the mere presence of a nonstationary alternative model can produce a sort of “Gresham’s law” phenomenon, in which a stationary model is driven out of consideration simply because use of the nonstationary model creates conditions that support its continued use. This volatility trap could be avoided if the policymaker had dog-

matic priors that ruled out the nonstationary alternative. It could also be avoided in a Lucas world, where policymakers fully understand (somehow) the endogeneity they confront. Unfortunately, neither alternative is persuasive. Instead, using the results of Cho and Kasa (2015), we argue that in the presence of potentially misspecified endogeneity it might be preferable to avoid model averaging altogether and instead base forecasts on a process of specification testing and *model selection*.

The results in Cho and Kasa (2016) were based on the previous work of Evans et al. (2013). Following Evans et al. (2013), we assumed the competing models are in the mind of a single agent. However, if this is the case, one wonders why the agent does not just expand the model space to include a single model that nests both. As noted by Timmermann (2006), nesting does not occur in practice, partly due to degrees of freedom considerations but mainly because competing models are based on differing information sets and heterogeneous prior beliefs among forecasters.

In the current paper, we extend the results of Cho and Kasa (2016) to this more relevant decentralized environment, where competing models reflect the beliefs of distinct agents. This multiple-agent setting raises some subtle issues. Our previous analysis focused on the policymaker's problem. He had to decide how to weigh the two forecasts. Here, with multiple agents, we must also consider the problem of the forecasters. Do they know their forecasts are being used by the policymaker, and are they aware the data might be endogenous? Do they know they are competing with another forecaster? If so, what are their beliefs about their rival's model? Would forecasters ever have an incentive to revise their beliefs about either their own model or another rival's model?

We show that Gresham's law of model averaging continues to apply in this more realistic decentralized setting. The agent using the nonstationary/drifted parameters model eventually dominates. Moreover, both forecasters maintain their beliefs about both their own model and their rival's model. This is interesting because *both* agents have misspecified models. However, the nature of their misspecifications turns out to be very difficult to detect statistically in finite samples.

In particular, the agent with the drifting parameters model has misspecified beliefs about his own model. Since the underlying model features constant parameters, the actual data-generating process is stationary. However, when expectational feedback is strong, mean reversion is weak, and so the agent with the drifting parameters model confronts the problem of testing the null of a unit root against a local alternative. These tests are well-known to have low power. In fact, we show that it would take *centuries* of data before the agent could reject his model with any confidence.

The agent with the constant parameters model has a correctly specified model. However, his beliefs about his *rival's* model are misspecified. In particular, he rationalizes the ongoing parameter variation in his rival's model as reflecting noise around a constant parameter value. In reality, his rival's model features mean-reverting dynamics that pull the estimates toward the rational expectations equilibrium. However, as before, this mean reversion is quite weak, and so the agent with the constant parameters model confronts the problem of testing the null of i.i.d. fluctuations against the alternative of weak autocorrelation. Again, we find that it

would take centuries of data before the agent could reject his beliefs about his rival's model with any confidence.

Although we frame our discussion within the context of a specific model, our results are of broader significance. Economists beyond a certain age will likely recall an era that is sometimes called the “unit root wars.” During the 1980s and 1990s there were heated debates among economists about whether macroeconomic time series were “trend stationary” or “difference stationary.” Econometricians designed ever more esoteric procedures to test these hypotheses. Although echoes of these debates still resound, for the most part the unit root wars have died out, without a clear victor. At the end of the day, what the unit root wars taught us is that it is very difficult to say anything with much confidence about stationarity. We simply do not have enough data.³

Although lack of a clear victor may have been disappointing to the participants, applied economists have largely taken the advice of Cochrane (1991) by avoiding the question altogether. Cochrane argues that the practically relevant question is which asymptotic distribution provides the better approximation to the actual (finite-sample) distribution, which is a question that must be decided on a case-by-case basis.

Our results sound a note of caution about this pragmatic approach to the question of stationarity. Cochrane's advice is based on the perspective of what Hansen (2014) calls the “outside econometrician,” meaning someone whose actions do not influence the system he is trying to learn about. The agents in our model are engaged in their own unit root war, and perhaps not surprisingly, they find that they cannot settle it either with the available data. However, they are *not* outside econometricians. Their beliefs about stationarity influence the data-generating process. We show that doubts about stationarity can produce outcomes that statistically confirm those doubts. This is a far more pessimistic conclusion than Cochrane's. In a self-referential world, persistent belief differences about stationarity actually matter.

The remainder of the paper is organized as follows. The next section outlines a simple model that conveys the basic idea. The essential feature of this model is that current outcomes depend on beliefs about future outcomes. We assume there are two forecasters who share a common reduced-form model but have differing beliefs about parameter stability. One thinks the parameters are stable; the other thinks parameters drift. We show how the forecasters can revise their beliefs using the Kalman filter. Section 3 considers the problem of the policymaker who must decide how to weigh the competing forecasts. Like Blinder in the opening quote, we assume he hedges his bets by averaging them, with weights that are recursively updated based on historical relative performance. We state our Gresham's law result, which shows that eventually the time-varying parameters (TVP) forecaster must dominate. We do not formally prove the result but merely provide the intuition. Section 4 turns to the problem of the forecasters and asks whether they would have any reason to modify or reject their models. Since they both have misspecified models, with an infinite sample they will both be able to reject their models. However, the empirically relevant question is how long it will take. We present Monte Carlo evidence that suggests it would take centuries. In the meantime, prices are substantially more volatile than they would otherwise be. Section 5 provides a brief discussion of our results and attempts to place them in context. Finally, Section 6 offers a few concluding remarks.

