Red Herrings and Revelations: The Destabilizing and Stabilizing Effects of Economic Theory

Paul Shea[1]

University of Kentucky

This Version: October 25, 2007

Abstract

This paper assumes that output depends on a large number of variables, but that economic theory has only revealed a subset of these variables. The variables that determine output include exogenous shocks, as well as endogenous variables that depend on both the exogenous variables and agents' expectations. This paper examines a policy maker's attempts to forecast output under two different types of expectations formation. I first analyze the case where the policy maker knows the structural coefficients associated with all known variables, and uses this knowledge to form expectations. Under these *structural coefficients expectations*, if economic theory reveals a new variable, then welfare will improve with probability between one-half and one. Under adaptive learning, however, the revelation of a new variable may worsen welfare more often than it improves welfare. This scenario occurs when the policy maker already knows most of the variables that determine output and when the model includes significant endogeneity. Furthermore, the policy maker is typically better off choosing to abandon structural coefficients expectations in favor of adaptive learning, regardless of how many variables theory has revealed.

JEL Classification: E27, E61, D81, D83.

Keywords: Adaptive Learning, Underparameterization, Bounded Rationality, Forecasting.

[1]E-mail:paul.shea@uky.edu

1. Introduction

The baseline approach for modeling the formation of expectations in macroeconomics is the assumption of rational expectations. Under rational expectations, agents know the model's reduced form solution, and use that solution to derive the mathematically optimal expectation. This likely requires that agents know the "true" model that has generated the data, are able to correctly calibrate that model, and can solve for its reduced form. The lengthy literature on adaptive learning criticizes rational expectations for these strong informational requirements. Adaptive learning typically relaxes these requirements by assuming that agents know which variables appear in a model's reduced form, but must estimate the reduced form coefficients through standard econometric techniques.[2]

Many popular macroeconomic structural models yield reduced form solutions that depend on only a handful of variables.[3] The assumption that agents know the exact set of variables in a model's reduced form, common in the adaptive learning literature, may therefore seem plausible. The simplicity of contemporary macroeconomic models, however, is likely due to a desire for tractability and clarity. A true model is likely to depend upon a larger set of variables, some of which are unknown or not emphasized in the literature. Like rational expectations, standard adaptive learning is therefore vulnerable to criticism for endowing its agents with excessive information.

Several papers examine this issue by modeling adaptive learning where agents use underparameterized models. Cho and Kasa (2006) examine the Sargent "Conquest" (1999) model where agents use adaptive learning to estimate the endogenous data generating process. In addition to uncertainty abut the coefficients, agents are also unsure of the correct specification. Agents therefore perform tests among a set of timeless, underparameterized models. Branch and Evans (2006) examine a cobweb model where agents choose among a set of underparameterized specifications.

[2]See Evans and Honkapohja (2001) for a detailed survey and discussion of the adaptive learning literature.

[3]In the ubiquitous New Keynesian literature, for example, the entire economy in reduced form often consists only of output, inflation, a short term interest rate, and a small set of fundamental shocks. For details on the New Keynesian model, see Woodford (2003).

The authors demonstrate that equilibria may occur where agents heterogeneously rely on different specifications. Evans and Ramey (2006) assume that agents use underparameterized, adaptive expectations in a model of the New Keynesian Phillips Curve. They allow agents to select the optimal weights on the previous observation versus the sample mean and show that the Lucas Critique (1976) still applies to much of the parameter space. Each of these papers assumes that agents choose among a set of underparameterized models with a timeless set of regressors.

The present paper departs from the related literature by assuming that an underlying economic theory process gradually reveals the significance of variables to policy makers. Agents therefore forecast using underparametrized specifications, but the degree of misspecification is diminishing over time.

This paper uses a simplified version of the Sargent "Conquest" (1999) model where the policy maker's objective is equivalent to making the best forecast of output. Output depends on a set of exogenous variables and a set of endogenous variables. The vector of endogenous variables depends on both the set of exogenous variables and agents' expectation of output. At any time, economic theory has revealed only subsets of both the exogenous and endogenous variables.

I consider two types of expectations formation. The first type is *structural coefficients expectations* where economic theory reveals not only a subset of relevant variables, but also the exact coefficients for all known variables in the structural model. Expectations are not fully rational, however, because agents do not know any coefficients associated with unrevealed variables. The second type of expectations formation is adaptive learning where agents use all revealed variables as regressors to estimate output.

This paper examines two questions related to the revelation of a new variable. The first question is whether a revelation improves the policy maker's forecasts and therefore welfare. Under structural coefficients expectations, the revelation of a new variable increases welfare with probability between one-half and one. Policy makers are therefore generally better off using newly revealed variables to forecast. There exists, however, a significant (though less than one-half) probability that a revelation may actually decrease welfare. I refer to this case as a *red herring*.[4] Under adaptive learning, the probability of a red herring depends on agent's prior state of knowledge. If theory has revealed a small number of variables, then the probability of a red herring is less than one-half. If the model includes endogeneity and theory has revealed a large number of variables, however, then the probability of a red herring will typically be greater than one-half. In this case, the revelation of a new variable likely worsens the policy maker's forecasts and overall welfare.

The adaptive learning literature typically assumes that agents must use learning because they lack the information needed to form rational expectations. The second question in this paper is whether agents are better off choosing to discard structural coefficients expectations in favor of adaptive learning? The results show that the probability of adaptive learning dominating structural coefficients expectations is always significantly greater than one-half. This result suggests that when theory reveals both a new variable and how that variable interacts with the other known variables, the policy maker should ignore the latter information and estimate the model itself.

The paper is organized as follows. Section 2 develops the model and examines the revelation of new variables under structural coefficients expectations. Section 3 studies the model's behavior under adaptive learning. Finally, Section 4 concludes.

2. The Model and Structural Coefficients Expectations

This paper relies on a version of Sargent's (1999) "Conquest Model":

$$u_t = u^* - \theta(y_t - x_t) \tag{1}$$

Equation 1 is the New Keynesian Phillips Curve. The policy maker chooses a policy, x_t, that determines the unemployment rate, u_t. The policy maker is unable to observe the partially exogenous variable, y_t, which is assumed to depend on a large number of observable variables.

[4] A red herring is a metaphor used to describe an object that distracts an investigation, diverts attention to a side issue, or provides useless but confusing information. Its origins date to pre-1900 England where a herring, reddened by salting and smoking, was used to confuse hounds pursuing a fox or other prey. See Quinion (2002) for more details.

$$y_t = Az_t + Bg_t \tag{2}$$

$$z_t = Cy_t^e + Dg_t \tag{3}$$

g_t is a *M x 1* vector of exogenous variables. Each element of g_t is independently and identically distributed, $N(0, \sigma_g^2)$. For simplicity, I assume that g_t exhibits no serial correlation. z_t is a *N x 1* vector of endogenous variables that depends on agent's expectation of y_t, as well as the exogenous variables. I assume that all agents, including the policy maker, possess identical expectations. C is a *N x 1* vector that describes the feedback between the endogenous variables and y_t^e. D is a *N x M* matrix that describes how the exogenous variables affect the endogenous variables. All agents observe only the expectation of y_t, not its actual value. I assume that endogeneity results from agents' behavior that is based on their expectations. z_t therefore depends on y_t^e, and not y_t.

The timing of the model works as follows: first, g_t is drawn. Second, z_t and y_t^e are simultaneously determined. Finally, y_t is determined based on g_t and z_t.

I focus on the model where the policy maker simply attempts to stabilize unemployment around u^*. This entails setting $x_t = y_t^e$. The policy maker's problem therefore reduces to forming the best forecast of y_t.

