ROBUSTIFYING SHILLER: DO STOCK PRICES MOVE *ENOUGH* TO BE JUSTIFIED BY SUBSEQUENT CHANGES IN DIVIDENDS?

KURT F. LEWIS AND CHARLES H. WHITEMAN†

ABSTRACT. This paper studies the consequences of a minor change to the present value formula for stock prices. In place of the squared-error-loss minimizing *expected* present value of future dividends, we use a predictor optimal for the min-max preference relationship appropriate in cases of ambiguity. With such "robust" predictions, the variance bound of Shiller (1981) and LeRoy-Porter (1981) is reversed in spectacular fashion in that prices are predicted to be far more volatile than what is observed in the data. We also investigate an intermediate "partially robust" case in which the degree of ambiguity is limited, and discover that such an intermediate model cannot be rejected in favor of an unrestricted time series model with more than twice as many parameters.

1. INTRODUCTION

In 1981, Shiller (1981) and LeRoy and Porter (1981) sparked a controversy whose legacy even today continues to occupy the attention of researchers in economics and finance. Their focus was on the simplest present value model of stock prices: that a stock price corresponds to the expected present value of the stock's dividends discounted at a constant rate. Using the orthogonality property of least squares projections (that projection errors are uncorrelated with information used in constructing the projection), they showed that the *actual* present value must be no less variable than its expectation, which according to the present value model is the current stock price. In the data, this variance bound is violated in dramatic fashion–stock prices are far more variable than subsequent realizations of dividend would appear to permit.

The controversy involves *why* prices are so volatile–are stock prices influenced by fads or "irrational exuberance"? Or, is something amiss with the volatility calculations, the treatment of dividends, or the assumption of a constant discount factor? Much work has been undertaken to pursue each of these latter three resolutions of the puzzle. For example, Flavin (1983) focussed on whether the sampling variability in the volatility calculations could be sufficient to generate the result even when the simple pricing model is true. Regarding dividends, Marsh and Merton (1986) argued that if the dividend process is difference stationary (rather than trend stationary, as Shiller had assumed), the variance bound is reversed. Time series work suggested, however, that the trend-stationary assumption was more plausible than the difference-stationary one (e.g. DeJong and Whiteman (1989)). Subsequent efforts included development of volatility tests that loosen stationarity assumptions (e.g., Campbell and Shiller (1987) and West (1988)).

Still, the controversy was unresolved. Perhaps if the nature of the econometric procedures or the assumption regarding the dividend process were not the culprits, the fault might lie with the economic environment. Indeed, the assumption of a constant discount factor was

Date: October, 2006.
JEL codes: G12, C53, C32.
† We thank Paul Muhly for a helpful conversation regarding the properties of Blaschke factors.

challenged quickly by LeRoy and LaCivita (1981) and Michener (1982), who showed that with a stochastic discount factor (the intertemporal marginal rate of substitution) appropriate for a representative risk-averse agent, some (but alas not enough) extra volatility would be generated in stock prices.

In 1985, the controversy morphed into the more general but closely related "equity premium" puzzle (Mehra and Prescott (1985)): the "excessively" volatile stock prices are associated with a return to stocks that is too much larger than the very low return to bonds to be explained in the context of the simplest asset pricing model unless the representative consumer is unrealistically risk averse. Hansen and Jagannathan (1991) developed a version of Shiller's variance bound for more general economies which has facilitated the study of more general specifications of preferences involving time-nonseparabilities (Constantinides (1990)) and state-nonseparabilities (Epstein and Zin (1989)) These modifications have not resolved the controversy either (Otrok, Ravikumar, and Whiteman (2002)).

We illustrate a different approach: it isn't sampling variability in volatility estimates, it isn't the dividend assumption, it isn't the discount factor, *it's the expectation.* We study the simplest, Shiller-style present value model, and imagine that the predictions of future dividends are not those of a forecaster minimizing squared error loss, but rather those of a forecaster facing and dealing with ambiguity regarding the economic environment. That is, the forecaster admits the possibility that the correct prediction model is unknown. To accommodate such an unpleasant but realistic situation, as much decision-theoretic literature suggests, the forecaster might behave in such a way as to mitigate the worst outcome that could conceivably occur.

To motivate this minimax loss function, consider the following experimental example of an agent's aversion to uncertainty and of actions taken to minimize downside risk that was originally proffered in Ellsberg (1961):

Experiment 1

> There are two urns, labeled Urn_I and Urn_{II}, each containing a total of 100 chips painted either black or white. The subject is allowed to examine the contents of Urn_I and discovers that it contains exactly 50 white chips and 50 black chips. The subject is not permitted to examine Urn_{II} other than to be told it contains 100 chips colored either black or white, but in an unknown ratio. The subject is given a list of four possible gambles:
>
> (A) The chip drawn from Urn_I is black.
> (B) The chip drawn from Urn_I is white.
> (C) The chip drawn from Urn_{II} is black.
> (D) The chip drawn from Urn_{II} is white.
>
> Winning a bet entitles the agent to a prize of value, for example \$100.

In studies replicating this experiment, the following preference orderings over gambles A through D have been observed: $A \sim B \succ C \sim D$.[1] As noted by Gilboa and Schmeidler (1989), who cite a version of this experiment in their seminal minimax decision theory work, it is easy to see that there is no probability measure supporting these preference orderings under expected utility maximization. Minimax decision theory is based on the following

[1] One example of a classroom experiment in which this result is exhibited is found in Study 1 of Fox and Tversky (1995), which used 141 undergraduates at Stanford University, responding to a questionnaire consisting of this and several other unrelated items that subjects completed for class credit.

explanation of this preference ordering: Aversion to the uncertainty regarding the contents of Urn_{II} causes the agent to treat the worst case scenario among a set of possibilities as the foundation of his prior probabilities on Urn_{II}. Thus, the agent's preferences represent a desire to minimize potential downside risk.

In the present value prediction problem, this means replacing ordinary expectations of future dividends with robust predictions. The standard reference for this in the economics literature is Hansen and Sargent (2005), who for the most part employ state space methods for calculating robust procedures. Such methods are convenient for numerical calculations, but make analytic derivations difficult. Because the environment is simple and we wish to illustrate the effect of robust decisions on economic outcomes, we utilize related "frequency domain" procedures that build on Whiteman (1985), Whiteman (1986), and Kasa (2001). As these procedures are arcane but not deep, we present details of the derivations to make the paper self-contained. Briefly, we exploit the robust decision-maker's aversion to serially correlated errors to derive the robust present value stock price analytically.

Our results indicate that robust predictions can be quite different from ordinary ("least squares") predictions, and the robust expected present value of dividends can be quite variable–so variable, in fact, that our initial univariate calculations (robust analogues of Shiller's) *reverse* the volatility relationship in dramatic fashion. It turns out, however, that this result implies that the robust predictor is hedging against a potential degree of model misspecification that, while possible, is extremely unlikely.

To obtain a reasonable limitation on the degree of robustness the predictor adopts, we employ a version of the "evil agent" game of Hansen and Sargent (2005). In this game, the predictor uses least squares methods, but plays a dynamic game against an evil "nature" who may deliver dividend process different from those the forecaster believes to characterize the data. Nature is constrained in how much noise she can add to the situation. With this formulation, the question becomes one of how much freedom nature would have to possess to cause the predictor agent to make present value predictions consistent with actual stock prices, and whether this would be implausible. Our estimates of a simple evil agent game suggest that the required freedom is not that great and that the other implications of the model might not be implausible. That is, the "moderately robust" present value model generates prices that *are* consistent with subsequent changes in dividends.

