Beyond Optimal Forecasting

Richard A. Ashley[1]

Department of Economics

Virginia Tech

November 4, 2006

Abstract

While the conditional mean is known to provide the minimum mean square error (MSE) forecast – and hence is optimal under a squared-error loss function – it must often in practice be replaced by a noisy estimate when model parameters are estimated over a small sample. Here two results are obtained, both of which motivate the use of forecasts biased toward zero (shrinkage forecasts) in such settings. First, the noisy forecast with minimum MSE is shown to be a shrinkage forecast. Second, a condition is derived under which a shrinkage forecast stochastically dominates the unbiased forecast over the class of loss functions monotonic in the forecast error magnitude. The appropriate amount of shrinkage from either perspective depends on a noisiness parameter which must be estimated, however, so the actual reduction in expected losses from shrinkage forecasting is an empirical issue. Simulation results over forecasts from a large variety of multiple regression models indicate that feasible shrinkage forecasts typically do provide modest improvements in forecast MSE when the noise in the estimate of the conditional mean is substantial.

---

[1]Mailing address: Economics Department (0316); Virginia Tech; Blacksburg, VA 24061. E-mail address: ashleyr@vt.edu

1. Introduction

It has long been known that unbiased forecasts are optimal on a squared error criterion – for example, see Granger and Newbold (1977). In other words, the conditional mean of a time series y_t provides the best forecast of y_t, in this limited sense of minimizing the mean square error, or MSE. Except for concerns about the adequacy of the squared error criterion itself, that is a very useful result. However, the conditional mean of y_t is almost never known: in practice it must be replaced with a more or less noisy estimate. This noise arises from the sampling errors inherent in estimating model parameters using finite (and often quite limited) data sets.[2]

It is shown in Section 2 below that the unbiased forecast is no longer squared-error optimal in this setting. Instead, the minimum-MSE forecast is shown to be a shrinkage of the unbiased forecast toward zero – a "mitigated" forecast, in the terminology of Armstrong (1978).

The optimal degree of shrinkage depends on the noisiness of the unbiased forecast – i.e., on the sampling variance in the estimate of the conditional mean. In particular, the amount of shrinkage which is optimal to apply to $\hat{y}_t$, an unbiased forecast of y_t, turns out to depend simply on the square of what one might call $\hat{y}_t$'s "sampling coefficient of variation" – the ratio of its error variance (as an estimator of the conditional mean) to the square of the conditional mean itself.

Since this coefficient of variation must itself be estimated, the practical significance of these results hinges on whether shrinkage forecasts based on estimated values of it provide

[2]Nowadays data sets are sometimes quite large, but the sample period over which the model specification can be taken to be stable is usually much smaller. Errors in estimating/forecasting input values of conditioning variables contribute to this noise in addition. These errors can be substantial also. Indeed, Ashley (1983) provides examples using actual macroeconomic forecasts where these errors overwhelm the model's forecasting ability altogether.

systematic improvements over the original unbiased forecast. Simulation results in Section 3 using a variety of multiple regression models indicate that modest improvements in the forecast mean square error (MSE) can be obtained in this way.

The forgoing results go beyond the usual optimal forecast – i.e., the conditional mean – yet still focus on optimizing expected loss for a particular loss function, the squared error loss function in this case. In contrast, the results developed in Section 4 transcend the framework of optimal forecasting altogether. In these results a standard stochastic dominance theorem from the microeconomics literature is exploited to yield a verifiable necessary and sufficient condition under which a particular shrinkage forecast stochastically dominates the unbiased forecast. Satisfaction of this condition implies that the expected loss from the shrinkage forecast is no larger than that of the unbiased forecast over the entire class of loss functions which are nondecreasing functions of the forecast error magnitude and further implies that its expected loss is strictly less than that of the unbiased forecast for at least one loss function in the class. Calculations using this condition and based on an assumption of normally distributed errors show that shrinkage forecasts in general dominate the unbiased forecast over this class of loss functions. Thus, it can be asserted that shrinkage forecasts are in principle an improvement over the conditional mean in a sense that goes beyond the concept of optimality.

Since the degree of shrinkage for which the shrinkage forecast dominates the unbiased forecast again depends on the square of $\hat{y}_t$'s sampling coefficient of variation, the practical significance of these results again hinges on whether shrinkage forecasts based on estimated values of this coefficient of variation still provide systematic improvements over the unbiased forecast. Simulations in Section 4 which (similar to those of Section 3) examine forecasts from a

variety of multiple regression models, again indicate that modest improvements in the forecast mean square error (MSE) can be obtained in this way.