Agents generally do not know all of the variables that affect y_t. At time *t*, economic theory has revealed the first *n* elements of z_t and the first *m* elements of g_t. In this section, I assume that theory has also revealed A^n (the first *n* elements of A), and B^m (the first *m* elements of B). Agents use this knowledge to form precise, but biased expectations of y_t. Because the policy maker does not know the full model, expectations are not fully rational. I therefore refer to this assumption as *structural coefficients expectations*. Later, I will assume that agents must use adaptive learning, where they rely on standard econometric techniques to estimate A^n and B^m. This modification will reduce bias, but will add noise to agents' expectations.

$$y_t^e = A^n z_t^n + B^m g_t^m \tag{4}$$

$$z_t^n = C^n y_t^e + D^n g_t \tag{5}$$

Substituting Equation 5 into Equation 4 yields the policy maker's expectation.

$$y_t^e = (1 - A^n C^n)^{-1}[A^n D^n g_t + B^m g_t^m] \tag{6}$$

I assume that the policy maker's loss, ℓ_t, depends on the squared gap between unemployment and u^*.[5]

$$\ell_t = (u_t - u^*)^2 = \theta^2(y_t^e - y_t)^2 \tag{7}$$

Without loss of generality, I set $\theta = 1$. The policy maker's loss may then be re-stated in terms of its knowledge (*n* and *m*) and the vector of exogenous shocks, g_t.

$$\ell_t = [((AC - I)(I - A^n C^n)^{-1} A^n D^n + AD + B)g_t ...$$

$$((AC - I)(I - A^n C^n)^{-1})B^m g_t^m]^2 \tag{8}$$

Proposition 1: If agents are fully rational ($m = M$ and $n = N$), then the loss, ℓ_t, equals zero.
Proof: If ($m = M$ and $n = N$), then $A^n = A$, $C^n = C$, and $B^m = B$. Equation 8 then reduces to $\ell_t = 0$.

Proposition 1 demonstrates that the policy maker's loss results from departures from full rationality. Structural coefficients expectations are one such departure, and I now examine how this assumption affects welfare.

In this section, I assume that agents use structural coefficients expectations instead of adaptive learning. Suppose, however, that agents did use adaptive learning to estimate y_t and that economic theory reveals an additional endogenous variable. Further suppose that agents are able to retroactively regress y_t on the new set of variables, $[g_t^m, z_t^n]$, including the newly revealed variable. Because this revelation changes the way agents form expectations, the Lucas Critique applies, and the retroactive regression coefficients will be biased relative to their values after the revelation.[6]

[5] The results of this paper are robust to any alternative assumption that also implies that the policy maker's loss is increasing in the gap between u_t and u^*.

[6] This bias is in addition to the bias that will result from omitted variables.

Under structural coefficients expectations, however, agents are assumed to rely on invariant structural coefficients to form their expectations, and the Lucas Critique does not apply. To simplify the analysis, I therefore examine structural coefficients expectations where $C = 0$, where 0 is a null vector. The appendix shows that the results that follow in this section are largely unchanged for cases where $C \neq 0$.

Setting $C = 0$, I re-write the policy maker's loss under structural coefficients expectations:

$$\ell_t = [(AD^{N,m} - A^n D^{n,m})g_t^m + (AD^{N,-m} - A^n D^{n,-m} + B^{-m})g_t^{-m}]^2 \tag{9}$$

where g_t^{-m} is a *1 x (M - m)* vector of unrevealed exogenous shocks, and $D^{N,m}$ is the first *N* rows and first *m* columns of D. Consider the loss caused by a single unknown exogenous variable, g_t^{m+1} where $g_t^i = 0$ for all $i \neq m+1$.

$$\ell_t^i = [(AD^{N,m+1} - A^n D^{n,m+1} + B^{m+1})g_t^{m+1}]^2 \tag{10}$$

Each exogenous variable affects y_t directly through the B matrix and indirectly through the endogenous variables. The total gap between y_t and its expectation caused by g_t^{m+1} is the sum of the direct effect (B^{m+1}) and indirect effect ($AD^{N,m+1} - A^n D^{n,m+1}$). The revelation of the $m+1^{th}$ exogenous variable reveals the direct effect to the policy maker. Because the direct effect of g_t^{m+1} is part of its total effect, this revelation is likely to provide useful information about the total effect and improve forecasting. With probability less than one-half, however, the direct effect may not be representative of the total effect and may worsen the policy maker's forecasts, resulting in a red herring. Similarly, the revelation of the n^{th} endogenous variable will likely increase welfare but will also result in a red herring with probability less than one-half.

Tedious but straightforward manipulation of Equation 10 allows the policy maker's expected loss for any [A,B,D] to be re-stated as a function of the individual elements of the relevant matrices.

$$E[\ell_t] = \sum_{j=1}^{m} \sigma_g^2 [\sum_{i=n+1}^{N} \sum_{k=n+1}^{N} A^{1,i} D^{i,j} A^{1,k} D^{k,j}] + ...$$

$$\sum_{j=m+1}^{M} \sigma_g^2 [\sum_{i=n+1}^{N} \sum_{k=n+1}^{N} A^{1,i} D^{i,j} A^{1,k} D^{k,j} + 2 \sum_{i=n+1}^{N} A^{1,i} D^{i,j} B^{1,j} + (B^2)^{1,j}] \tag{11}$$

Suppose that economic theory reveals the m^{th} exogenous variable. Equation 11 demonstrates that the change in the expected loss equals:

$$E[\ell_t]^{m+1,n} - E[\ell_t]^{m,n} = -\sigma_g^2[(B^2)^{1,m+1} + 2B^{1,m+1}\sum_{i=n+1}^{N} A^{1,i}D^{i,m+1}] \quad (12)$$

Likewise, if economic theory reveals the n^{th} endogenous variable, the change in the expected loss equals:

$$E[\ell_t]^{m,n+1} - E[\ell_t]^{m,n} = -\sigma_g^2[\sum_{j=1}^{M}((A^{1,n+1}D^{n+1,j})^2 + 2A^{1,n+1}D^{n+1,j}\sum_{k=n+2}^{N} A^{1,k}D^{k,j})...$$

$$+2\sum_{j=m+1}^{M} A^{1,n+1}D^{n+1,j}B^{1,j}] \quad (13)$$

I assume that agents know, ex-ante, the distributions of each element of [A, B, D]. Each element of [A, B, D] is independently and identically distributed, $N(0, \sigma_q^2)$, where $q = A, B, D$. I further assume orthogonality across matrices by using the following variance-covariance matrix:

$$\begin{bmatrix} \sigma_a^2 & 0 & 0 \\ 0 & \sigma_b^2 & 0 \\ 0 & 0 & \sigma_d^2 \end{bmatrix} \quad (14)$$

Proposition 2: Ex-ante, the revelation of the m^{th} exogenous variable results in a decreased expected loss.

Proof: Under the assumption of orthogonality of Equation 14, Equation 12 reduces to $E[\ell_t]^{m+1,n} - E[\ell_t]^{m,n} = -\sigma_g^2\sigma_b^2 < 0$

Proposition 3: Ex-ante, the revelation of the n^{th} endogenous variable results in a decreased expected loss.

Proof: Under the assumption of orthogonality of Equation 14, Equation 13 reduces to $E[\ell_t]^{m+1,n} - E[\ell_t]^{m,n} = -\sigma_g^2\sigma_a^2\sigma_d^2 < 0$

Propositions 2 and 3 show that, ex-ante, the policy maker is always better off relying on a newly discovered variable because it is likely to result in a welfare improvement. Equations 12 and 13 demonstrate, however, that for specific draws of [A, B, D], newly revealed variables may result in an increased expected loss. Both newly revealed exogenous and endogenous variables may therefore be red herrings.

Proposition 4: For exogenous variables, the probability of a red herring approaches zero as $\sigma_b^2 \to \infty$, $\sigma_a^2 \to 0$, or $\sigma_d^2 \to 0$.

Proof: By Equation 12, as $\sigma_b^2 \to \infty$, $E[\ell_t]^{m+1,n} - E[\ell_t]^{m,n} \to -\infty$. Likewise, as $\sigma_a^2 \to 0$ or $\sigma_d^2 \to 0$, $E[\ell_t]^{m+1,n} - E[\ell_t]^{m,n} \to -\sigma_b^2 < 0$

Proposition 5: For exogenous variables, the probability of a red herring approaches 1/2 as $\sigma_b^2 \to 0$, $\sigma_a^2 \to \infty$, or $\sigma_d^2 \to \infty$.

Proof: By Equation 12, if $\sigma_b^2 \to 0$, $\sigma_a^2 \to \infty$, or $\sigma_d^2 \to \infty$, then $E[\ell_t]^{m+1,n} - E[\ell_t]^{m,n} \to 0$. Because of the model's symmetry, $prob(E[\ell_t]^{m+1,n} - E[\ell_t]^{m,n} > 0) \to 1/2$.