2. The Setup and the Variance Bound

Shiller (1981) begins with the simple present value model

$$p_t = E_t \sum_{k=0}^{\infty} \gamma^k d_{t+k}. \tag{1}$$

where $\gamma = 1/(1+r)$ with r the (assumed constant) real rate of interest, p_t and d_t are the real stock price and divident at time t, and E_t denotes conditional expectation given information available at t. Shiller writes the model in terms of the ex post rational price series p_t^*, which is defined as the present value of the *actual* subsequent real dividends:

$$p_t^* = \sum_{k=0}^{\infty} \gamma^k d_{t+k} \tag{2}$$

Clearly, as this requires us to have the *actual* dividend sequence out into the infinite future, it is impossible to observe p_t^* without error. Shiller notes that with a long enough dividend sequence, we can observe an approximate p_t^* by choosing a terminal date, making some assumptions about the dividend series after that terminal date and then constructing the p_t^* series via backwards recursion usinng[2]

$$p_t^* = \gamma(p_{t+1}^* + d_{t+1}). \tag{3}$$

In any case, from (1) and (2), we have

$$p_t = E_t(p_t^*) \tag{4}$$

which implies

$$p_t^* = p_t + \epsilon_t$$

where ϵ_t is orthogonal to information available at time t (including p_t). Then $var(p_t^*) = var(p_t) + var(\epsilon_t)$, and we have

$$\sigma(p) \leq \sigma(p^*), \tag{5}$$

That actual p is vastly more variable than measured p^* is demonstrated in Shiller's Figure 1 for the SP Composite Index from 1871-1979, as replicated below. The p^* in the figure was generated by the backwards recursion on (3) and an estimated $r = 0.048$, implying $\gamma = 0.943$. For the data in the figure, the estimated standard deviation of dividends $\sigma(d)$ is 1.12 while $\sigma(p)$ is 42.74 and $\sigma(p^*)$ is 7.24. Note that prices volatility is about 34 times that of dividends.

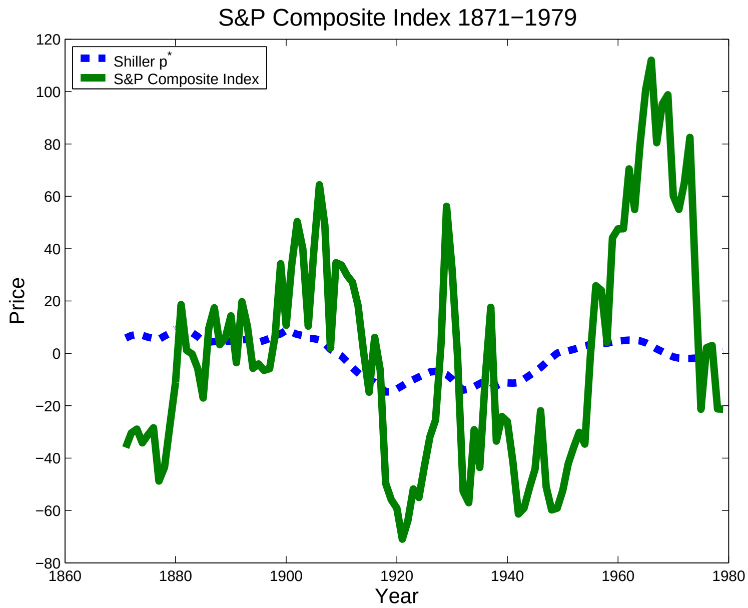

Figure 1. Shiller's Figure 1

[2]The assumptions that Shiller makes about dividends after the terminal date are that they are smooth and grow at the exponential growth rate taken out of the original data.

3. The Least Squares Prediction

To provide a benchmark for the robust present value calculation to be presented below, in this section we present a first-principles derivation of the stochastic process for stock prices implied by the present value model and the assumption that the dividend process is stationary after detrending. This benchmark case involves computation of the so-called "Hansen and Sargent (1980) formula".[3] The robust calculation of the ensuing section is an example of the "robust Hansen-Sargent formula" of Kasa (2001).

We will assume that the agent's information set includes current and past observations on a trend-stationary dividend process; in fact, we will assume that the market forecasters know that the data generating process for the detrended dividends (i.e., the "model") coincides with Wold representation, which we write

$$d_t = \sum_{j=0}^{\infty} q_j \varepsilon_{t-j} = q(L)\varepsilon_t \qquad E(\varepsilon_t) = 0\,, \quad E(\varepsilon_t^2) = 1 \tag{6}$$

using the lag operator L and defining $q(L)$ implicitly. Using (6) and (1), we have

$$p_t = E_t \sum_{j=0}^{\infty} \gamma^j d_{t+j} = E_t\left(\frac{q(L)}{1-\gamma L^{-1}}\right)\varepsilon_t = E_t(p_t^*). \tag{7}$$

The coincidence of the Wold representation and the data generation process means that ε_t is both statistically and economically "fundamental" for d_t. Thus (6) not only *represents* dividends via the Wold representation, but also *generates* dividends: (6) is the "correct" model. We defer to the next section the case in which the representative investor is not so sure about this.

The least-squares minimization problem the representative agent faces is to find a stochastic process p_t to minimize the expected squared prediction error $E(p_t - p_t^*)^2$. In terms of the information known at date t, the agent's task is to find a linear combination of current and past dividends, or, equivalently, of current and past dividend innovations ε_t that is "close" to p_t^*. Writing $p_t = f(L)\varepsilon_t$, the problem becomes one of finding the coefficients f_j in $f(L) = f_0 + f_1 L + f_2 L^2 + ...$ to minimize $E(f(L)\varepsilon_t - p_t^*)^2$. Using the technique in Whiteman (1985), this problem has an equivalent, frequency-domain representation

$$\min_{f(z)\in H^2} \frac{1}{2\pi i}\oint \left|\frac{q(z)}{1-\gamma z^{-1}} - f(z)\right|^2 \frac{dz}{z} \tag{8}$$

where H^2 denotes the Hardy space of square-integrable analytic functions on the unit disk, and $\oint$ denotes (counterclockwise) integration about the unit circle. The requirement that $f(z) \in H^2$ ensures that the forecast is causal, and contains no future values of the ε's. Expression (8) may seem exotic, but is quite simple: noting that on the unit circle $z = e^{i\omega}$, making the substitution and integrating with respect to ω from 0 to 2π, the expression calls for minimizing the average value (or area under) the spectral density of the prediction error $p_t - p_t^*$. Of course this follows from the fact that the variance of a process is equal to the integral of its spectral density. In the next section, we shall study a forecaster who seeks a

[3] A survey of these methods can be found in Whiteman (1983).

function $f(z)$ to minimize not the average value on the unit circle, but the *maximum* value on the unit circle.

The first-order conditions for choosing f_j are, for $j = 0, 1, 2, ...,$

$$\begin{aligned}
&\frac{\partial}{\partial f_j}\left[\frac{1}{2\pi i}\oint\left|\frac{q(z)}{1-\gamma z^{-1}} - f(z)\right|^2 \frac{dz}{z}\right] \\
&\quad = -\frac{1}{2\pi i}\oint z^j\left[\frac{q(z^{-1})}{1-\gamma z} - f(z^{-1})\right] + z^{-j}\left[\frac{q(z^{-1})}{1-\gamma z^{-1}} - f(z)\right]\frac{dz}{z} \\
&\quad = -\frac{1}{2\pi i}\oint z^j\left[\frac{q(z^{-1})}{1-\gamma z} - f(z^{-1})\right]\frac{dz}{z} - \frac{1}{2\pi i}\oint z^{-j}\left[\frac{q(z^{-1})}{1-\gamma z^{-1}} - f(z)\right]\frac{dz}{z} = 0. \qquad (9)
\end{aligned}$$