2 A SIMPLE MODEL

We begin by outlining a special case of the model considered by Cho and Kasa (2016), in which a price at time t , p_t , is determined according to

$$(1) \quad p_t = \delta + \alpha E_t p_{t+1} + \sigma \epsilon_t,$$

where $\alpha \in (0,1)$ is a discount rate. This is a key parameter in the ensuing analysis. It determines the strength of expectational feedback. For simplicity, the ϵ_t shock is assumed to be Gaussian white noise. This model has been a workhorse in the macroeconomic learning literature. It can be interpreted as an asset-pricing model with constant fundamentals. The unique stationary rational expectations equilibrium is

$$(2) \quad p_t = \frac{\delta}{1-\alpha} + \sigma \epsilon_t.$$

Hence, rational expectations predicts that prices exhibit i.i.d. fluctuations around a fixed mean with a variance of σ^2 .

2.1 Learning

Note that rational expectations is an *equilibrium* concept. It says nothing about how, or even whether, this equilibrium could ever be attained. The original architects of the rational expectations revolution (John Muth, Robert Lucas, Edward Prescott, and Thomas Sargent) defended it by arguing that a process of adaptive, out-of-equilibrium learning would eventually produce convergence to a rational expectations equilibrium. Applications assume this process of learning has already taken place. However, formal analysis of this conjecture did not begin until the 1980s. Due to the presence of feedback, which makes the data endogenous, addressing the convergence question is not a simple matter of applying the law of large numbers.

There is now a well-developed literature that provides conditions under which a wide variety of economic models can be expected to converge to rational expectations equilibria.⁴ For the particular model in equation (1), the crucial restriction is that $\alpha < 1$. That is, feedback cannot be too strong. Fortunately, this imposes no additional restrictions in this case, since theory requires $\alpha < 1$.

Our paper questions these existing convergence results. These results are based on imputing dogmatic priors to agents. The typical case presumes agents are convinced that parameters are constant; they are just not sure what the constants are. They rule out a priori the possibility that parameters drift. A more recent literature, based on so-called “constant gain” learning algorithms, presumes agents have a dogmatic prior that parameters drift. Here agents are not even given the chance to learn the constant parameter rational expectations equilibrium, because their priors say this is a zero probability event.

Here, agents are less dogmatic. In particular, the policymaker is open-minded. He puts positive weight on both possibilities and then revises his beliefs using Bayes’ rule as the data come in. One might suspect, based on a naive application of “grain of truth” sort of arguments,

that eventually he will learn the constant parameter rational expectations equilibrium since his prior puts positive weight on this possibility. This will indeed be the case if feedback is relatively weak. However, we show that if $\alpha > 1/2$, which is the empirically relevant case, the policymaker's beliefs will eventually converge to the time-varying parameters model. The usual grain-of-truth argument does not apply here because *both* models are misspecified due to the presence of feedback.

To be more precise, suppose there are two forecasters who share the following state-space model for prices:

$$(3) \quad p_t = \beta_t + \sigma \epsilon_t$$

$$(4) \quad \beta_t = \beta_{t-1} + \sigma_v \nu_t,$$

where it is assumed that $\text{cov}(\epsilon, \nu) = 0$. Note that the rational expectations equilibrium is a special case of this, with

$$\sigma_v = 0 \text{ and } \beta = \frac{\delta}{1-\alpha}.$$

For now, suppose the beliefs of the forecasters about parameter stability are dogmatic. Later, in Section 4, we allow the forecasters to question their priors. In particular, suppose one forecaster, whom we call \mathcal{M}_0 , believes parameters are constant, that is,

$$\mathcal{M}_0: \sigma_v^2 = 0,$$

while the other forecaster, \mathcal{M}_1 , is convinced parameters are time-varying, that is,

$$\mathcal{M}_1: \sigma_v^2 > 0.$$

Both forecasters use the Kalman filter to revise their beliefs about β_t :

$$(5) \quad \beta_{t+1}(i) = \beta_t(i) + \left(\frac{\Sigma_t(i)}{\sigma^2 + \Sigma_t(i)} \right) (p_t - \beta_t(i)).$$

The only difference is in how they update their beliefs about the variance of β_t :

$$(6) \quad \Sigma_{t+1}(0) = \Sigma_t(0) - \frac{\Sigma_t^2(0)}{\sigma^2 + \Sigma_t(0)}$$

$$(7) \quad \Sigma_{t+1}(1) = \Sigma_t(1) - \frac{\Sigma_t^2(1)}{\sigma^2 + \Sigma_t(1)} + \sigma_v^2.$$

Note that the parameter update, equation (5), assumes that β_t is based on time- $(t-1)$ information. This assumption is made to avoid simultaneity between beliefs and observations.⁵