Proposition 4 considers three cases. If each element of B has a large variance, then the direct effect of the m^{th} exogenous variable dominates its indirect effect. The revelation of the direct effect which accompanies the revelation of the m^{th} exogenous variable is therefore necessarily informative about its total effect. If A is a null matrix, then y_t does not depend on z_t and the direct and total effects of g^m are one and the same. Finally, if D is a null matrix, then z_t always equals zero and there is no indirect effect.[7]

Proposition 5 considers the inverse cases. Recall that agents know the distribution of [A, B, D]. If $\sigma_b^2 \to 0$, then the revelation that an element of B equals zero is uninformative and does not affect forecasting. As $\sigma_b^2 \to 0$, however, it is also the case that the welfare loss associated with a red herring approaches zero. If $\sigma_a^2 \to \infty$ or $\sigma_d^2 \to \infty$, then the indirect effect of the revelation of g_t^m dominates its direct effect. In this case, the welfare loss associated with a red herring approaches infinity as its probability approaches one-half.

Proposition 6: For endogenous variables, the probability of a red herring approaches zero as $\sigma_a^2 \to \infty$ or as $\sigma_d^2 \to \infty$.

[7] This latter result is dependent on the assumption that $C = 0$.

Proof: By Equation 13, as $\sigma_a^2 \to \infty$ or as $\sigma_d^2 \to \infty$, $E[\ell_t]^{m+1,n} - E[\ell_t]^{m,n} \to -\infty$.

Proposition 7: For endogenous variables, the probability of a red herring approaches one-half as $\sigma_a^2 \to 0$, $\sigma_d^2 \to 0$, or as $\sigma_b^2 \to \infty$.

Proof: By Equation 13, if $\sigma_a^2 \to 0$, $\sigma_d^2 \to 0$,or $\sigma_b^2 \to \infty$, then $E[\ell_t]^{m+1,n} - E[\ell_t]^{m,n} \to 0$. Because of the model's symmetry, $prob(E[\ell_t]^{m+1,n} - E[\ell_t]^{m,n} > 0) \to 1/2$.

Proposition 6 is the analog to Proposition 4. It considers cases where the importance of the endogenous variables dominates that of the exogenous variables. In these cases, the information conveyed by the revelation of an endogenous variable is necessarily informative and red herrings cannot occur. From Proposition 4, the probability of a red herring occurring on an exogenous variable approaches zero as $\sigma_b^2 \to \infty$. It is not the case, however, that the probability of a red herring on an endogenous variable approaches one-half as $\sigma_b^2 \to 0$. Proposition 7 considers three cases where the importance of the exogenous variables dominates that of the endogenous variables. In these cases, the revelation of an endogenous variable is not informative and the probability of a red herring approaches one-half. If $\sigma_a^2 \to 0$ or $\sigma_d^2 \to 0$, then the welfare loss associated with a red herring approaches zero. If $\sigma_b^2 \to \infty$, however, then the welfare loss associated with a red herring also approaches infinity.

I now simulate the model to quantify the probability of a red herring when an exogenous or endogenous variable is revealed to policy makers. For these simulations, I assume that $N = M = 30$. In the baseline case, I set $\sigma_a^2 = \sigma_b^2 = \sigma_d^2 = 0.1$. I then take 20,000 random draws from [A, B, D] and evaluate the likelihood of a red herring for all possible combinations of *m* and *n*. To access the effect of the variance terms on the model, I also simulate low variance cases where σ_a^2, σ_b^2, or σ_d^2 equal 0.02 and high variance cases where σ_a^2, σ_b^2, or σ_d^2 equal 0.50. Table 1 displays the probability of a red herring when the 15^{th} exogenous variable is revealed for each possible value of *n*.

(FIGURE 1 HERE)

Figure 1 plots the likelihood that the 15^{th} exogenous variable is a red herring for different values of *n*. The middle line represents the baseline case. The probability of a red herring approaches zero as *n* approaches N. The top line represents the calibration where σ_B^2 is low (0.02), and the bottom line represents the calibration where σ_B^2 is high (0.5). As predicted by Propositions 4 and 5, the probability of a red herring decreases as σ_B^2 increases. Although not shown in either Figure 1 or Figure 2, the behavior of the high σ_A^2 and σ_D^2 calibrations are nearly identical to the low σ_B^2 case. Likewise, the low σ_A^2 and σ_D^2 calibrations are nearly identical to the high σ_B^2 case.

(FIGURE 2 HERE)

Figure 2 graphs the likelihood that the m^{th} exogenous variable is a red herring for $n = 15$. Once again, higher values of σ_B^2 decrease the likelihood of a red herring.

(FIGURE 3 HERE)

Figure 3 plots the likelihood that the 15^{th} endogenous variable is a red herring for different vales of *m*. The likelihood of a red herring is significantly higher when economic theory reveals an endogenous variable than when it reveals an exogenous variable. As predicted by Proposition 7, higher values of σ_B^2 now increase the likelihood of a red herring.[8]

(FIGURE 4 HERE)

[8] As for exogenous variables, the model with high values of σ_A^2 or σ_D^2 usually behave like it does for low values of σ_B^2 and vice versa. One exception, however, is the calibration where σ_D^2 is low. That calibration yields significantly lower probabilities of red herrings than any other calibration.

Figure 4 graphs the likelihood that the n^{th} endogenous variable is a red herring for $m = 15$. The results are similar to those shown in Figure 3.[9]

3. Adaptive Learning

Under structural coefficients expectations, when economic theory reveals a new variable, it also reveals the associated structural coefficients. This is reminiscent of rational expectations where agents are assumed to know a model's exact reduced form solution and use that solution to form their expectations.

The adaptive learning literature often argues that the informational requirements of rational expectations are too extreme. An infinite number of structural models may generate any reduced form. To know the coefficients in the reduced form, agents likely must know which model has generated the data, know the structural coefficients, and be able to solve for its reduced form solution. Adaptive learning relaxes these informational requirements by assuming that agents must use standard econometric techniques to estimate the model's reduced form. It is therefore typically the case that agents *must* use adaptive learning because they simply lack the information to use rational expectations. If it is unrealistic to assume that economic theory reveals the structural coefficients associated with a newly revealed variable, then the standard justification for assuming adaptive learning may certainly be applied to the present paper.

There is, however, an additional justification for adaptive learning in this paper. In this model, agents would ideally use rational expectations where they know the structural coefficients on all $N + M$ endogenous and exogenous variables. The assumption that agents know only a fraction of these variables, however, implies that agents cannot use rational expectations, but must instead rely on structural coefficients expectations. Whereas rational expectations are certain to yield better forecasts than adaptive learning, it is possible that adaptive learning will outperform structural coefficients expectations. In this case, agents are better off *choosing* to use adaptive learning

[9]As in Figure 3, the calibration where σ_D^2 is low yields significantly lower probabilities of red herrings than any other calibration.

instead of structural coefficients expectations.

This section addresses two questions. First, how likely are red herrings under the assumption that agents must use adaptive learning? Second, how often would agents optimally choose to discard structural coefficients expectations in favor of adaptive learning?

Under rational expectations, agents form more accurate expectations than under adaptive learning. The possibility that agents may be better off under adaptive learning arises only because agents may possess structural coefficients expectations, not rational expectations. Several papers develop different cases where agents may prefer adaptive learning to rational expectations. The most common approach allows agents to possess rational expectations if and only if they incur an additional cost.[10] If the cost of rational expectations is sufficiently high, then some or all agents will prefer adaptive learning to rational expectations. Adam (2005) develops a model where agents learn adaptively and choose between a correctly specified model (relative to rational expectations) and an underparameterized model. If agents use the former specification, adaptive learning asymptotically converges to rational expectations. If agents use the latter specification, however, the learning process causes both models to be misspecified and it is possible that the latter model yields a better forecast on average. Learning the misspecified model is optimal, however, only if agents have already been using that model. Agents would have been better off had they used the first specification all along.

To model adaptive learning, I assume that agents know the form of the structural model, but do not know the exact coefficients.

$$y_t^e = a z_t^n + b g_t^m \tag{15}$$

Under adaptive learning, I assume that agents use recursive least squares to obtain a_t and b_t. This is similar to running an OLS regression of y_t on z_t and g_t, and updating that regression each period as new data becomes available. Equation 15 represents agents' perceived law of motion (PLM).