Now change variables in the second contour integral to $w = z^{-1}$, implying $dw = -1(z^{-2})$ and $dz/z = -dw/w$, with integration with respect to w being clockwise. Upon changing variables in the first integral and multiplying by -1 to reverse integration back to clockwise, the the two integrals in (9) become identical and the equality collapses to

$$-\frac{2}{2\pi i}\oint z^{-j}\left[\frac{q(z)}{1-\gamma z^{-1}} - f(z)\right]\frac{dz}{z} = 0. \qquad (10)$$

Now define

$$H(z) = \frac{q(z)}{1-\gamma z^{-1}} - f(z) \qquad (11)$$

so that (10) becomes

$$-\frac{2}{2\pi i}\oint z^{-j}H(z)\frac{dz}{z} = 0. \qquad (12)$$

This equality requires that for $j = 0, 1, 2, \ldots,$ twice the coefficient H_j in the Laurent expansion of $H(z)$ valid for $|z| = 1$ equal zero. Multiplying by z^j and summing over *all* $j = 0, \pm1, \pm2, ...,$ we find that

$$H(z) = \sum_{-\infty}^{-1}$$

where $\sum_{-\infty}^{-1}$ denotes an unknown function involving only negative powers of z. Then by recalling the definition of $H(z)$, we have

$$\frac{q(z^{-1})}{1-\gamma z^{-1}} - f(z) = \sum_{j=-\infty}^{-1}$$

which is known as a "Wiener-Hopf" equation. Application of the "plussing" operator to both sides of the equation yields:[4]

[4]The plussing operator is a linear annihilator operator that means "ignore negative powers of z."

$$\left[\frac{q(z)}{1-\gamma z^{-1}}\right]_+ - [f(z)]_+ = 0$$

implying

$$f(z) = \left[\frac{q(z)}{1-\gamma z^{-1}}\right]_+ = \left[\frac{zq(z)}{z-\gamma}\right]_+ \tag{13}$$

which points to the fact that $f(z)$ is, by construction, one-sided in non-negative powers of z. We now use a method highlighted in Appendix A of Hansen and Sargent (1980) to determine the form of the function in (13). First, note that the function being "plussed" in (13) is well-behaved $|z| < 1$ except for a single simple pole at γ. By examining the Laurent expansion of $q(z)$ around γ, we are able to determine that the principle part (that is, the part of the Laurent expansion containing negative powers of z) is $P(z) = \gamma q(\gamma)/(z-\gamma)$. Second, we note that "plussing" involves simply subtracting off those parts. That is:

$$[A(z)]_+ = A(z) - P(z)$$

where $P(z)$ is the principle part of the Laurent series expansion of $A(z)$. This implies

$$f(z) = \left[\frac{q(z)}{1-\gamma z^{-1}}\right]_+ = \left[\frac{zq(z)}{z-\gamma}\right]_+ = \frac{zq(z)-\gamma q(\gamma)}{z-\gamma}.$$

To illustrate how the formula works, suppose detrended dividends are described by a first-order autoregression; i.e., that $q(L) = (1-\rho L)^{-1}$. Then

$$p_t = f(L)\varepsilon_t = \frac{Lq(L)-\gamma q(\gamma)}{L-\gamma}\varepsilon_t = \left(\frac{1}{1-\rho\gamma}\right) d_t.$$

In this simple first order case,

$$\sigma(p) = \left(\frac{1}{1-\rho\gamma}\right)\sigma(d). \tag{14}$$

With $\gamma = 0.943$, as estimated from the S&P data, the largest $\sigma(p)$ can be for stationary dividends ($|\rho| < 1$) is about 22 times dividends. Estimating ρ from the same data, the ratio is about 11, far short of the factor of 34 needed to match the observed volatility.

It is instructive to note that while the pricing formula (13) makes p_t the best least squares predictor of p_t^*, the prediction errors $p_t - p_t^*$ will not be serially uncorrelated. Indeed

$$\begin{aligned}
p_t - p_t^* &= \gamma\left\{\frac{Lq(L)-\gamma q(\gamma)}{L-\gamma} - \frac{q(L)}{1-\gamma L^{-1}}\right\}\varepsilon_t \\
&= \frac{-\gamma^2 q(\gamma)}{L-\gamma}\varepsilon_t \;=\; -\gamma^2 q(\gamma)\frac{L^{-1}}{1-\gamma L^{-1}}\varepsilon_t \\
&= -\gamma^2 q(\gamma)\left\{\varepsilon_{t+1} + \gamma\varepsilon_{t+2} + \gamma^2\varepsilon_{t+3} + \dots\right\}.
\end{aligned}$$

Thus the prediction errors will be described by a highly persistent (γ is close to unity) first-order autoregression. But because this autoregression involves *future* ε_t's, the serial correlation structure of the errors cannot be exploited to improve the quality of the prediction of p_t^*. The reason is that the predictor "knows" the *model* for price setting (the present value formula) and the dividend process; the best predictor $p_t = E_t p_t^*$ of p_t^* "tolerates" the serial correlation because the (correct) model implies that it involves *future* ε_t's and therefore cannot be predicted. If one only had data on the errors (and did not know the model that generated them), they would appear (rightly) to be characterized by a first-order autoregression; fitting an AR(1) (i.e., the best *linear* model) and using it to "adjust" p_t by accounting for the serial correlation in the errors $p_t - p_t^*$ would decrease the quality of the estimate of p_t^*. The reason is the usual one that the Wold representation for $p_t - p_t^*$ is not the economic model of $p_t - p_t^*$, and (correct) models always beat Wold representations. This also serves as a reminder of circumstances under which one should be willing to tolerate serially correlated errors: when one knows the model that generated them, and the model implies that they are as small as they can be made.

4. The Robust Prediction Case

What happens in case the individual making the prediction of future dividends does not know for certain that dividends are generated as in (6)? This notion of ambiguity was introduced in the linear, time-invariant context we are studying in the engineering literature by Zames (1981), and has been studied more generally in the economics literature by Gilboa and Schmeidler (1989), Hansen and Sargent (2005) and others. In our setup, the ambiguity would be manifested in possible departures from the moving average representation (6). Following the development Kasa (2001) used in a related context, suppose the dividend process is given by

$$d_t = [q(L) + \Delta(L)]\,\varepsilon_t \tag{15}$$

where $\Delta(L)\varepsilon_t$ is a "perturbation" of the original dividend process. Then if the forecaster uses (13), the actual squared error loss $\mathcal{L}^A = E[p_t - p_t^*]^2$ will be given by

$$\begin{aligned}
\mathcal{L}^A &= \mathcal{L}^q + \|\Delta(z)\|_2^2 + \frac{2}{2\pi i}\oint \Delta(z)\left[\frac{q(z)}{1-\gamma z^{-1}}\right]_- \frac{dz}{z} \\
&= \mathcal{L}^q + \|\Delta(z)\|_2^2 + \frac{2}{2\pi i}\oint \Delta(z)\left(\frac{\gamma q(\gamma)}{z-\gamma}\right)\frac{dz}{z} \\
&= \mathcal{L}^q + \|\Delta(z)\|_2^2 + 2q(\gamma)\,[\Delta(\gamma) - \Delta(0)]
\end{aligned}$$

where $\mathcal{L}^q$ is the loss when dividends are indeed generated by $d_t = q(L)\varepsilon_t$. The second line follows from the linear annihilator operator $[\cdot]_-$ which means "ignore positive powers of z." This leaves only the principle part of the element in the brackets, which was shown earlier to be $\gamma q(\gamma)/(z-\gamma)$. The third equality is a result of the application of Cauchy Residue Theorem. The expression for $\mathcal{L}^A$ indicates that the actual loss could be much larger than $\mathcal{L}^q$ even for a small perturbation provided $q(\gamma)$ is large. This result, combined with the knowledge that the true dividend process is hard to come by, suggests that the forecast should be constructed with greater robustness to model misspecification.