It turns out that both forecasters' estimates converge to the same (rational expectations) value, $\beta_t(i) \rightarrow \delta/(1-\alpha)$. This is not too surprising since they are both forecasting the same thing using the same data and the same reduced-form model. However, there is an important

difference in the nature of the convergence. The estimates of \mathcal{M}_0 converge in a relatively strong sense, based on the law of large numbers. As the sample size grows, it becomes increasingly unlikely that $\beta_t(0)$ will deviate from $\delta/(1-\alpha)$ by any given amount. In contrast, the estimates of \mathcal{M}_1 converge in a weaker, distributional sense, based on the central limit theorem. His estimates never settle down. They fluctuate persistently around the rational expectations value. These persistent fluctuations reflect this forecaster’s desire to remain alert to potential parameter instability.

One might suspect that since the underlying structural parameters (δ, α) are constant, \mathcal{M}_0 is “right” and \mathcal{M}_1 is “wrong,” so that if someone were able to cast and revise a (weighted) vote on who’s right, eventually \mathcal{M}_0 would win. In fact, precisely the opposite occurs.

3 MODEL AVERAGING

We now turn to the problem of the policymaker. The policymaker does not construct models himself. He is presented with competing forecasts and must decide how to use them. Like Alan Blinder in the opening quote, we assume he does this by averaging them. Specifically, let π_t denote the current probability assigned by the policymaker to \mathcal{M}_1 , the TVP model, and let $\beta_t(i)$ denote the current parameter estimate for \mathcal{M}_i . The policymaker’s time- t forecast becomes

$$E_t p_{t+1} = \pi_t \beta_t(1) + (1 - \pi_t) \beta_t(0).$$

Substituting this into the actual law of motion for prices in equation (1) implies that the forecasters’ parameter estimates evolve according to

$$(8) \quad \beta_{t+1}(i) = \beta_t(i) + \left(\frac{\Sigma_t(i)}{\sigma^2 + \Sigma_t(i)} \right) \left\{ [\delta + \alpha[\pi_t \beta_t(1) + (1 - \pi_t) \beta_t(0)] - \beta_t(i)] + \sigma \epsilon_t \right\}.$$

Notice the presence of feedback here. The evolving beliefs of the forecasters depend on the beliefs of the policymaker. The more confidence the policymaker has in \mathcal{M}_i , the more likely \mathcal{M}_i will prevail. For now, we suppose the forecasters are unaware of this feedback when constructing their models.

Since $\beta_t(i) \rightarrow \delta/(1-\alpha)$ for \mathcal{M}_i , where $i = 1, 2$, the only real question is what happens to π_t . We assume the policymaker is a Bayesian. He views the current value of π_t as a prior and then updates it using Bayes’ rule,

$$(9) \quad \frac{1}{\pi_{t+1}} - 1 = \frac{A_{t+1}(0)}{A_{t+1}(1)} \left(\frac{1}{\pi_t} - 1 \right),$$

where

$$(10) \quad A_t(i) = \frac{1}{\sqrt{2\pi(\sigma^2 + \Sigma_t(i))}} \exp \left[-\frac{(p_t - \beta_t(i))^2}{2(\sigma^2 + \Sigma_t(i))} \right]$$

is the time- t likelihood of model i .

The ultimate fate of this economy has now been converted to a mathematical problem. Equations (6), (7), (8), (9), and (10) are a system of seven nonlinear stochastic difference equations in $(\pi_t, \beta_t(i), \Sigma_t(i), A_t(i))$, where $i = 1, 2$. Solving even one nonlinear stochastic difference equation can be challenging, so this would seem to be a hopeless endeavor. The key to making this system tractable is to exploit the fact that subsets of the variables evolve on different timescales. By appropriately averaging over each subset, we can simplify the analysis to one of studying the interactions between lower dimensional subsystems. In particular, variables that operate on a relatively slow timescale can be fixed at their current values, while variables that operate on a relatively fast timescale can be fixed at the means of their stationary distributions. In Cho and Kasa (2016) we show that this economy features a hierarchy of *four* timescales. The data operate on a relatively fast calendar timescale. Estimates of the TVP model evolve on a slower timescale determined by the parameter innovation variance. Estimates of the constant parameter model evolve even more slowly, on a timescale determined by the inverse of the historical sample size. Finally, the model weight, π_t , evolves on a variable timescale but spends most of its time in the neighborhood of either 0 or 1, where it evolves on a timescale that is even slower than that of the constant parameter model.

Exploiting these timescale separation methods, Cho and Kasa (2016) prove the following result.