[10]See Evans and Ramey (1992), Brock and Hommes (1997), and Branch and McGough (2005) for prominent examples from this literature.

By inserting agents' PLM into Equations 2 and 3, it is possible to obtain the economy's actual law of motion (ALM).

$$y_t = ACaz_t^n + (ACb + AD^m + B^m)g_t^m + (AD^{-m} + B^{-m})g_t^{-m} \tag{16}$$

The mapping from the PLM to the ALM may therefore be written as:

$$T\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} ACa \\ ACb + AD^m + B^m \end{bmatrix} \tag{17}$$

Because the ALM depends on variables not included in the PLM, and z_t^n is correlated with the omitted variables, g_t^{-m}, the optimal projection of the ALM on the PLM yields biased regression coefficients. Under structural coefficients expectations, $a = A^n$ and $b = B^m$. Structural coefficients expectations therefore provide noiseless coefficients, but at the expense of ignoring the correlation between z_t^n and g_t^{-m}, and thus being a suboptimal projection. Because the policy maker only cares about forecasting y_t, and is unconcerned about the bias of *a* and *b*, adaptive learning asymptotically dominates structural coefficients expectations. While the learning process is ongoing, however, adaptive learning will be noisier than structural coefficients expectations and the latter may therefore be preferable.[11]

Under adaptive learning, it is not typically the case that *a* and *b* will converge to their structural coefficients expectations, or rationale xpectations values. This is therefore an example of what Evans and Honkapohja (2001) refer to as a *restricted perceptions equilibrium.*[12] A restricted perceptions equilibrium is optimal in the class of PLMs that agents are considering, but may be inferior to other types of PLMs. In this model, a restricted perceptions equilibrium implies that agents are forming the best possible econometric estimate, given that they do not know all the variables included in the model.

To determine the fixed points of the learning process, it is useful to re-state z_t^n as a function of the exogenous variables.

[11]The simulations that compare welfare under adaptive learning and structural coefficients expectations assume constant-gain learning where the learning process is persistent.

[12]Branch and Evans (2006) analyze restricted perceptions equilibria in a fixed, underparameterized model.

$$z_t^n = (I - C^n a)^{-1}[(C^n b + D^{n,m})g_t^m + D^{n,-m}g_t^{-m} \tag{18}$$

where $D^{n,m}$ is the first *n* rows and *m* columns of D, and $D^{n,-m}$ is the first *n* rows and last *M* - *m* columns of D.

To calculate the fixed points of the learning process, I project the ALM onto the PLM. It is convenient to write both z_t^n and y_t as functions of g_t^m and g_t^{-m}.

$$z_t^n = \alpha g_t^m + \beta g_t^{-m}$$

where α and β are given by Equation 18.

$$y_t = \chi g_t^m + \delta g_t^{-m} \tag{19}$$

where $\chi = AC(1 - aC^n)^{-1}(aD^{n,m} + b) + AD^m + B^m$ and $\delta = AC(1 - aC^n)^{-1}aD^{n,-m}AD^{-m} + B^{-m}$.

$$Var\begin{bmatrix} z_t^n \\ g_t^m \end{bmatrix} = \sigma_g^2 \begin{bmatrix} \alpha\alpha' + \beta\beta' & \alpha \\ \alpha & I(m) \end{bmatrix} = \tag{20}$$

$$Cov\begin{bmatrix} \begin{bmatrix} z_t^n \\ g_t^m \end{bmatrix} & , y_t \end{bmatrix} = \sigma_g^2 \begin{bmatrix} \alpha\chi' + \beta\delta' \\ \chi' \end{bmatrix} = \tag{21}$$

The OLS regression is thus given by:

$$T\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \alpha\alpha' + \beta\beta' & \alpha \\ \alpha & I(m) \end{bmatrix}^{-1} \begin{bmatrix} \alpha\chi' + \beta\delta' \\ \chi' \end{bmatrix} \tag{22}$$

Because *a* and *b* appear in in the right hand side of Equation 22, solving for the fixed point is complex. After using Gauss to solve for the fixed point, I begin the learning process at this point. For simplicity, I assume that agents do not include an intercept in their regression. If agents do use an intercept, the fixed point for the intercept equals zero.

The system is stable under adaptive learning if the learning coefficients remain in the neighborhood of their restricted perceptions equilibrium values. To evaluate stability under learning, I use the related concept of E-Stability. Evans and Honkapohja (2001) demonstrate that under general conditions, a model is stable under learning if and only if it is E-Stable. E-Stability maps from the PLM to the ALM using the E-Stability differential equation.

$$d/d\tau \begin{bmatrix} a \\ b \end{bmatrix} = T \begin{bmatrix} a \\ b \end{bmatrix} - \begin{bmatrix} a \\ b \end{bmatrix} \tag{23}$$

If each eigenvalue of the Jacobian of the right hand side of Equation 23 has real parts less than zero when evaluated at its fixed point, then the model is E-Stable. Evaluating this condition using the ALM, Equation 16, and the PLM, Equation 15, shows that the necessary and sufficient condition for E-Stability is that $AC < 1$.

There are two sources of red herrings under adaptive learning. The first source is related to the added noise that an additional regressor adds to the process. It is necessarily true that including an additional regressor results in an equal or better in-sample fit than before its inclusion. Likewise the new restricted perceptions equilibrium performs better than the older equilibrium at their respective fixed points. Often, however, the inclusion of an additional regressor will introduce enough noise to the estimate so that the enlarged set of regressors provides a worse out of sample fit than the original set of regressors. In this case, forecasts worsen, welfare decreases, and a red herring occurs. Later simulations will demonstrate, however, that this source, by itself, results in red herrings with probability less than one-half.

If the model includes endogeneity, then the Lucas Critique is a second source of red herrings. I assume that once theory reveals the n^{th} endogenous variable, agents are able to collect past data for that variable and regress y on z^{n+1} and g^m. Because y_t depends on its expectation, however, the discovery of an additional endogenous variable changes that underlying statistical relationship. Agents' initial estimates of a_t and b_t may therefore be significantly different than their values in the new restricted perceptions equilibrium. While the learning process returns a_t and b_t to the neighborhood of their restricted perceptions equilibrium values, agents will make poor forecasts. If their discount factor is sufficiently small, then the revelation of the new variable will worsen intertemporal utility. Later simulations show that the combination of both sources of red herrings will often result in their occurrence with probability greater than one-half.

To simulate adaptive learning, I assume that agents estimate Equation 2 using recursive least squares. Agents' *(n+m) x 1* vector of regressors, $\phi = [a_t, b_t]'$, at time *t* is updated according to:

$$\phi_t = \phi_{t-1} + \gamma R_t^{-1} \phi'_{t-1} (y_{t-1} - a_{t-1} z_{t-1}^n - b_{t-1} g_{t-1}^m) \tag{24}$$

$$R_t = (1 - \gamma) R_{t-1} + \gamma (\phi_{t-1} \phi'_{t-1}) \tag{25}$$

This type of recursive least squares is not identical to running an OLS regression each period. The gain (γ) represents the weight placed on the most recent observation. Under standard OLS, the gain equals the inverse of the sample size and each observation counts equally. Equations 24 and 25 are an example of constant-gain learning where agents place extra weight on more recent observations.[13] Constant-gain learning is a popular way to model learning when the model that agents estimate is subject to structural change, such as the revelation of a new variable. In the absence of constant-gain learning, agents would be even more vulnerable to the Lucas Critique for large sample sizes.

Under structural coefficients expectations, the revelation of a new endogenous variable increases agents' expected loss with probability between zero and one-half. I now examine the welfare effects of discovering a new endogenous variable under adaptive learning. I begin the learning process at the fixed point of the restricted perceptions equilibrium, given by Equation 22. I initialize the learning process by simulating 2500 random draws of g_t while using a gain equal to 0.005. After a 2500 period burn, I calculate the policy maker's loss for 500 additional periods under three different scenarios. In the first scenario, economic theory reveals the $n + 1^{th}$ endogenous variables and agents use this additional variable to forecast y_t. In the second scenario, economic theory does not reveal the new variable and the model's behavior is unchanged from the first 2500 observations. In the third scenario, I assume that economic theory does reveal the new variable but that agents possess structural coefficients expectations as discussed in Section 2. I then calculate the loss under all three cases.