The problem with a misspecified model is that the "wrong" sequence of "errors" ε_t could "excite" $\Delta(L)$ in such a way that very large prediction errors occur. To guard against this, the predictor might wish to make forecasts that minimize the maximum possible squared error loss rather than the average or expected squared error loss.

The robust predictor solves

$$\min_{f(z)\in H^\infty} \max_{|z|=1} \left| \frac{q(z)}{1-\gamma z^{-1}} - f(z) \right|^2 \Leftrightarrow \min_{f(z)\in H^\infty} \max_{|z|=1} \left| \frac{zq(z)}{z-\gamma} - f(z) \right|^2 . \tag{16}$$

Unlike in the least squares case (8), where $f(z)$ was restricted to the class H^2 functions finitely square integrable on the unit circle, the restriction now is to the class of functions with finite maximum modulus on the unit circle, and the H^2 norm has been replaced by H^∞ norm.

To begin the solution process, note that there is considerable freedom in designing the minimizing function $f(z)$: it must be well-behaved (i.e., must have a convergent power series in nonnegative powers of z on the unit disk), but is otherwise unrestricted. Further note that $zq(z)/(z-\gamma)$ can be thought of as the associated Laurent expansion, which is of the form

$$\frac{zq(z)}{z-\gamma} = \frac{b_{-1}}{z-\gamma} + b_0 + b_1(z-\gamma) + b_2(z-\gamma)^2 + ...$$

Intuitively, while in the least squares case $f(z)$ is set to "cancel" all the terms of this series except the first, here the object is to set $f(z)$ to minimize a different function of the prediction errors. Now define the "Blaschke factor" $B_\gamma(z) = (z-\gamma)/(1-\gamma z)$ and note that

$$\left| \frac{z-\gamma}{1-\gamma z} \right|^2 = \frac{(z-\gamma)(z^{-1}-\gamma)}{(1-\gamma z)(1-\gamma z^{-1})} = \frac{(z-\gamma)z^{-1}(1-\gamma z)}{(1-\gamma z)z^{-1}(z-\gamma)} = 1.$$

Multiplying the objective by the Blaschke factor thus does not alter its value on the unit circle, but the factor does cancel the pole at γ, yielding

$$\min_{\{f(z)\}} \sup_{|z|=1} \left| \frac{zq(z)}{1-\gamma z} - \frac{z-\gamma}{1-\gamma z} f(z) \right|^2 .$$

Defining

$$T(z) = \frac{zq(z)}{1-\gamma z} \tag{17}$$

we have

$$\min_{f\in H^\infty} \sup_{|z|=1} |T(z) - B_\gamma(z) f(z)| \Leftrightarrow \min_{f\in H^\infty} \|T(z) - B_\gamma(z) f(z)\|_\infty . \tag{18}$$

Define the function inside the $\|$'s as

$$\phi(z) = T(z) - B_\gamma(z) f(z) \tag{19}$$

and note that $\phi(\gamma) = T(\gamma)$. Thus the problem of finding $f(z)$ reduces to the problem of finding the smallest $\phi(z)$ satisfying $\phi(\gamma) = T(\gamma)$:

$$\min_{\phi\in H^\infty} \|\phi(z)\|_\infty \text{s.t. } \phi(\gamma) = T(\gamma)$$

Theorem 4.1. *(Kasa, 2001) The solution to* (20) *is the constant function* $\phi(z) = T(\gamma)$.

Proof. To see this, first note that the norm of a constant function is the modulus of the constant itself. This is written as

$$\|\phi(z)\|_\infty = \|T(\gamma)\|_\infty = |T(\gamma)|^2. \tag{20}$$

Next, suppose that there exists another function $\Psi(z) \in H^\infty$, with $\Psi(\gamma) = T(\gamma)$ and also

$$\|\Psi(z)\|_\infty < \|\phi(z)\|_\infty. \tag{21}$$

Recall the definition of the L^∞ norm, and using equations (20) and (21):

$$\|\Psi(z)\|_\infty = \sup_{|z|=1} |\Psi(z)|^2 < |T(\gamma)|^2. \tag{22}$$

The maximum modulus theorem states that a function f which is analytic on the disk U achieves its maximum on the boundary of the disk. That is

$$\sup_{z\in U} |f(z)|^2 \leq \sup_{z\in\partial U} |f(z)|^2. \tag{23}$$

Therefore, we can see that

$$\sup_{|z|<1} |\Psi(z)|^2 \leq \sup_{|z|=1} |\Psi(z)|^2 < |T(\gamma)|^2. \tag{24}$$

However, one of the values on the interior of the unit disk is $z = \gamma$, which can be plugged in to the far LHS of equation (24) to get the result

$$|\Psi(\gamma)|^2 \leq \sup_{|z|=1} |\Psi(z)|^2 < |T(\gamma)|^2 \implies |\Psi(\gamma)|^2 < |T(\gamma)|^2. \tag{25}$$

This contradicts the requirement that $\Psi(\gamma) = T(\gamma)$. Therefore, we have verified that there does not exist another function $\Psi(z) \in H^\infty$ such that $\Psi(\gamma) = T(\gamma)$ and $\|\Psi(z)\|_\infty < \|\phi(z)\|_\infty$. □

Now that we have a form for $\phi(z)$, we can use it to find a formula for $f(z)$. Recalling the form of $f(z)$ and completing some tedious algebra, we obtain

$$f(z) = \frac{T(z) - \phi(z)}{B_\gamma(z)} = \frac{zq(z) - \gamma q(\gamma)}{z - \gamma} + \frac{\gamma^2}{1-\gamma^2} q(\gamma)$$

which is the least-squares solution plus a constant. This means that after the initial period, the impulse response function for the robust predictor is identical to that of the least squares predictor. In the initial period, the least squares impulse response is $q(\gamma)$, while the robust impulse response is larger: $q(\gamma)/(1-\gamma^2)$. Recalling that γ is the discount factor, and therefore close to unity, the robust impulse response can be considerably larger than that of the least squares response. Relatedly, the volatility of prices in the robust case will be larger as well. For example, in the first-order autoregressive case studied above,

$$p_t = f(L)\varepsilon_t = \frac{1}{1-\rho\gamma}d_t + \frac{\gamma^2}{(1-\gamma^2)(1-\rho\gamma)}\varepsilon_t \tag{26}$$

from which the variance can be calculated as

$$\sigma^2(p_t) = \left(\frac{1}{1-\rho\gamma}\right)^2 \sigma^2(d_t) + \frac{2\gamma^2 - \gamma^4}{(1-\rho\gamma)^2(1-\gamma^2)^2}. \tag{27}$$

Using the data from Figure 1, we have the values for $\sigma(d)$, r, ρ and γ that we have been using throughout this example, equation (27) gives us the result that $\sigma(p) = 89.52$.

The standard deviation of the actual price sequence in the SP dataset is 42.74. Thus when the agent is robust to the most misspecification possible, the resulting price volatility will have a standard deviation over twice as high as would be needed to exhibit the excess volatility seen in the data: *the robust present value model predicts prices that are substantially more volatile than those seen in the data.* The "robust puzzle" is therefore why prices are so *smooth*. The reversal of the volatility relationship is apparent in the figure below.