Section 5 concludes the paper with a discussion of how these results can be implemented in practice.

2. Squared-Error Optimal Shrinkage Forecasts

In this section it is shown that the mean square error of an unbiased, but noisy, forecast is always improved by shrinking it to some degree toward zero.

In particular, suppose that y_t is to be forecast for period T+1 and that the expected value of y_{T+1} can be modeled using a finite data set of consisting of T observations on k explanatory variables and contained in the T × k matrix X. Let $\hat{y}_{T+1}$ be an unbiased estimate of $E\{y_{T+1}|X\}$ provided by this model. {For example, using the notation of the multiple regression model example developed in Section 3, $\hat{y}_{T+1}$ is just $x_{T+1}^t \hat{\beta}^{ols}$, where x_{T+1}^t is a row vector containing the observations on the k explanatory variables for period T+1.} Then

$$\hat{y}_{T+1} = E\{y_{T+1}|X\} + v_{T+1} = \mu_y + v_{T+1} \quad (1)$$

defines μ_y and also defines υ_{T+1}, the error the model makes in estimating the conditional mean of y_{T+1}. Clearly, $E\{v_{T+1}|X\}$ equals zero since $\hat{y}_{T+1}$ is assumed to be an unbiased estimate of $E\{y_{T+1}|X\}$. This error υ_{T+1} might reasonably be called the "sampling error" in $\hat{y}_{T+1}$ since – as the multiple regression model example used in Section 3 makes abundantly clear – it is due to the sampling errors made in estimating the model parameters.[3] The forecast $\hat{y}_{T+1}$ is here called "noisy" insofar as σ_v^2, the variance of this sampling error, is positive.

Next, note that y_{T+1} itself is decomposable into its conditional mean plus an error:

$$y_{T+1} = E\{y_{T+1}|X\} + \epsilon_{T+1} = \mu_y + \epsilon_{T+1} \quad (2)$$

[3]This comment assumes that the observations in the X matrix are taken to be fixed; where they are stochastic, their own sampling variation contributes to the variation in v_{T+1} also.

where $E\{\epsilon_{T+1} | X\}$ equals zero by construction. One might sensibly call ϵ_{T+1} the "instrinsic forecast error" since this is the forecast error that would remain if $E\{y_{T+1} | X\}$ could be obtained without error.

It is assumed here that ϵ_{T+1} and υ_{T+1} are uncorrelated. Since the model for $E\{y_{T+1} | X\}$ could be easily improved by adding additional linear terms if this were not so, this is a reasonable assumption to make. In the context of the multiple regression model example used in Section 3, ϵ_{T+1} and υ_{T+1} are uncorrelated so long as the model error term is serially uncorrelated.

Combining equations (2) and (3), the forecast error, $\hat{y}_{T+1} - y_{T+1}$ can be written:

$$\hat{y}_{T+1} - y_{T+1} = (\mu_y + \nu_{T+1}) - (\mu_y + \epsilon_{T+1}) = \nu_{T+1} - \epsilon_{T+1}$$

yielding

$$\begin{aligned} MSE(\hat{y}_{T+1}) &= E\{(\hat{y}_{T+1} - y_{T+1})^2\} \\ &= E\{\nu_{T+1}^2 - 2\nu_{T+1}\epsilon_{T+1} + \epsilon_{T+1}^2\} \\ &= \sigma_\nu^2 + \sigma_\epsilon^2 \end{aligned}$$

Now consider instead the "shrinkage forecast" $\lambda\hat{y}_{T+1}$, where λ is a parameter to be chosen. It is not necessary to restrict the values of λ to the interval [0,1) but the term "shrinkage" might be inapposite otherwise. The errors made by this forecast are:

$$\begin{aligned} \lambda\hat{y}_{T+1} - y_{T+1} &= \lambda(\mu_y + \nu_{T+1}) - (\mu_y + \epsilon_{T+1}) \\ &= [\lambda - 1]\mu_y + \lambda\nu_{T+1} - \epsilon_{T+1} \end{aligned}$$