[13]For more details on constant-gain learning, see Sargent (1999), or Evans and Honkapohja (2001).

To limit computational time, I do not simulate every pair of *m* and *n*. Instead I simulate every pair where *m* equals 2,5,8... and *n* initially equals 2,5,8,... I conduct 500 simulations for each pair of *m* and *n*. My baseline calibration assumes $\sigma_j^2 = 0.1$ for $j = A, B, C, D$ and that agents use a discount factor of 0.99. For comparison, I consider several other calibrations. I simulate high variance cases where $\sigma_j^2 = 0.3$, and low variance cases where $\sigma_j^2 = 0.02$. For the low σ_C^2 case, however, I set $\sigma_C^2 = 0.00$ to completely isolate the effect of the Lucas Critique. Tables 1-3 display the probability of a red herring where the revelation of a new endogenous variable decreases welfare.

(Table 1 HERE)

(Table 2 HERE)

(Table 3 HERE)

As shown by Tables 1-3, when agents know a small number of the variables that matter, the probability of a red herring tends to be significant, but less than one-half. In these cases, the inclusion of an additional variable ultimately improves forecasting by enough to overcome both the Lucas Critique and the additional noise that it introduces. As agents know more variables, however, the probability of a red herring rises above one-half in the baseline and most other cases. As agents know more of the model, an additional variable contributes less information and is less likely to overcome the additional noise that it introduces. Additionally, because z_t depends on g_t, the model exhibits significant multicolinearity. As agents know more variables, multicolinearity increases, decreasing the rate of convergence to the new restricted perceptions equilibrium.

The case where $\sigma_C^2 = 0.00$ eliminates all endogeneity and suppresses the Lucas Critique. In this case, the probability of a red herring never exceeds one-half. By contrast, the high endogeneity case where $\sigma_C^2 = 0.3$ results in very high probabilities of red herrings.

These results suggest that the policy maker should include newly revealed variables when its initial knowledge is poor but should avoid using them to forecast when it knows most of the variables that matter. In practice, however, it may be difficult for the policy maker to know how many variables it knows versus how many theory has not yet revealed. Likewise, if the public is aware of the newly revealed variable and is able to form its own expectations, it may not be possible for the policy maker to ignore the newly revealed variable.

It is also of interest to examine whether the policy maker should choose to use adaptive learning if lstructural coefficients expectations are also feasible. Tables 4-6 display the probability that structural coefficients expectations outperform adaptive learning after the revelation of an additional endogenous variable.[14]

(Table 4 HERE)

(Table 5 HERE)

(Table 6 HERE)

Structural coefficients expectations are more precise than adaptive learning but ignore the correlation between z_t^n, and the omitted variables, g_t^{-m}. These results show that, under all scenarios, it is worthwhile for the policy maker to choose adaptive learning and accept the additional noise. If economic theory reveals both a new variable and the associated structural coefficients, then the policy maker is better off ignoring the latter information.

[14]The results do not differ significantly if adaptive learning is compared to structural coefficients expectations without the revelation of a new endogenous variable.

4. Conclusion

For tractability and clarity, macroeconomic models typically include an unrealistically small number of variables. This paper examines the implications of a policy maker having to forecast while only knowing a subset of the relevant variables. I rely on a simple framework where the policy maker's objective is equivalent to forming the best forecast of output while relying on an underparameterized specification.

Under structural coefficients expectations, economic theory is benign in that the revelation of a new variable is always likely to improve forecasting and welfare. Ignoring the question of whether structural coefficients expectations are feasible, however, policy makers should abandon this method of forecasting in favor of adaptive learning because the latter is always likely to outperform the former.

If a policy maker is using adaptive learning, either by choice or necessity, then the role of economic theory is less clear. If most of the relevant variables are unknown, then the revelation of a new variable is likely to improve forecasting and improve welfare. If most of the relevant variables are unknown, however, then the revelation will likely decrease welfare.

References

Adam, K. 2005. "Learning to Forecast and Cyclical Behavior of Output and Inflation." *Macroeconomic Dynamics,* 9: 1-27.

Branch, W, and G. Evans. 2006. "Intrinsic Heterogeneity in Expectation Formation." *Journal of Economic Theory,* 127: 264-295.

Branch, W, and B. McGough. 2005. "Replicator Dynamics in a Cobweb Model with Rationally Heterogeneous Expectations." *Working Paper.*

Brock, W, and C. Hommes. 1997. "A Rational Route to Randomness." *Econometrica,* 65(5): 1059-1095.

Evans, G, and G. Ramey. 2006. "Adaptive Expectations, Underparameterization and the Lucas Critique." *Journal of Monetary Economics,* 53: 249-264.

Evans, G, and G. Ramey. 1992. "Expectation Calculation and Macroeconomic Dynamics." *The American Economic Review,* 82: 207-224.

Evans, G, and S. Honkapohja. 2001. *Learning and Expectations in Macroeconomics*. Princeton, NJ: Princeton University Press.

Cho, I, and K. Kasa. 2006. "Learning and Model Validation." *Working Paper*.

Lucas, R. 1976. "Econometric Policy Evaluation: a Critique." *in K. Brunner and A. Meltzer, eds, The Phillips Curve and the Labor Market, Vol. 1 of the Carnegie-Rochester Conferences in Public Policy, a Supplimental Series to the Journal of Monetary Economics (North Holland-Amsterdam)*: 19-46.

Woodford, M. 2003. *Interest and Prices*. Princeton, NJ and Oxford: Princeton University Press.

Quinion, M. 2002. "The Lure of the Red Herring." *WorldWideWords*.

Sargent, T. 1999. *The Conquest of American Inflation*. Princeton, NJ: Princeton University Press.

Appendix A: Structural Coefficients Expectations with Endogeneity

I now consider the model under structural coefficients expectations with endogeneity where $C \neq 0$ generally. While this addition adds additional complexity to the analysis, the basic results are unaffected. Under structural coefficients expectations, theory reveals not only new endogenous variables, but also the exact nature of the endogeneity. Policy makers can therefore easily account for new endogeneity in their expectations.

With endogeneity, the loss function becomes:

$$\ell_t = [(AD + B)g_t - (AC - 1)(A^n C^n - 1)^{-1}(A^n D^n g_t + B^m g_t^m]^2 \tag{26}$$

The change in the expected loss upon the revelation of the m^{th} exogenous variable equals:

$$E[\ell_t]^{m+1,n} - E[\ell_t]^{m,n} = -\sigma_g^2[(B^2)^{1,m+1} + 2B^{1,m+1} \sum_{i=n+1}^{N} A^{1,i} D^{i,m+1} + ...$$
$$(AC - 1)(A^n C^n - 1)^{-1} B^{1,m+1}(-(AC - 1)(A^n C^n - 1)(2\sum_{i=1}^{n} A^{1,i} D^{i,m+1} + B^{1,m+1}) + ...$$
$$2(\sum_{i=1}^{n} A^{1,i} D^{i,m+1} - sum_{i=n+1}^{N} A^{1,i} D^{i,m+1})] \tag{27}$$

The change in the expected loss upon the revelation of the n^{th} endogenous variable equals:

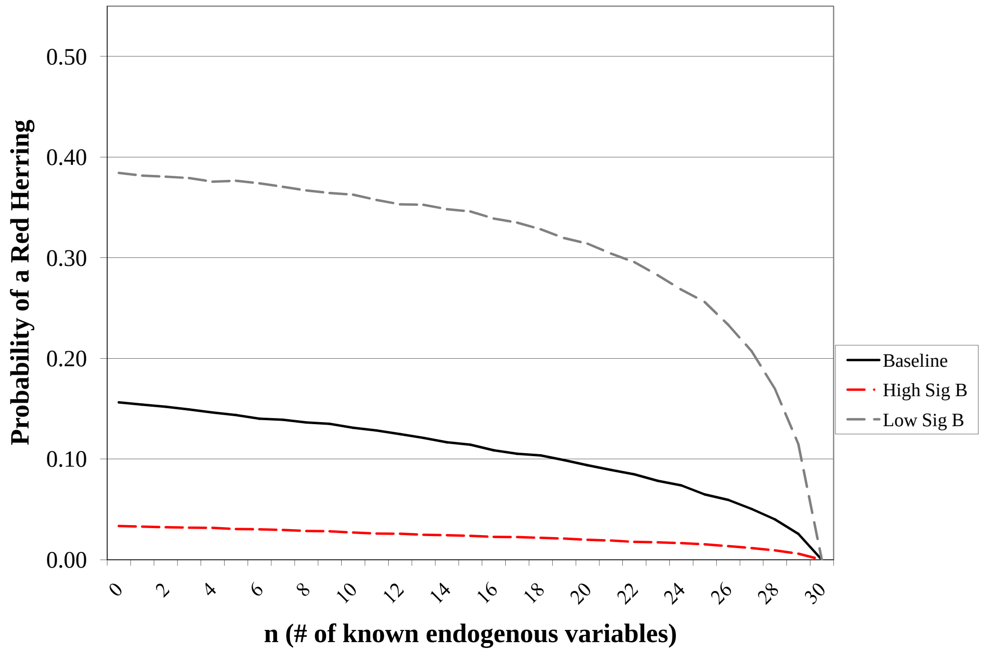
$$E[\ell_t]^{m,n+1} - E[\ell_t]^{m,n} = -\sigma_g^2[\sum_{j=1}^{M}((A^{1,n+1} D^{n+1,j})^2 + 2A^{1,n+1} D^{n+1,j} \sum_{k=n+2}^{N} A^{1,k} D^{k,j})...$$
$$+2\sum_{j=m+1}^{M} A^{1,n+1} D^{n+1,j} B^{1,j}] + (AC - 1)(A^n C^n - 1)((AC - 1)(A^n C^n - 1)(-\sum_{j=1}^{M} A^{1,n+1} D^{n+1,j}...$$
$$(A^{1,n+1} D^{n+1,j} + 2\sum_{i=1}^{n} A^{1,i} D^{1,j}) - 2\sum_{j=1}^{m} B^{1,j} A^{1,n+1} D^{n+1,j}) + ...$$
$$2\sum_{j=1}^{M} A^{1,n+1} D^{n+1,j})(\sum_{1=1}^{n} A^{1,i} D^{1,j} - \sum_{1=n+2}^{N} A^{1,i} D^{1,j} - ...$$
$$2\sum_{j=m+1}^{M} B^{1,j} A^{1,n+1} D^{n+1,j} - 2\sum_{j=1}^{m} B^{1,j} A^{1,n+1} D^{n+1,j})] \tag{28}$$

I repeat the simulations from Section 2 for the model with endogeneity. The results are similar, as reported by Figures 5-8.

(FIGURES 5-8 HERE)

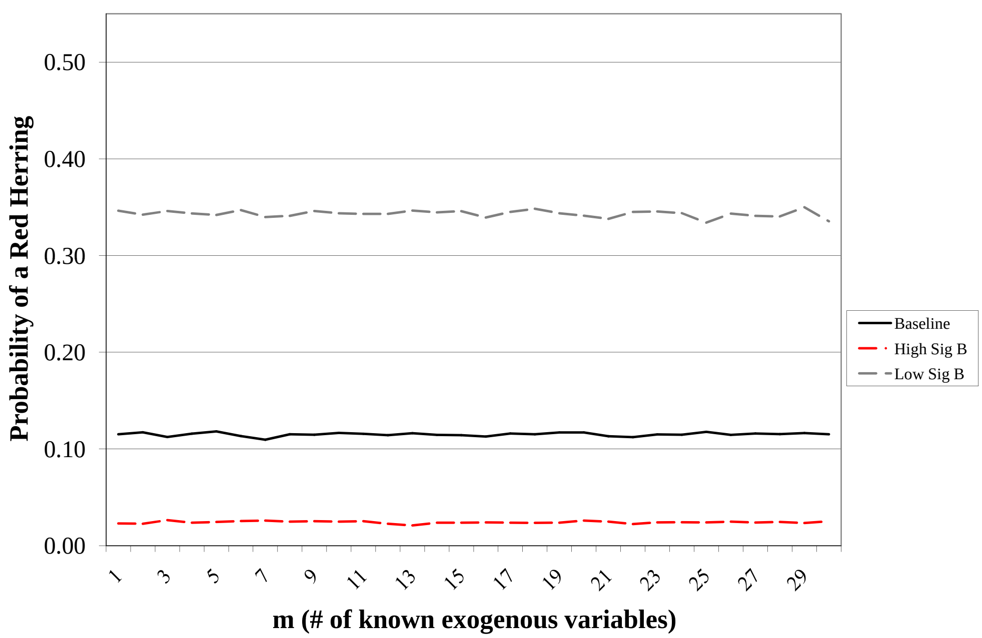
Figures and Tables

Figure 1: Effect of Discovering the Fifteenth Exogenous Variable



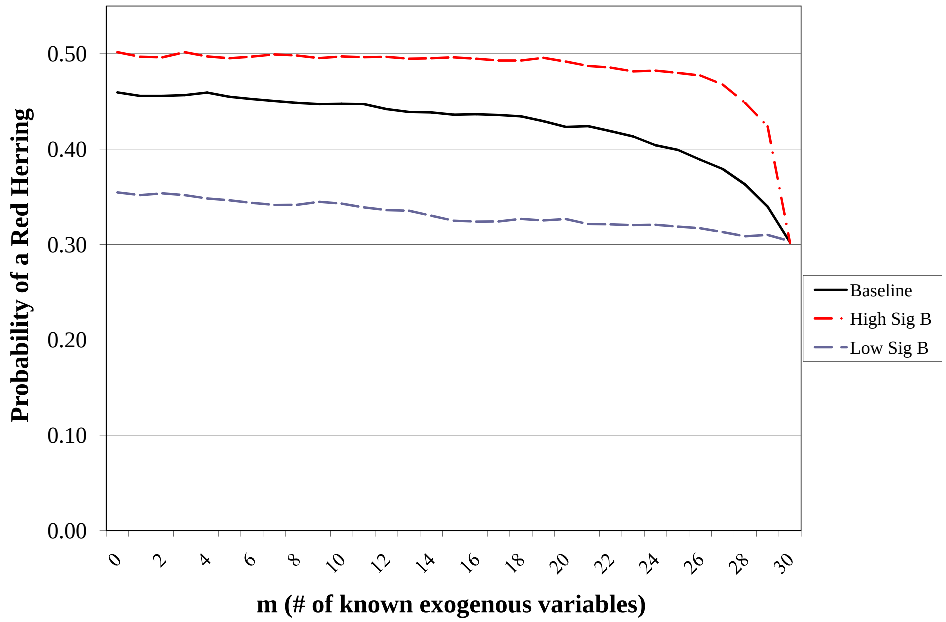
chesches

Figure 2: Effect of Discovering the m*th* Exogenous Variable (n = 15)



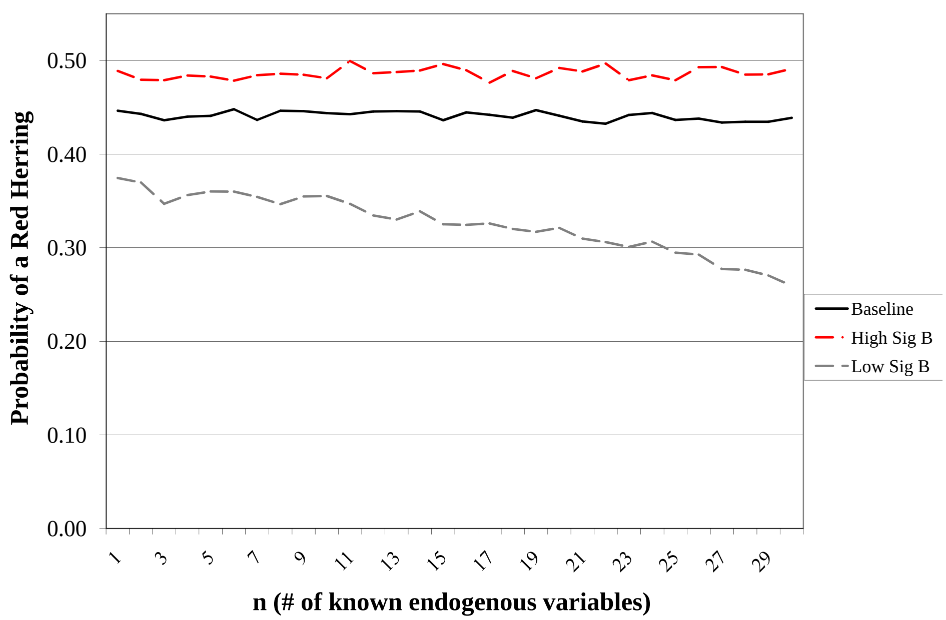
chesches

Figure 3: Effect of Discovering the Fifteenth Endogenous Variable



chesches

Figure 4: Effect of Discovering the n*th* Endogenous Variable (m = 15)