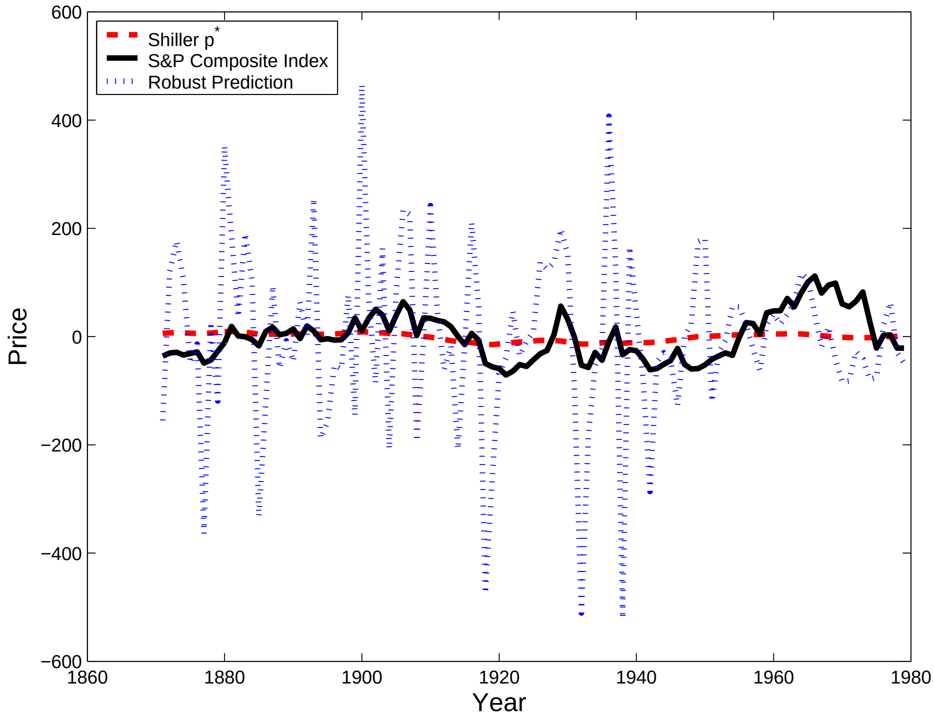

Figure 2. Robust Price Result

5. The Evil Agent Game

In this section we investigate an intermediate case in which the investor-predictor is behaves robustly relative to a restricted set of possible models for dividends. In particular, following Hansen and Sargent (2005), we consider a a game played between the predicting agent and a malicious nature in which the degree of nature's malevolence is restricted. This restriction comes in the form of a cost associated with delivering excessively "noisy" dividends to the agent. The solution procedure, taken from Whiteman (1986), requires only slight modification from the development in sections 3 and 4.

5.1. **The Investing Agent's Problem.** Imagine that the dividend process investing agents perceive is written as

$$d_t = [q(L) + m(L)]\,\varepsilon_t \equiv C(L)\varepsilon_t.$$

Thus, the investor agent's problem becomes to choose an analytic function $f(z)$ in order to

$$\min_{f(z)} \frac{1}{2\pi i} \oint \left| \frac{C(z)}{1-\gamma z^{-1}} - f(z) \right|^2 \frac{dz}{z}. \tag{28}$$

As above, the optimization in equation (28) leads to the following Weiner-Hopf equation:

$$\frac{C(z)}{1-\gamma z^{-1}} - f(z) = \sum_{-\infty}^{-1}. \tag{29}$$

5.2. **The Evil Agent's Problem.** The Evil Agent's problem in this game is to make it as difficult as possible for the investing agent to make this prediction. However, there are restrictions on how much power the Evil Agent (EA) has to exercise. The EA has control over $m(z)$, which, as seen above, is a component of the dividend process that Investing Agents (IA) take as given. The following optimization shows the EA's problem.

$$\max_{C(z)} \frac{1}{2\pi i} \oint \left| \frac{C(z)}{1-\gamma z^{-1}} - f(z) \right|^2 - \theta \left| \frac{C(z)-q(z)}{1-\gamma z^{-1}} \right|^2 \frac{dz}{z} \tag{30}$$

5.3. **Solving the Evil Agent Game.** First, we find the Weiner-Hopf equation that results from the EA's optimization problem.

$$\begin{aligned} 0 = \frac{1}{2\pi i} \oint \Bigg\{ & \frac{z^j}{1-\gamma z^{-1}} \left[\frac{C(z^{-1})}{1-\gamma z} - f(z^{-1}) \right] - \frac{\theta z^j}{1-\gamma z^{-1}} \left[\frac{C(z^{-1}) - q(z^{-1})}{1-\gamma z} \right] \\ & + \frac{z^{-j}}{1-\gamma z} \left[\frac{C(z)}{1-\gamma z^{-1}} - f(z) \right] - \frac{\theta z^{-j}}{1-\gamma z} \left[\frac{C(z)-q(z)}{1-\gamma z^{-1}} \right] \Bigg\} \frac{dz}{z} \end{aligned}$$

$$\longrightarrow \quad \frac{(1-\theta)C(z)}{1-\gamma z} - \left[\frac{(1-\gamma z^{-1})}{1-\gamma z} f(z) \right] + \frac{\theta q(z)}{1-\gamma z} = \sum_{-\infty}^{-1}. \tag{31}$$

We recall from (29) that the Investor Agent's Weiner-Hopf equation can be written

$$C(z) = (1-\gamma z^{-1}) f(z) + \sum_{-\infty}^{-1}.$$

After applying the plussing operator, this leaves

$$C(z) = \left[(1-\gamma z^{-1}) f(z) \right]_+. \tag{32}$$

Using (32) in (31) and applying the plussing operator again, we have

$$\frac{1-\theta}{1-\gamma z}\left[(1-\gamma z^{-1})f(z)\right]_{+}-\left[\frac{(1-\gamma z^{-1})}{1-\gamma z}f(z)\right]_{+}+\frac{\theta q(z)}{1-\gamma z}=0. \tag{33}$$

In order to proceed, we need to solve for the "plussed" elements of equation (33). The first "plussed" term is equivalent to

$$\begin{aligned}\left[(1-\gamma z^{-1})f(z)\right]_{+}&=f(z)-\gamma\left[\frac{f(z)}{z}\right]_{+}\\&=f(z)-\gamma\left[\frac{f(z)-f_0}{z}\right]\\&=(1-\gamma z^{-1})f(z)+\frac{\gamma f_0}{z}.\end{aligned} \tag{34}$$

Similarly,

$$\left[\frac{(1-\gamma z^{-1})}{1-\gamma z}f(z)\right]_{+}=\frac{(1-\gamma z^{-1})}{1-\gamma z}f(z)+\frac{\gamma f_0}{z}. \tag{35}$$

Using (34) and (35) in (33), we obtain (via somewhat length algebraic manipulation):

$$f(z)=\frac{zq(z)-\gamma q(\gamma)}{z-\gamma}+\frac{\gamma^2}{\theta-\gamma^2}q(\gamma). \tag{36}$$

It is seen from (36) that $f(z)$ takes a recognizable form. The first term in $f(z)$ is the solution to the standard prediction problem in the event that the forecaster is attempting to minimize MSE, rather than playing the game against an Evil Agent. Therefore, we can see that a forecaster using robust methods will end up using a function $f(z)$ that looks like an MSE forecast plus an extra term having to do with what he is trying to be robust against. Note that the value of θ will control just how large a part the second term in (36) will play in the forecast. Recall from (30) that θ is the Lagrange multiplier on the constraint the EA faces. Changing the value for θ is interpreted as loosening or tightening the constraint faced by the EA. As we increase θ, we make it more costly (in terms of the trade-off within his optimization) for the EA to add noise to the system. As we decrease θ, we make it less costly.

It turns out that there are two values for θ which make $f(z)$ immediately recognizable. The following relationship is clear from (36).

$$\begin{cases}\theta\to\infty & f(z)\to\text{MSE Result}\\ \theta\downarrow 1 & f(z)\to H^{\infty}\text{-norm Result}\end{cases}$$

The H^{∞}-norm result is one that has been shown earlier, and is a case where forecasters look at trying to generate a forecast when faced with the worst possible model misspecification. A natural question at this point would be to ask why the worst-case scenario does not occur at $\theta=0$, as that is the value for θ that people naturally associate with a non-binding constraint and therefore total freedom for the EA. The answer to this mystery lies in the saddle-point nature of this optimization problem. The second order conditions for finding a maximum are violated for values of $\theta\in(0,1)$. Therefore, the lower limit on θ is equal to 1.