Theorem 3.1. $\forall \epsilon > 0, \forall T \geq 1$, define

$$T_1^\epsilon = \#\{t \leq T | \pi_t \geq 1 - \epsilon\}$$

as the number of periods $t \leq T$ when the policymaker believes in \mathcal{M}_1 with probability more than $1 - \epsilon$. If $\pi_0 \neq 0$ and $1/2 < \alpha < 1$, then $\forall \epsilon > 0$,

$$\lim_{\sigma_v^2 \rightarrow 0} \lim_{T \rightarrow \infty} E \frac{T_1^\epsilon}{T} = 1.$$

This result says that as long as feedback is sufficiently strong ($\alpha > 1/2$) and assumed parameter drift is sufficiently weak ($\sigma_v^2 \rightarrow 0$), the TVP will eventually dominate. Note the order of limits is important here. By letting $T \rightarrow \infty$ first, what we are doing is comparing asymptotic distributions as the parameter σ_v^2 changes. The same strategy is used when using stochastic stability arguments to select among multiple Nash equilibria (Kandori, Mailath, and Rob, 1993).⁶ Of course, from a practical standpoint, the relevant question is what happens for strictly positive values of σ_v^2 . Using simulations, Cho and Kasa (2016) show that convergence is continuous and that values of σ_v^2 that generate significant steady-state “excess volatility” are associated with a value of π_t well in excess of 1/2. The following section provides additional results along these lines.

The intuition for why the TVP model eventually dominates is the following: When the weight on the TVP model is close to 1, the world is relatively volatile (due to feedback). This makes the constant parameters model perform relatively poorly since it is unable to track the feedback-induced time-variation in the data. Of course, the tables are turned when the weight

on the TVP model is close to zero. Now the world is relatively tranquil, and the TVP model's additional noise puts it at a disadvantage. However, as long as this noise isn't too large, the TVP model can exploit its ability to respond to rare sequences of shocks that generate "large deviations" in the estimates of the constant parameters model. That is, during tranquil times, the TVP model is lying in wait, ready to pounce on large-deviation events. These events provide a foothold for the TVP model, which due to feedback allows it to regain its dominance.

Interestingly, the TVP model asymptotically dominates here because it is better able to react to the volatility that it itself creates. Although \mathcal{M}_1 is misspecified from the perspective of the rational expectations equilibrium, this equilibrium must be *learned* via some adaptive process. Our result shows that this learning process can be subverted by the mere presence of a misspecified alternative, even when the correctly specified model would converge if considered in isolation. This result therefore echoes the conclusions of Sargent (1993), who notes that adaptive learning models often need a lot of "prompting" before they converge. Elimination of misspecified alternatives can be interpreted as a form of prompting.

4 PERSISTENT DISAGREEMENT

The policymaker in the previous analysis is quite sophisticated. Although he is ignorant of the underlying structural model, he is aware of his own ignorance and recognizes that in the presence of model uncertainty it is wise to give all models a chance. Moreover, he is not dogmatic. As evidence accumulates, he revises his beliefs.

The same cannot be said of the forecasters. Although they revise coefficient estimates, they dogmatically adhere to their beliefs about parameter stability. Moreover, they operate in a vacuum, in the sense that they are totally unaware of each other's presence and ignore the fact that the policymaker is actually using their forecasts. Here we extend the analysis of Cho and Kasa (2016) by assuming the forecasters are a bit more sophisticated.⁷

In doing this, we assume the forecasters follow Neyman-Pearson principles. That is, they stick with a null model unless sufficient evidence mounts against it. This is not as schizophrenic as it might seem. Bayesian decision theory presumes agents have full confidence in their priors. If they didn't, they would have different priors. Policymakers must decide how best to use the information they receive. In doing this, it makes sense to adhere to Bayesian principles. However, agents engaged in the construction of economic models confront a less structured and more open-ended problem. They must remain alert to the possibility that *all* current models are misspecified and be prepared to expand their priors in response. In this environment, we think Neyman-Pearson behavior makes sense.⁸

As before, we can summarize the beliefs of \mathcal{M}_0 and \mathcal{M}_1 as a pair of perceived state space models. The main difference is that now we must include each forecaster's belief about his rival's model. These beliefs take the form of conjectures about the other forecaster's beliefs about stationarity. \mathcal{M}_1 thinks \mathcal{M}_0 uses a constant parameter model, while \mathcal{M}_0 thinks \mathcal{M}_1 uses a *random* (not drifting) coefficients model with i.i.d. fluctuations around a constant mean. Specifically, the perceived observation equation for \mathcal{M}_0 is

$$(11) \quad p_t = (1 - \pi_t)\beta_t(0) + \pi_t\bar{\beta}_t(1) + \sigma\epsilon_t,$$

while the state transition equation is

$$\begin{aligned} \beta_t(0) &= \beta_{t-1}(0) \\ \bar{\beta}_t(1) &= \bar{\beta}_{t-1}(1) \end{aligned}$$

where an overbar is used to represent an agent's belief about the other agent's model. To simplify notation, define

$$y_t(0) = \begin{bmatrix} p_t \\ \beta_t(1) \end{bmatrix}, \quad \xi_{t-1}(0) = \begin{bmatrix} \beta_t(0) \\ \bar{\beta}_t(1) \end{bmatrix}, \quad H_t = \begin{bmatrix} (1 - \pi_t) \\ \pi_t \end{bmatrix}, \quad Q(0) = 0$$

and

$$R(0) = \text{cov} \begin{bmatrix} \sigma\epsilon_t \\ \epsilon_{1,t} \end{bmatrix}.$$