chesches

Table 1

Probability of a Red Herring When the 15^{th} Endogenous Variable is Revealed

		High Variances				**Low Variances**			
m	BL	σ_A^2	σ_B^2	σ_C^2	σ_D^2	σ_A^2	σ_B^2	σ_C^2	σ_D^2
2	0.44	0.40	0.43	0.40	0.46	0.43	0.39	0.40	0.44
5	0.42	0.41	0.47	0.47	0.44	0.41	0.48	0.41	0.50
8	0.43	0.39	0.46	0.46	0.44	0.44	0.46	0.43	0.47
11	0.43	0.42	0.46	0.42	0.51	0.42	0.43	0.47	0.52
14	0.43	0.48	0.46	0.46	0.50	0.49	0.39	0.46	0.44
17	0.44	0.48	0.47	0.47	0.50	0.44	0.45	0.45	0.50
20	0.49	0.45	0.55	0.48	0.53	0.46	0.47	0.49	0.51
23	0.50	0.49	0.52	0.51	0.62	0.49	0.45	0.47	0.56
26	0.53	0.54	0.59	0.54	0.60	0.51	0.52	0.44	0.56
29	0.56	0.62	0.65	0.68	0.67	0.56	0.53	0.49	0.69

Table 2

Probability of a Red Herring When the n^{th} Endogenous Variable is Revealed (m = 14)

		High Variances				Low Variances			
n	BL	σ_A^2	σ_B^2	σ_C^2	σ_D^2	σ_A^2	σ_B^2	σ_C^2	σ_D^2
3	0.30	0.34	0.40	0.35	0.38	0.32	0.35	0.33	0.41
6	0.36	0.36	0.41	0.38	0.41	0.35	0.37	0.36	0.40
9	0.41	0.37	0.42	0.40	0.45	0.40	0.33	0.39	0.46
12	0.43	0.45	0.46	0.46	0.49	0.43	0.43	0.41	0.48
15	0.43	0.48	0.46	0.46	0.50	0.49	0.39	0.46	0.44
18	0.43	0.44	0.50	0.48	0.51	0.48	0.46	0.45	0.50
21	0.47	0.50	0.49	0.51	0.51	0.51	0.47	0.49	0.53
24	0.45	0.53	0.51	0.48	0.55	0.45	0.42	0.46	0.54
27	0.49	0.51	0.53	0.52	0.56	0.50	0.45	0.39	0.54
30	0.30	0.29	0.51	0.48	0.55	0.27	0.07	0.00	0.53

Table 3

Probability of a Red Herring When the n^{th} Endogenous Variable is Revealed (m = n-1)

		High Variances				Low Variances			
n	BL	σ_A^2	σ_B^2	σ_C^2	σ_D^2	σ_A^2	σ_B^2	σ_C^2	σ_D^2
3	0.37	0.32	0.41	0.40	0.37	0.36	0.28	0.33	0.39
6	0.40	0.35	0.43	0.38	0.44	0.41	0.32	0.35	0.39
9	0.39	0.35	0.41	0.43	0.46	0.39	0.34	0.40	0.40
12	0.44	0.44	0.46	0.45	0.49	0.41	0.40	0.40	0.45
15	0.43	0.48	0.46	0.46	0.50	0.49	0.39	0.46	0.44
18	0.44	0.44	0.47	0.50	0.55	0.45	0.46	0.47	0.55
21	0.54	0.49	0.56	0.57	0.58	0.54	0.46	0.44	0.57
24	0.52	0.54	0.58	0.64	0.61	0.52	0.49	0.49	0.62
27	0.55	0.59	0.65	0.65	0.71	0.59	0.50	0.45	0.72
30	0.45	0.45	0.65	0.67	0.72	0.46	0.14	0.00	0.68

Table 4

Probability that Structural Coefficients Expectations Dominate Adaptive Learning (n = 15)

		High Variances				Low Variances			
m	BL	σ_A^2	σ_B^2	σ_C^2	σ_D^2	σ_A^2	σ_B^2	σ_C^2	σ_D^2
2	0.23	0.24	0.32	0.27	0.34	0.24	0.19	0.18	0.32
5	0.24	0.22	0.30	0.32	0.31	0.28	0.20	0.22	0.36
8	0.25	0.22	0.33	0.31	0.33	0.27	0.19	0.17	0.34
11	0.25	0.21	0.33	0.29	0.35	0.25	0.18	0.25	0.37
14	0.26	0.26	0.33	0.32	0.35	0.26	0.16	0.24	0.33
17	0.20	0.25	0.30	0.32	0.35	0.22	0.15	0.21	0.34
20	0.22	0.16	0.32	0.29	0.32	0.22	0.13	0.15	0.33
23	0.21	0.13	0.28	0.25	0.34	0.21	0.09	0.14	0.29
26	0.19	0.12	0.27	0.25	0.30	0.17	0.09	0.09	0.27
29	0.16	0.11	0.33	0.22	0.39	0.32	0.05	0.10	0.38

Table 5

Probability that Structural Coefficients Expectations Dominate Adaptive Learning (m = 14)

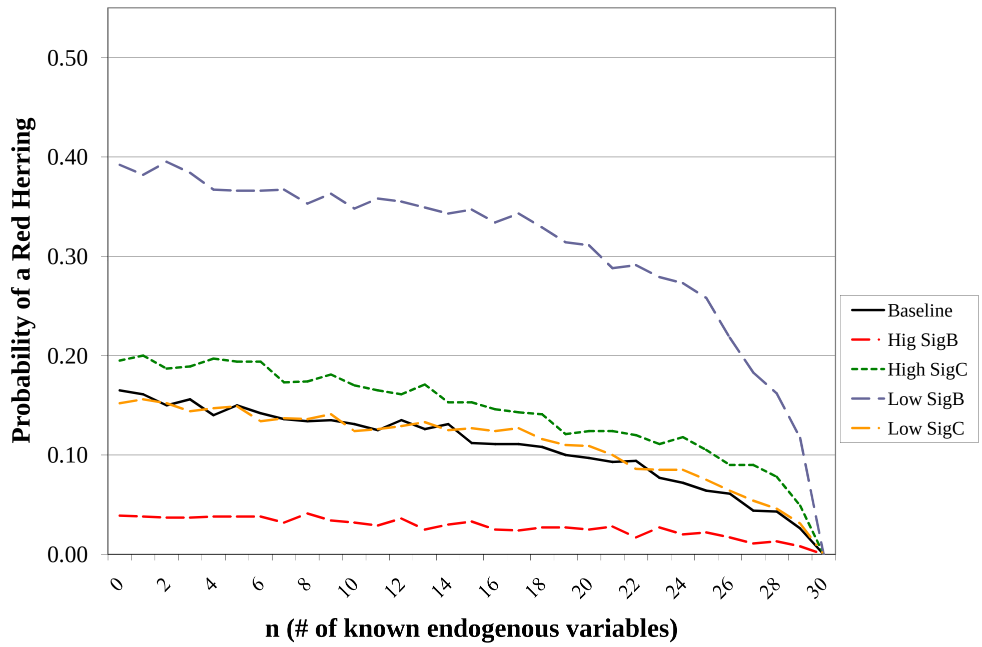
		High Variances				Low Variances			
n	BL	σ^2_A	σ^2_B	σ^2_C	σ^2_D	σ^2_A	σ^2_B	σ^2_C	σ^2_D
3	0.16	0.12	0.34	0.24	0.33	0.27	0.09	0.18	0.40
6	0.18	0.16	0.26	0.24	0.34	0.20	0.12	0.18	0.30
9	0.21	0.20	0.30	0.27	0.34	0.24	0.14	0.18	0.35
12	0.27	0.16	0.33	0.33	0.35	0.24	0.17	0.21	0.33
15	0.26	0.26	0.33	0.32	0.35	0.26	0.16	0.24	0.33
18	0.25	0.19	0.32	0.29	0.33	0.25	0.18	0.22	0.33
21	0.22	0.20	0.26	0.25	0.25	0.22	0.16	0.15	0.28
24	0.16	0.19	0.23	0.22	0.25	0.17	0.10	0.13	0.21
27	0.15	0.15	0.18	0.15	0.17	0.14	0.09	0.06	0.18
30	0.04	0.04	0.04	0.03	0.03	0.04	0.02	0.00	0.04