Now that a form for $f(z)$ has been uncovered, the next question involves the final form of $C(z)$, that is: what do dividends look like to the investor agent? Further, we would like to know how much noise is being added by the EA in equilibrium. We pursue the form of $C(z)$ first by recalling equations (32) and (34), which gave us

$$C(z) = \left[(1-\gamma z^{-1})f(z)\right]_+ . \tag{37}$$

Substituting from (36) into (37) results in another "plussing" problem:

$$C(z) = \left[(1-\gamma z^{-1})\left\{\frac{q(z)-\gamma z^{-1}q(\gamma)}{1-\gamma z^{-1}}\right\} + (1-\gamma z^{-1})\left\{\frac{\gamma^2}{\theta-\gamma^2}q(\gamma)\right\}\right]_+$$

which when solved using the methods shown above yields

$$C(z) = q(z) + \frac{\gamma^2}{\theta-\gamma^2}q(\gamma). \tag{38}$$

5.4. **Additional Noise from the Evil Agent.** Now that we have the form of $C(z)$, we can use the Cauchy Integral formula to solve for the present value of the noise added in by the Evil Agent. We will denote the present value of the noise by η.

$$\begin{aligned}
\eta &= \frac{1}{2\pi i}\oint \left|\frac{C(z)-q(z)}{1-\gamma z^{-1}}\right|^2 \frac{dz}{z} \\
&= \frac{\gamma^4 q^2(\gamma)}{(\theta-\gamma^2)^2}\frac{1}{2\pi i}\oint \frac{dz}{(1-\gamma z^{-1})(1-\gamma z)z} \\
&= \frac{\gamma^4 q^2(\gamma)}{(\theta-\gamma^2)^2(1-\gamma^2)}
\end{aligned}$$

6. Empirical Results

Do parameterizations of the Evil Agent game do a good job of fitting the data? We begin with the simplest possible univariate setup and then proceed to more realistic and complex environments. In the univariate case, we use the dividend process that was used in Shiller (1981). Thus

$$q(L) = \frac{1}{1-\rho L} \tag{39}$$

where $\rho = 0.95$. The analytical results allow us to use any dividend process; this particular choice reflects our desire to keep the initial model comparison between the results of Shiller (1981) and this work as easy as possible.

6.1. **Univariate θ Estimation.** As a first pass, we took the function for price prediction and asked, "given the data, what value of θ would create price volatility in our model equal to that in the data?" This was done by assuming the dividends were generated by the AR(1) process in Shiller (1981); the required $\theta = 1.62$. To put this into context, we examine what this implies the investor must be thinking about dividends.

6.2. **Detectability of θ.** In the EA game, the final dividend process being targeted by the IA is expressed by equation (38). This means that the variance calculation of the resulting dividend process can be written:

$$
\begin{aligned}
var\left[C(L)\varepsilon_t\right] &= \sigma_\varepsilon^2 \left\{ \left(1 + \frac{\gamma^2}{\theta - \gamma^2} q(\gamma)\right)^2 + c_1^2 + c_2^2 + c_3^2 + \dots \right\} \\
&= \sigma_\varepsilon^2 \left\{ \left(1 + \frac{\gamma^2}{\theta - \gamma^2} q(\gamma)\right)^2 + \rho^2 + \rho^4 + \rho^6 + \dots \right\} \\
&= \sigma_\varepsilon^2 \left\{ \left(1 + \frac{\gamma^2}{\theta - \gamma^2} q(\gamma)\right)^2 + \frac{\rho^2}{1-\rho^2} \right\} \qquad (40)
\end{aligned}
$$

Now, compare the result in equation (40) with the variance of dividends if they were produced by the AR(1) process given above:

$$var[q(L)\varepsilon_t] = \frac{\sigma_\varepsilon^2}{1-\rho^2}. \qquad (41)$$

To compare the two, we calculate

$$
\begin{aligned}
\frac{var[C(L)\varepsilon_t]}{var[q(L)\varepsilon_t]} &= \frac{\sigma_\varepsilon^2 \left\{ \left(1 + \frac{\gamma^2}{\theta-\gamma^2} q(\gamma)\right)^2 + \frac{\rho^2}{1-\rho^2} \right\}}{\frac{\sigma_\varepsilon}{1-\rho^2}} \\
&= (1-\rho^2) \left\{ \left(1 + \frac{\gamma^2}{\theta - \gamma^2} q(\gamma)\right)^2 + \frac{\rho^2}{1-\rho^2} \right\} \\
&= (1-\rho^2)\left(1 + \frac{\gamma^2}{\theta - \gamma^2} q(\gamma)\right)^2 + \rho^2 \\
&= 606.55,
\end{aligned}
$$

which indicates that the investing agent is guarding against a dividend process that is *quite* different from what has been observed. The problem with this comparison is that it does not permit compromise: is there a θ that gets price variability "close" without making dividends too variable? To study this, we will need to estimate a system which simultaneously fits both the price and dividend processes.

6.3. **Bivariate Estimation.** We begin by specifying a general bivariate moving average representation for d_t and p_t.:

$$\begin{pmatrix} d_t \\ p_t \end{pmatrix} = \begin{pmatrix} A(L) & B(L) \\ C(L) & D(L) \end{pmatrix} \begin{pmatrix} \varepsilon_{d_t} \\ \varepsilon_{p_t} \end{pmatrix}.$$

In order to accommodate the cross-equation restrictions implied by the present value relationship, we identify the system by restricting the innovations to be uncorrelated and unit variance Guassian processes. It is possible to describe both the least-squares prediction problem and the Evil Agent game within this structure. In fact, the calculations of the previous sections are now applied column by column, so that the elements of $C(L)$ are functions of the elements of $A(L)$, and $D(L)$ is a function of $B(L)$. We choose to specify dividends as the sum of a persistent component (like Shiller's AR(1) process) and a transitory component. Then the restricted moving average representation in the least squares case is

$$\begin{pmatrix} A_{LS}(L) & B_{LS}(L) \\ C_{LS}(L) & D_{LS}(L) \end{pmatrix} = \begin{pmatrix} A_0 & \frac{B_0}{1-bL} \\ \frac{LA(L)-\gamma A(L)}{L-\gamma} & \frac{LB(L)-\gamma B(\gamma)}{L-\gamma} \end{pmatrix} = \begin{pmatrix} A_0 & \frac{B_0}{1-bL} \\ A_0 & \frac{B_0}{1-b\gamma}\left(\frac{1}{1-bL}\right) \end{pmatrix}$$

while in the Evil Agent game the MA is

$$\begin{pmatrix} A_{EA}(L) & B_{EA}(L) \\ C_{EA}(L) & D_{EA}(L) \end{pmatrix} = \begin{pmatrix} A_0 + \frac{\gamma^2}{\theta-\gamma^2}A_0 & \frac{B_0}{1-bL} + \frac{\gamma^2}{\theta-\gamma^2}\left(\frac{B_0}{1-b\gamma}\right) \\ \frac{LA(L)-\gamma A(L)}{L-\gamma} + \frac{\gamma^2}{\theta-\gamma^2}A(\gamma) & \frac{LB(L)-\gamma B(\gamma)}{L-\gamma} + \frac{\gamma^2}{\theta-\gamma^2}B(\gamma) \end{pmatrix}$$
$$= \begin{pmatrix} \frac{A_0\theta}{\theta-\gamma^2} & \frac{B_0}{1-bL} + \frac{\gamma^2}{\theta-\gamma^2}\left(\frac{B_0}{1-b\gamma}\right) \\ \frac{A_0\theta^2}{(\theta-\gamma^2)^2} & \frac{B_0\theta-\gamma^2 B_0 bL}{(1-bL)(\theta-\gamma^2)(1-b\gamma)} \end{pmatrix}.$$

6.4. **State Space Formulations.** By shifting into the state space, we can make use of the powerful set of tools available via Kalman filtering. This requires that each system be written in terms of a state and observer system. In order to estimate the system, we construct a state-space formulation of each of these vector MA systems which will allow us to use the Kalman filter to perform maximum likelihood estimation. The estimation will find parameter values for A_0, B_0 and b in the least-squares system, and A_0, B_0, b and θ in the Evil Agent game system.