Similarly, we can write the perceived observation equation of \mathcal{M}_1 as

$$(12) \quad p_t = (1 - \pi_t)\bar{\beta}_t(0) + \pi_t\beta_t(1) + \sigma\epsilon_t$$

and his perceived state transition equation as

$$(13) \quad \bar{\beta}_t(0) = \bar{\beta}_{t-1}(0)$$

$$(14) \quad \beta_t(1) = \beta_{t-1}(1) + \epsilon_{v,t}.$$

Define

$$y_t(1) = \begin{bmatrix} p_t \\ \beta_t(0) \end{bmatrix}, \quad \xi_{t-1}(1) = \begin{bmatrix} \bar{\beta}_t(0) \\ \beta_t(1) \end{bmatrix}, \quad \text{and } Q(1) = \text{cov} \begin{bmatrix} 0 \\ \epsilon_{v,t} \end{bmatrix}$$

and

$$R(1) = \text{cov} \begin{bmatrix} \sigma\epsilon_t \\ \epsilon_{0,t} \end{bmatrix}.$$

As before, the evolving conditional mean and variance of the posteriors are described by the Kalman filter:

$$(15) \quad \xi_t(i) = \xi_{t-1} + P_{t-1}(i)H_t(H_t'P_{t-1}H_t + R(i))^{-1}(y_t - H_t'\xi_{t-1})$$

$$(16) \quad P_t(i) = P_{t-1}(i) - P_{t-1}(i)H_t(H_t'P_{t-1}(i)H_t + R(i))^{-1}H_t'P_{t-1}(i) + Q(i).$$

Note that the forecasters are assumed to know the policymaker's current model weight, π_t .

The Kalman filter generates a sequence of hidden state estimates that can then be substituted into the perceived observation equations to generate sequences of price forecasts:

$$(17) \quad \begin{aligned} \hat{p}_t(0) &= (1 - \pi_t)\beta(0)_t + \pi_t\bar{\beta}(1)_t \\ \hat{p}_t(1) &= (1 - \pi_t)\bar{\beta}(0)_t + \pi_t\beta(1)_t. \end{aligned}$$

As before, the policymaker takes these two forecasts and averages them:

$$(18) \quad \hat{p}_t = (1 - \pi_t)\hat{p}_t(0) + \pi_t\hat{p}_t(1).$$

The actual time- t price is then determined by the (unknown) structural model in equation (1):

$$(19) \quad p_t = \delta + \alpha\hat{p}_t + \sigma\epsilon_t.$$

Notice that if equations (17) are substituted into (18), which is then substituted into the actual price equation (19), we observe that the agents' models in (11) and (12) suffer from an additional form of misspecification, since each forecaster fails to recognize that his own forecast embodies a form of model averaging in its response to the other forecaster's forecast. However, this misspecification disappears in the limit.

It turns out that if \mathcal{M}_0 and \mathcal{M}_1 do not reject their models, our previous Gresham's law result continues to apply: $\pi_t \rightarrow 1$, and the economy is plagued by self-confirming volatility.⁹ However, because the two models are now more symmetric, the rate of convergence is slower. Hence, the key question is whether the two forecasters would have any reason to reject their models. Since both models are misspecified, we know that with an *infinite* sample they will eventually reject them. However, we also know the sun will vaporize our planet before this occurs, so the real question is how long it will take before they reject them.

To address this question we perform a simple Monte Carlo experiment. After a fixed interval, we allow each forecaster to test the specification of his model.¹⁰ In the case of \mathcal{M}_0 , this involves testing the null hypothesis that the error term in his model is i.i.d. against the alternative of (first-order) autocorrelation. The persistent mean-reverting learning dynamics of \mathcal{M}_1 make the alternative true. In the case of \mathcal{M}_1 , this involves testing the null hypothesis that $\beta_t(1)$ is a random walk, against the alternative of stationarity. His own learning dynamics, along with the stationarity of the underlying model, again make the alternative true.¹¹

Table 1 reports the results. We consider two sample lengths, $T = 100$ and $T = 600$. We set $\alpha = 0.95$ and $\sigma^2 = 0.10$, which suggests an annual time unit. The important parameter is σ_v^2 , which reflects \mathcal{M}_1 's beliefs about parameter drift. Our theorems pertain to the limit, as $\sigma_v^2 \rightarrow 0$, but we know that at this limit the two models are equivalent and no extra volatility arises. However, if σ_v^2 is too big, it becomes unlikely that convergence to $\pi = 1$ occurs. The only way to proceed is to try out values and see what happens. Table 1 assumes $\sigma_v^2 = 0.0005$. Note, this is orders of magnitude smaller than σ^2 . Good practice recommends reducing significance levels as the sample size increases. However, in Table 1 we assume the forecasters maintain a 5 percent significance level, even when the sample size increases to $T = 600$. This biases the results toward rejection. Finally, the numbers in Table 1 report averages across $N = 600$ replications. Each replication is initialized randomly in the neighborhood of $\pi \approx 1/2$.

Table 1
Monte Carlo Simulations

T	Reject(0)	Reject(1)	PR($\pi = 1$)	var(P)
100	0.022	0.492	0.358	0.109
600	0.480	0.221	0.705	0.153

NOTE: Results are averages across $N = 600$ replications. Parameters: $\sigma^2 = 0.10$; $\sigma_v^2 = 0.0005$; $\alpha = 0.95$.