Table 6

Prob. that Structural Coefficients Expectations Dominate Adaptive Learning (m = n-1)

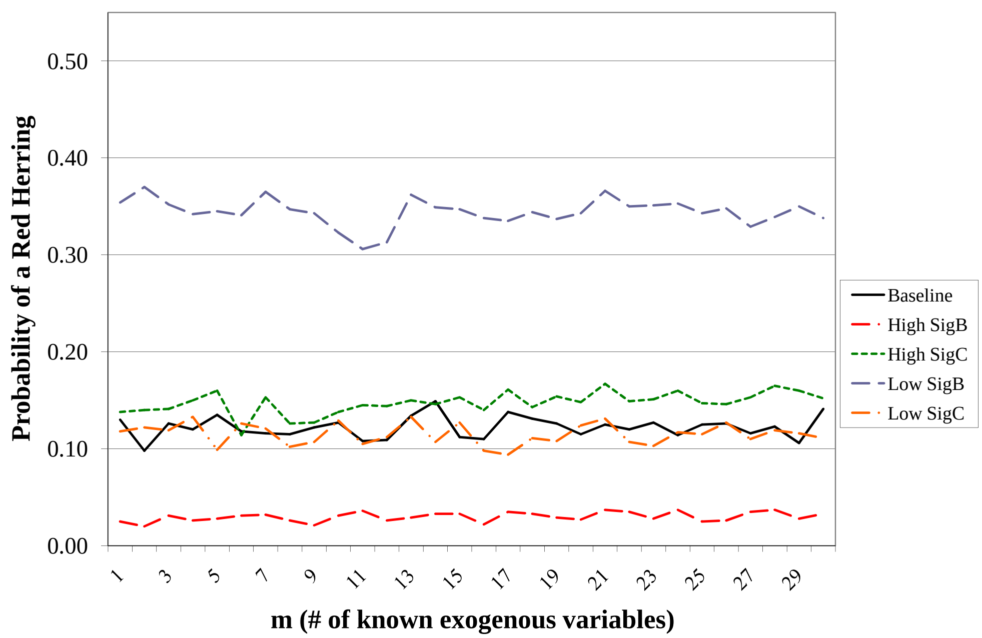
		High Variances				Low Variances			
n	BL	σ_A^2	σ_B^2	σ_C^2	σ_D^2	σ_A^2	σ_B^2	σ_C^2	σ_D^2
3	0.21	0.18	0.28	0.27	0.26	0.23	0.16	0.18	0.31
6	0.22	0.21	0.25	0.24	0.35	0.23	0.11	0.15	0.32
9	0.21	0.17	0.29	0.27	0.36	0.22	0.16	0.19	0.30
12	0.26	0.26	0.35	0.31	0.35	0.25	0.18	0.20	0.34
15	0.26	0.26	0.33	0.32	0.35	0.26	0.16	0.24	0.33
18	0.19	0.20	0.26	0.28	0.34	0.22	0.16	0.18	0.31
21	0.17	0.15	0.23	0.24	0.27	0.17	0.09	0.11	0.23
24	0.15	0.12	0.18	0.22	0.17	0.14	0.07	0.08	0.19
27	0.10	0.11	0.15	0.14	0.15	0.10	0.06	0.06	0.15
30	0.15	0.20	0.18	0.17	0.14	0.16	0.07	0.00	0.14

Figure 5: Effect of Discovering the Fifteenth Exogenous Variable



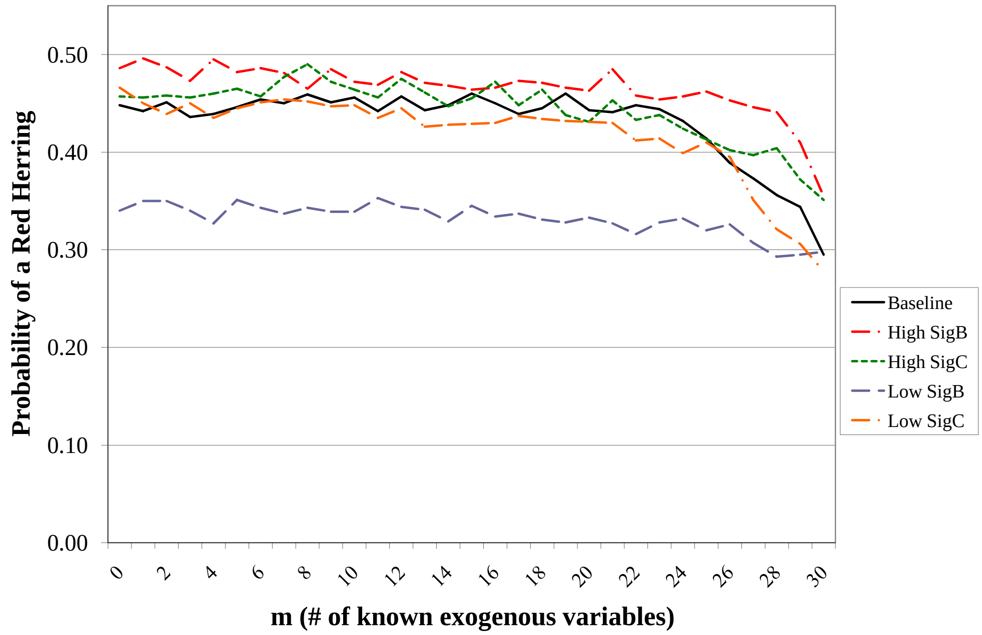
chesches

Figure 6: Effect of Discovering the m*th* Exogenous Variable (n = 15)



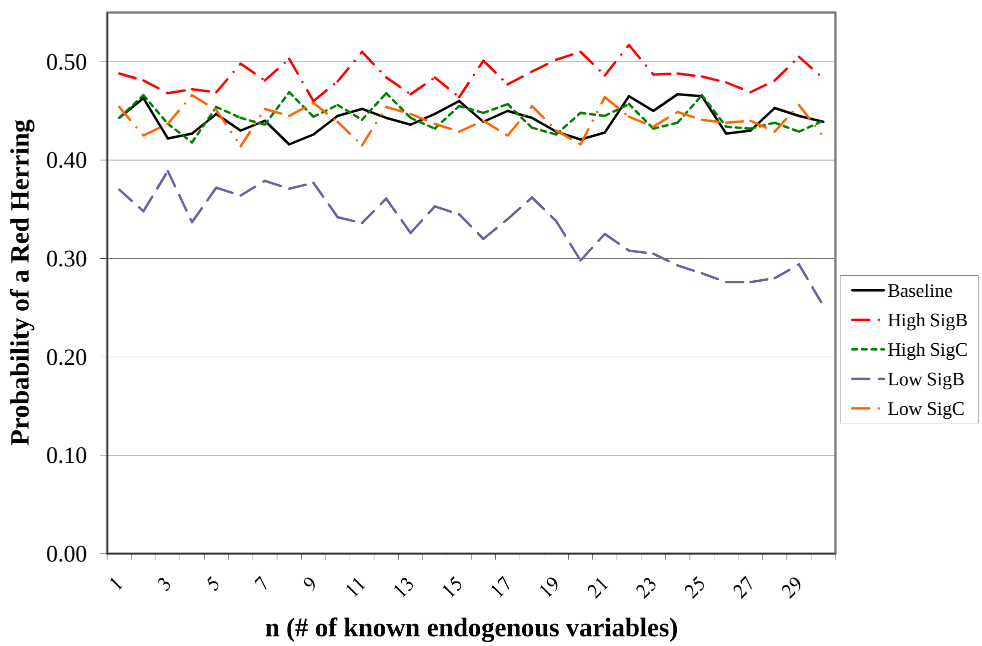
chesches

Figure 7: Effect of Discovering the Fifteenth Endogenous Variable



chesches

Figure 8: Effect of Discovering the n*th* Endogenous Variable (m = 15)



chesches