6.4.1. *Least-Squares State Space.* The natural state of the system is the persistent component of dividends:

$$s_t = bs_{t-1} + \varepsilon_{p_t} \tag{42}$$

$$\begin{pmatrix} d_t \\ p_t \end{pmatrix} = \begin{pmatrix} B_0 \\ \frac{B_0}{1-b\gamma} \end{pmatrix} s_t + \begin{pmatrix} A_0 & 0 \\ A_0 & 0 \end{pmatrix} \begin{pmatrix} \varepsilon_{d_t} \\ \varepsilon_{p_t} \end{pmatrix} \tag{43}$$

where equation (42) is the state equation and (43) is the observer equation. With this forumlation, the Kalman filter can be used to evaluate the likelihood using a standard procedure (see, e.g., Hamilton (1994)). We find the following estimates of the parameters:

$$A_0 = 1.9894\,, \quad B_0 = 1.2478\,, \quad b = 0.9998.$$

The log-likelihood of this model is -718.48. This will provide a benchmark by which to evaluate the Evil Agent game model.

6.4.2. *Evil Agent State Space.* The state space for the Evil Agent game system can be written in the following way:

$$s_t = bs_{t-1} + \varepsilon_{p_t}$$

$$\begin{pmatrix} d_t \\ p_t \end{pmatrix} = \begin{pmatrix} B_0 \\ \frac{B_0}{1-b\gamma} \end{pmatrix} s_t + \begin{pmatrix} \frac{A_0\theta}{\theta-\gamma^2} & \frac{\gamma^2 B_0}{(\theta-\gamma^2)(1-b\gamma)} \\ \frac{A_0\theta^2}{(\theta-\gamma^2)^2} & \frac{\gamma^2 B_0}{(\theta-\gamma^2)(1-b\gamma)} \end{pmatrix} \begin{pmatrix} \varepsilon_{d_t} \\ \varepsilon_{p_t} \end{pmatrix}.$$

This system is estimated in exactly the same way as the least-squares system, thus demonstrating the flexibility of state space methods for problems such as these. The estimation of the system results in the following maximum likelihood estimates:

$$A_0 = 1.1693 * 10^{-5}\,, \quad B_0 = 1.1922\,, \quad b = 0.9998\,, \quad \theta = 10.3037.$$

The log-likelihood of this model is -712.03. The improvement is significant over the least squares model. Furthermore, when compared to the univariate estimate of θ, we that these parameter values do not make dividends too variable. Because the estimated A_0 is so small, both dividends and prices are dominated by the persistent component, and thus the relevant ratio in this case is

$$\frac{var[B_{EA}(L)\varepsilon_t]}{var[B_{LS}(L)\varepsilon_t]} = (1-\rho^2)\left(1 + \frac{\gamma^2}{\theta - \gamma^2}q(\gamma)\right)^2 + \rho^2 = 1.2375.$$

Unlike the univariate case, this value for θ is very promising in terms of generating prices and dividends which are simultaneously "close" to those in the data. The dividends in the model would be roughly only 12% more volatile than those in the data.

The problem, however, lies not in the volatility ratio, but in the overall fit of the model. While the evil agent model beats the least squares model, neither fare very well against a slightly less restrictive model. In particular, the system

$$s_t = b s_{t-1} + \varepsilon_{p_t} \tag{44}$$

$$\begin{pmatrix} d_t \\ p_t \end{pmatrix} = \begin{pmatrix} B_1 \\ B_2 \end{pmatrix} s_t + \begin{pmatrix} B_3 & 0 \\ B_4 & 0 \end{pmatrix} \begin{pmatrix} \varepsilon_{d_t} \\ \varepsilon_{p_t} \end{pmatrix} \tag{45}$$

achieves a value of the log likelihood about 100 higher than the evil agent system, indicating that there is a long way to go in fitting d_t and p_t jointly.

6.5. **More Sophisticated Data-Generating Processes.** The answer to this challenge lies in the use of more sophisticated processes for dividends and prices. The results of the ARMA(3,1) process below demonstrates the power of the Investor-Evil Agent Game. Consider the following bivariate ARMA(3,1) process.

$$\begin{pmatrix} d_t \\ p_t \end{pmatrix} = \begin{pmatrix} A(L) & B(L) \\ C(L) & D(L) \end{pmatrix} \begin{pmatrix} \varepsilon_{d_t} \\ \varepsilon_{p_t} \end{pmatrix}.$$

where

$$A(L) = \frac{\rho_0}{(1-\rho_3 L)(1-\rho_4 L)}\,, \quad B(L) = \frac{\mu_0(1-\mu_1 L)}{(1-L)(1-\mu_3 L)(1-\mu_4 L)}. \tag{46}$$

As seen earlier, the fully specified game between an investor and the Evil Nature results in the following processes for $C(L)$ and $D(L)$.

$$C(L) = \frac{LA(L) - \gamma A(\gamma)}{L - \gamma} + \frac{\gamma^2}{\theta - \gamma^2} A(\gamma)\,, \quad D(L) = \frac{LB(L) - \gamma B(\gamma)}{L - \gamma} + \frac{\gamma^2}{\theta - \gamma^2} B(\gamma). \tag{47}$$

We were led to this specification of $A(L)$ and $B(L)$ by an exploration of the likelihood prompted by difficulties in estimating the model with general ARMA(3,1) specifications for $A(L)$ and $B(L)$. We suspect that these difficulties were caused by near cancellations of roots of the numerator and denominator polynomials of our specified ARMAs, together with the presence of a highly persistent autoregressive component. Both maximum likelihood estimation and exploration of a diffuse-prior Bayesian posterior by Markov Chain Monte Carlo methods were much better behaved with the more parsimonious specification.

To determine the relevant unrestricted alternative to (46) and (47), note that for the given $A(L)$ and $B(L)$, the cross-equation restrictions of the evil agent setup cause $C(L)$ to be ARMA(2,2) and $D(L)$ to be ARMA(3,3). Thus the unrestricted model specification has

$$A(L) = \frac{\alpha_0}{(1 - \alpha_1 L)(1 - \alpha_2 L)}\,, \quad B(L) = \frac{\beta_0(1 - \beta_1)}{(1 - L)(1 - \beta_3 L)(1 - \beta_4 L)}. \tag{48}$$

$$C(L) = \frac{\chi_0(1 + \chi_1 L)(1 + \chi_2 L)}{(1 - \chi_3 L)(1 - \chi_4 L)}\,, \quad D(L) = \frac{\delta_0(1 - \delta_1)(1 - \delta_2)(1 - \delta_3)}{(1 - L)(1 - \delta_5 L)(1 - \delta_6 L)}. \tag{49}$$

It is helpful to note the way in which the models (the game model, the unrestricted model, and the least-squares model) are nested. The least-squares model is nested within the game model, by placing a restriction on one parameter: the LS model restricts θ to be ∞. The unrestricted model nests the game model – the cross equation restrictions represent specific restrictions on the values of the χ's and δ's in $C(L)$ and $D(L)$. Because of these nesting relationships, comparison between models can be accomplished with a simple likelihood ratio test.