From the second column, we see that when $T = 100$, \mathcal{M}_0 has no incentive to reject his model. In fact, the rejection probability is less than the size of the test! (which likely reflects sampling variability). On the other hand, \mathcal{M}_1 has nearly a 50 percent chance of rejecting his model. This discrepancy is explained by the fourth column, which reports the proportion of time that $\pi_t \in (0.95, 1)$ at the end of the sample. Evidently, there is little tendency for $\pi_t \rightarrow 1$ when sample sizes are relatively small. This is not surprising. When $T \approx 100$, the implicit gain of \mathcal{M}_0 's updated equation is not that different from \mathcal{M}_1 's. When $\pi_t \approx 0$, it becomes relatively easy for \mathcal{M}_1 to reject his model, since the data are being primarily generated by \mathcal{M}_0 's model, which features constant parameters. Still, a 50 percent rejection probability is nothing to write home about since in fact his model is misspecified. Finally, the last column reports the variance of prices. Remember, in a rational expectations equilibrium the variance would be 0.10. Although the resulting variance is significantly larger than assumed parameter drift, it is only about 10 percent larger than the rational expectations value. Again, this reflects the fact that in small samples there is a high probability of being in the $\pi = 0$ state.

Things become more interesting as T increases. From the second row we see that when $T = 600$, \mathcal{M}_1 's rejection probability actually *decreases* to 0.221. This happens because the likelihood that $\pi_T \approx 1$ almost doubles—to 0.705. At the same time, \mathcal{M}_0 now finds it easier to reject his model. Again, this is because the other guy's model is more likely to be generating the data. Observe that prices are now considerably more volatile. The variance exceeds the rational expectations value by more than 50 percent.

The results in Table 1 suggest that with reasonably high probability, disagreements about important aspects of models can persist for hundreds of years. Granted, $600 \neq \infty$, so an economic theorist might not find these results convincing. However, to policymakers and market participants, 600 years is an eternity.

5 DISCUSSION

Here we briefly attempt to place our results in context and respond to a potential criticism of the way we have modeled beliefs about parameter instability.

5.1 Merging

Economists have long been interested in the question of whether rational individuals can “agree to disagree.” The classic reference is Aumann (1976). Aumann proved that disagree-

ment is impossible, even when agents have different information, as long as they share a common prior and their posteriors (and their prior) are common knowledge. Blackwell and Dubins (1962) show that common priors are to be expected as long as agents initially agree on what is possible and what is impossible (i.e., their priors are mutually absolutely continuous). Kalai and Lehrer (1993) used these results to show that agents can learn to play Nash equilibria.

Our results are more closely aligned with recent work that introduces “frictions” in an effort to overcome these classic asymptotic merging results. Fudenberg and Levine (1993) show that disagreement can persist about off-equilibrium path events. Kurz (1994) shows that disagreement can persist when agents consider only a limited set of moments. Acemoglu, Chernozhukov, and Yildiz (2016) show that disagreement about underlying state variables can persist when agents’ models are not identified. Perhaps most closely related to our work, Esponda and Pouzo (2016) show that disagreement can persist when agents’ models are misspecified.

The common thread running through all this prior work is the presumption that agents have access to infinite samples. Persistence means *forever*. Clearly, it is of interest to know when disagreement will never disappear. However, that doesn’t mean disagreements that take centuries to resolve are uninteresting or unimportant. We think finite samples are an interesting friction. Perhaps the reason it hasn’t been more thoroughly studied is simply that results will likely be less precise, being necessarily probabilistic and more naturally stated using Neyman-Pearson language. Nonetheless, there are results on learning rates and large deviations that could profitably be used to characterize sets of models that agents are unable to discriminate among with a given sample size. Rather than speak of merging to a common singleton model, the task would be to characterize statistical “equivalence classes” of models. Although there is a recent literature along these lines on partial identification within econometrics, it has yet to make its way to the endogenous data setting of macroeconomics. The work of Hansen and Sargent (2008) on robustness and detection error probabilities provides the closest example of what we have in mind.

5.2 Stability

We suspect most Bayesians would argue that our policymaker was asking for trouble from the beginning, since his prior was nonconvex. It consisted of two discrete points, $\sigma_v^2 = 0$ and some other strictly positive value. Repeated switching between $\pi = 0$ and $\pi = 1$ would be interpreted as evidence that the data prefer an intermediate value. Why not let the agents estimate σ_v^2 directly? True, his filtering problem would then be nonlinear, but that’s what computers are for. We have two responses to this critique. First, we agree that if a *single agent* is viewed as both formulating models and using them, then a nonconvex prior makes little sense. However, as discussed previously, actual policy environments are more decentralized and feature distinct agents undertaking distinct tasks using distinct information and their own distinct priors about how the economy works. Nesting these heterogeneous beliefs within a single convex “hypermodel” is wishful thinking. Second, the claim that allowing agents to estimate σ_v^2 directly would avoid the volatility trap ignores the fact that the data are endogenous here. In Cho and Kasa (2016), we pursue an idea from Evans and Honkapohja (1993) and compute

Nash equilibrium values of σ_v^2 . That is, we look for values of σ_v^2 that are best responses to themselves. Not surprisingly, $\sigma_v^2 = 0$ is a Nash equilibrium. If the world is tranquil, then constant parameter models are optimal. However, we also show that there exist strictly positive values of σ_v^2 that are (stable) Nash equilibria, even when agents are allowed to incrementally adjust their own estimates of σ_v^2 . When the world is turbulent, it is individually rational to use a time-varying parameter model, even if you know that if everyone else were to use a constant parameter model a more stable Pareto superior outcome would result.

6 CONCLUSION

One of the occupational hazards of being an economist is enduring the inevitable jokes about disagreement among economists. We've all heard the one about economists being laid end-to-end and still not reaching a conclusion, or the more recent one about economics being the only field where two people can get a Nobel prize for saying opposite things. Like most jokes, these jokes have an element of truth to them. Economists *do* disagree with each other. However, most economic models pretend that they don't.

Does this disagreement matter? Physicists and biologists disagree too, so why should we care whether economists disagree? In this paper we've argued there is an important difference between economics and the natural sciences. Disagreement among physicists presumably doesn't alter the laws of physics. Economics isn't like that. Competing economic theories, when put into practice, have the capacity to create their own self-confirming reality. This paper discussed one particular example involving beliefs about parameter stability, but there are no doubt many others.

Given that the stakes are high, what is the best strategy for coping with this disagreement? We think our paper offers two main lessons. First, we think that examples like this should tip the scales back in favor of theory in economics. In this era of big data and machine learning, one often hears the argument that economic theory has become passé. Why not let the data decide? Our paper shows why these arguments are naive. Letting the data decide is a fine strategy when the data are exogenous, but that's not the case in economics. The agents in our economy suffer bad outcomes because they do not fully understand the nature of the endogeneity they confront. Letting the data decide produces the wrong decision. If a bright young theorist came along and suggested that learning itself might generate parameter instability, the self-confirming volatility trap could be avoided. Second, we think our paper tips the scales in favor of model selection, as opposed to model averaging. Although most econometrics texts outline methods for testing and selecting models, they do so apologetically, with the warning that such procedures lack coherent decision-theoretic foundations. Bayesian decision theory tells us that we should hedge our bets by averaging across models. Our paper shows why model selection might not be such a bad idea after all. The basic problem with averaging is that it forces models to *compete*. Competition is great when the rules are fixed and fair. But here, the time-varying parameter model can effectively alter the rules of the game in its own favor. Again, if econometricians fully understood the endogeneity they confronted, this wouldn't be an issue. When priors are correctly specified, Bayesian methods produce good outcomes. When they're not, Neyman-Pearson methods begin to look better. ■

NOTES

- ¹ See, e.g., Bates and Granger (1969) and Timmerman (2006).
- ² See Evans and Honkapohja (2001) for a comprehensive survey.
- ³ This is not an issue of the number of observations but rather of the length of the period. Perron and Shiller (1985) show that one cannot increase statistical power by sampling a given period more frequently.
- ⁴ Evans and Honkapohja (2001) provide the definitive summary of this literature.
- ⁵ See Evans and Honkapohja (2001) for further discussion.
- ⁶ Taking limits in the reverse order would be uninteresting, since the models become identical as $\sigma_v \rightarrow 0$.
- ⁷ As noted in the Introduction, if we assume the competing models are in the mind of the policymaker himself, then the previous analysis makes more sense. This is how Evans et al. (2013) interpret their analysis. However, actual forecasting and policy environments are more decentralized and feature a division of labor between policymakers and model builders.
- ⁸ We are not alone. For example, Gilboa, Postlewaite, and Schmeidler (2008) argue that Neyman-Pearson behavior is actually more consistent with recent developments in decision theory than is Bayesian decision theory. We should also note that Bayesian practitioners are aware of the importance of testing priors against the data (e.g., Geweke, 2010). These efforts blur the distinction between Bayesian and frequentist behavior.
- ⁹ A formal proof is available upon request.
- ¹⁰ Note, we assume they do this retrospectively, not repeatedly in real time, as in the “monitoring structural change” approach of Chu, Stinchcombe, and White (1996). Hence, standard testing procedures can be applied.
- ¹¹ Note, \mathcal{M}_0 tests his beliefs about his rival’s model, while \mathcal{M}_1 tests his beliefs about his own model. This imparts maximal power to the tests, since \mathcal{M}_0 ’s beliefs about his own model are correct, while \mathcal{M}_1 ’s beliefs about \mathcal{M}_0 ’s model are also correct.