6.6. **The Model Estimates.**

6.6.1. *The Least-Squares Model.* The estimation of the least-squares model is summarized below.

Parameter	**Estimate**	**Standard Error**
ρ_0	0.6694	0.046
ρ_3	0.9181	0.046
ρ_4	0.1676	0.104
μ_0	0.1667	0.035
μ_1	-0.4785	0.033
μ_3	-0.0390	0.064
μ_4	0.9784	0.016

Log-Likelihood Value: -589.63

Table I. Least-Squares Model Parameter Estimates

6.6.2. *The Game Model.* The estimation of the game model is described below.

Parameter	**Estimate**	**Standard Error**
ρ_0	0.5844	0.039
ρ_3	0.6561	0.112
ρ_4	0.1505	0.162
μ_0	0.2704	0.063
μ_1	-0.6767	0.038
μ_3	0.5249	0.157
μ_4	0.7639	0.122
θ	2.1645	0.135

Log-Likelihood Value: -577.77

Table II. Game Model Parameter Estimates

As can be seen by looking at the estimates of the maximum likelihood of both models, the likelihood ratio test produces a rejection of the "restricted" model, the LS model, in favor of the game model. This rejection is at the 99% level.

6.6.3. *The Unrestricted Time-Series Model.* The estimation of the unrestricted model is described below.

Parameter	**Estimate**	**Standard Error**
α_0	0.5924	0.040
α_1	0.5862	0.225
α_2	0.3091	0.262
β_0	0.1448	0.059
β_1	0.9988	0.981
β_3	0.0451	0.364
β_4	0.0425	0.410
χ_0	4.9994	1.001
χ_1	0.1782	0.077
χ_2	0.9999	0.996
χ_3	-0.4469	0.090
χ_4	0.5127	0.044
δ_0	19.0235	1.042
δ_1	0.6666	0.064
δ_2	-0.6250	0.216
δ_3	-0.6227	0.200
δ_5	-0.2505	0.025
δ_6	0.8350	0.084

Log-Likelihood Value: -569.32

Table III. The Unrestricted Model Parameter Estimates

Due to the fact that the unrestricted model has a total of ten fewer restrictions than the game model, the likelihood ratio test critical value (at 95%) is approximately 18.3. The test statistic in this case is 16.9, well inside the region in which we fail to reject the more restrictive game model in favor of the pure time-series model listed above.

6.6.4. *Analysis of Resuls.* The results show that we reject the LS model in favor of the Game model. This is significant, but not totally unexpected, given that we use an additional parameter. However, we later see that untying the rational expectation cross-equation restrictions, creating the unrestricted model — giving the flexibility of *ten* additional free parameters — this much less restrictive framework generates less of a gain over the Game model than the Game model achieved over the LS model using a single additional parameter.

7. Conclusion

In most modern economic models, agents deal in risk rather than uncertainty. In reality, economic decision-makers are forced to account for both. This paper has placed the agents in the model on the same footing as the authors of the model: the real world contains data generating processes (DGP) for which we have estimates, but not certainties. Through the mechanism of robust prediction and control, agents deal with this uncertainty by making decisions that attempt to be insensitive to misspecifications of the DGP, something that economic decision-makers in the real world could be doing as well to combat this uncertainty.

For the present value model of stock prices, the application of robust decision-making yields a model whose behavior more closely mimics that of the actual data. With robust predictions, asset prices display the "excess volatility" seen in actual stock prices. Thus not only is uncertainty regarding the DGP realistic, it also suggests the resolution of economic puzzle in a plausible manner. The possible resolution of the excess volatility puzzle by such a simple modification in such a simple model suggests that the modification might bear fruit in other, more complex models.

References

Campbell, J. Y., and R. J. Shiller (1987): "Cointegration and Tests of Present Value Models," *Journal of Political Economy*, 95, 1062–1088.

Constantinides, G. M. (1990): "Habit Formation: A Resolution of the Equity Premium Puzzle," *The Journal of Political Economy*, 98(3), 519–543.

DeJong, D. N., and C. H. Whiteman (1989): "The Temporal Stability of Dividends and Stock Prices: Evidence from the Likelihood Function," University of Iowa Department of Economics Working Paper, Number 89-03.

Ellsberg, D. (1961): "Risk, Ambiguity, and the Savage Axioms," *The Quarterly Journal of Economics*, 75(4), 643–669.

Epstein, L. G., and S. Zin (1989): "Substitution, Risk Aversion, and the Temporal Behaviorof Consumption and Asset Returns: A Theoretical Framework," *Econometrica*, 57(4), 937–969.

Flavin, M. (1983): "Excess Volatility in the Financial Markets: A Reassessment of the Empirical Evidence," *The Journal of Political Economy*, 91, 929–956.

Fox, C. R., and A. Tversky (1995): "Ambiguity Aversion and Comparative Ignorance," *The Quarterly Journal of Economics*, 11(3), 585–603.

Gilboa, I., and D. Schmeidler (1989): "Maxmin Expected Utility with Non-Unique Prior," *The Journal of Mathematical Economics*, 18, 141–153.

Hamilton, J. D. (1994): *Time Series Analysis*. Princeton University Press.

Hansen, L. P., and R. Jagannathan (1991): "Implications of security market data for models of dynamic economies," *Journal of Political Economy*, 99, 225–262.

Hansen, L. P., and T. J. Sargent (1980): "Formulating and Estimating Dynamic Linear Rational Expectations Models," *The Journal of Economic Dynamics and Control*, 2, 7–46.

——— (2005): *Misspecification in Recursive Macroeconomic Theory*. Princeton University Press.

Kasa, K. (2001): "A Robust Hansen-Sargent Prediction Formula," *Economic Letters*, 71, 43–48.

LeRoy, S. F., and C. J. LaCivita (1981): "Risk Aversion and the Dispersion of Asset Prices," *The Journal of Business, Vol. 54, No. 4. (Oct., 1981), pp. 535-547*, 54(4), 535–547.

LeRoy, S. F., and R. D. Porter (1981): "The Present Value Relation: Tests Based on Implied Variance Bounds," *Econometrica*, 64, 555–574.

Marsh, T. A., and R. C. Merton (1986): "Dividend Variability and Variance Bounds Tests for the Rationality of Stock Market Prices," *American Economic Review*, 76, 483–498.

Mehra, R., and E. C. Prescott (1985): "The Equity Premium: A Puzzle," *The Journal of Monetary Economics*, 15(2), 145–161.

Michener, R. W. (1982): "Variance Bounds in a Simple Model of Asset Pricing," *The Journal of Political Economy*, 90(1), 166–175.

Otrok, C., B. Ravikumar, and C. H. Whiteman (2002): "Habit Formation: A Resolution to the Equity Premium Puzzle?," *The Journal of Monetary Economics*, 49, 1261–1288.

Shiller, R. J. (1981): "Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?," *The American Economic Review*, 71(3), 421–436.

West, K. D. (1988): "Dividend Innovations and Stock Price Volatility," *Econometrica*, 56(1), 37–61.

Whiteman, C. H. (1983): *Linear Rational Expectations: A User's Guide*. University of Minnesota Press, Minneapolis, MN.

——— (1985): "Spectral Utility, Wiener-Hopf Techniques, and Rational Expectations," *The Journal of Economic Dynamics and Control*, 9, 225–240.

——— (1986): "Analytical Policy Design Under Rational Expectations," *Econometrica*, 54(6), 1387–1405.

Zames, G. (1981): "Feedback and Optimal Sensitivity: Model Reference Transformations, Multiplicative Seminorms, and Approximate Inverses," *IEEE Transactions on Automatic Control*, 16, 301.