REFERENCES

- Acemoglu, Daron; Chernozhukov, Victor, and Muhamet, Yildiz. “Fragility of Asymptotic Agreement under Bayesian Learning.” *Theoretical Economics*, January 2016, 11(1), pp. 187-225; <https://doi.org/10.3982/TE436>.
- Aumann, Robert J. “Agreeing to Disagree.” *Annals of Statistics*, November 1976, 4(6), pp. 1236-39; <https://doi.org/10.1214/aos/1176343654>.
- Bates, J.M. and Granger, C.W.J. “The Combination of Forecasts.” *Journal of the Operational Research Society*, December 1969, 20(4), pp. 451-68; <https://doi.org/10.1057/jors.1969.103>.
- Blackwell, David and Dubins, Lester. “Merging of Opinions with Increasing Information.” *Annals of Mathematical Statistics*, September 1962, 33(3), pp. 882-86; <https://doi.org/10.1214/aoms/1177704456>.
- Blinder, Alan S. *Central Banking in Theory and Practice*. Cambridge: MIT Press, 1998.
- Cho, In-Koo and Kasa, Kenneth. “Learning and Model Validation.” *Review of Economic Studies*, January 2015, 82(1), pp. 45-82; <https://doi.org/10.1093/restud/rdu026>.
- Cho, In-Koo and Kasa, Kenneth. “Gresham’s Law of Model Averaging.” Working Paper No. 6, Simon Fraser University, 2016.
- Chu, Chia-Shang James; Stinchcombe, Maxwell and White, Halbert. “Monitoring Structural Change.” *Econometrica*, September 1996, 64(5), pp. 1045-65; <https://doi.org/10.2307/2171955>.
- Cochrane, John H. “A Critique of the Application of Unit Root Tests.” *Journal of Economic Dynamics and Control*, April 1991, 15(2), pp. 275-84; [https://doi.org/10.1016/0165-1889\(91\)90013-Q](https://doi.org/10.1016/0165-1889(91)90013-Q).
- Esponda, Ignacio and Pouzo, Demian. “Berk-Nash Equilibrium: A Framework for Modeling Agents with Misspecified Models.” *Econometrica*, May 2016, 84(3), pp. 1093-30; <https://doi.org/10.3982/ECTA12609>.

Cho and Kasa

- Evans, George W. and Honkapohja, Seppo. "Adaptive Forecasts, Hysteresis and Endogenous Fluctuations." Federal Reserve Bank of San Francisco *Economic Review*, 1993, 1, pp. 3-13.
- Evans, George W. and Honkapohja, Seppo. *Learning and Expectations in Macroeconomics*. Princeton, NJ: Princeton University Press, 2001; <https://doi.org/10.1515/9781400824267>.
- Evans, George W.; Honkapohja, Seppo; Sargent, Thomas J. and Williams, Noah. "Bayesian Model Averaging, Learning, and Model Selection," in T.J. Sargent and J. Vilmunen, eds., *Macroeconomics at the Service of Public Policy*. New York: Oxford University Press, 2013, pp. 99-119; <https://doi.org/10.1093/acprof:oso/9780199666126.003.0007>.
- Fudenberg, Drew, and Levine, David K. "Self-Confirming Equilibrium." *Econometrica*, May 1993, 61(3), pp. 523-45; <https://doi.org/10.2307/2951716>.
- Geweke, John. *Complete and Incomplete Econometric Models*. Princeton: Princeton University Press, 2010; <https://doi.org/10.1515/9781400835249>.
- Gilboa, Itzhak; Postlewaite, Andrew W. and Schmeidler, David. "Probability and Uncertainty in Economic Modeling." *Journal of Economic Perspectives*, Summer 2008, 22(3), pp. 173-88; <https://doi.org/10.1257/jep.22.3.173>.
- Hansen, Lars P. "Uncertainty Outside and Inside Economic Models." *Journal of Political Economy*, October 2014, 122(5), pp. 945-87; <https://doi.org/10.1086/678456>.
- Hansen, Lars P. and Sargent, Thomas J. *Robustness*. Princeton: Princeton University Press, 2008.
- Kalai, Ehud and Lehrer, Ehud. "Rational Learning Leads to Nash Equilibrium." *Econometrica*, September 1993, 61(5), pp. 1019-45; <https://doi.org/10.2307/2951492>.
- Kandori, Michihiro; Mailath, George J. and Rob, Rafael. "Learning, Mutation and Long Run Equilibria in Games." *Econometrica*, January 1993, 61(1), pp. 29-56; <https://doi.org/10.2307/2951777>.
- Kurz, Mordecai. "On the Structure and Diversity of Rational Beliefs." *Economic Theory*, November 1994, 4(6), pp. 877-900; <https://doi.org/10.1007/BF01213817>.
- Lucas, Robert E. Jr. "Expectations and the Neutrality of Money." *Journal of Economic Theory*, April 1972, 4(2), pp. 103-24; [https://doi.org/10.1016/0022-0531\(72\)90142-1](https://doi.org/10.1016/0022-0531(72)90142-1).
- Perron, Pierre and Shiller, Robert J. "Testing the Random Walk Hypothesis: Power versus Frequency of Observation." *Economics Letters*, 1985, 18, pp. 381-86; [https://doi.org/10.1016/0165-1765\(85\)90058-8](https://doi.org/10.1016/0165-1765(85)90058-8).
- Sargent, Thomas J. *Bounded Rationality in Macroeconomics*. Oxford and New York: Oxford University Press and Clarendon Press, 1993.
- Timmermann, Allan. "Forecast Combinations," in G. Elliott, C. Granger and A. Timmermann, eds., *Handbook of Economic Forecasting*. Volume 1. Amsterdam: Elsevier, 2006, pp. 135-196; [https://doi.org/10.1016/S1574-0706\(05\)01004-9](https://doi.org/10.1016/S1574-0706(05)01004-9).