Note that $\lambda\hat{y}_{T+1}$ is a biased forecast if λ is unequal to one, since ν_{T+1} and ϵ_{T+1} have mean zero. But if λ is in the interval [0,1), then its error variance is smaller than that of the unbiased forecast, $\hat{y}_{T+1}$. Since the mean square error of a forecast can be decomposed into the sum of its error variance and the square of its bias, this variance reduction raises the possibility that $\lambda\hat{y}_{T+1}$ might have smaller MSE than the unbiased forecast, $\hat{y}_{T+1}$. And it turns out that that $\lambda\hat{y}_{T+1}$ **does** have smaller MSE than the unbiased forecast so long as λ is not too small. In fact, Theorem 1 below shows that the minimum-MSE forecast is always a shrinkage forecast whenever σ_ν^2 is strictly positive:

Theorem 1

Given μ_y, ν_{T+1}, and ϵ_{T+1} as defined above in equations 1 and 2, the shrinkage forecast which minimizes the mean square forecasting error is $\lambda^*\hat{y}_{T+1}$, where λ^* is

$$\lambda^* = \frac{\mu_y^2}{\mu_y^2 + \sigma_\nu^2} = \frac{1}{1 + cv_\nu^2}$$

and $cv_\nu \equiv \sigma_\nu / |\mu_y|$ is what might sensibly be called $\hat{y}_{T+1}$'s coefficient of variation due to its sampling error ν_{T+1}.

Proof: See Appendix 1.

Note that λ^* is clearly less than one so long as σ_ν^2 is strictly positive. In other words, the minimum-MSE forecast of y_{T+1} is shrunk toward zero so long as $E\{y_{T+1} | X\}$, the conditional mean of y_{T+1}, is estimated with error.

In fact, since $dMSE(\lambda\hat{y}_{T+1})/d\lambda$ is positive for all values of $\lambda > \lambda^*$, the value of

$\mathrm{MSE}(\lambda \hat{y}_{T+1})$ is strictly less than $\mathrm{MSE}(\hat{y}_{T+1})$ for all values of λ in the interval $[\lambda^*, 1)$. Evidently, the drop in the error variance as one shrinks down from $\hat{y}_{T+1}$ has a greater impact on the forecast MSE, at least at first, than does the increase in the shrinkage-induced bias.

It is interesting to note that the optimal shrinkage factor (λ^*) depends only on the "noisiness" of $\hat{y}_{T+1}$ as an estimator of the conditional mean $E\{y_{T+1} | X\}$, as quantified by cv_y: it does not depend on σ_ϵ^2, the variance of the intrinsic forecast error, $y_{T+1} - E\{y_{T+1} | X\}$.

Of course, cv_y is not known in practice, so it must be estimated. As the multiple regression example in the next section demonstrates, it is not particularly difficult to estimate cv_y. In fact, one can easily obtain such an estimate using the usual estimated standard error for the unbiased forecast. But if cv_y is too severely under-estimated, then $\lambda^* \hat{y}_{T+1}$ will not have smaller MSE than $\hat{y}_{T+1}$. Moreover, where λ^* is replaced in this way by a sample estimate $\hat{\lambda}^*$, the variance of $\hat{\lambda}^* \hat{y}_{T+1}$ will exceed the variance of $\lambda^* \hat{y}_{T+1}$ because of the sampling variance in $\hat{\lambda}^*$. The practical question thus becomes: can cv_y be estimated sufficiently well that $\hat{\lambda}^* \hat{y}_{T+1}$ is an improvement on $\hat{y}_{T+1}$? This question is addressed in the next section.

3. A Simulation Study of the Performance of the Estimated MSE-minimizing Shrinkage Forecast in Multiple Regression Models

This section examines the performance of the minimum-MSE shrinkage forecast, $\hat{\lambda}^* \hat{y}_{T+1}$, derived above, where y_{T+1} is forecast using a multiple regression model based on k observed explanatory variables and estimated over T sample observations. In this case, as will become evident below, the usual estimate of the standard error of the unbiased forecast can be used to construct a straightforward consistent estimator of the forecast coefficient of variation (cv_y) needed to obtain $\hat{\lambda}^*$ using Theorem 1. The simulation results quoted here quantify the circumstances and degree to which this estimate of cv_y is sufficiently accurate so that the shrinkage forecast $\hat{\lambda}^* \hat{y}_{T+1}$ has lower MSE than the unbiased forecast, $\hat{y}_{T+1}$. Obviously, many real-world forecasts are not obtained from multiple regression models. But many are and, in any case, the results obtained for this type of forecasting model shed light on the applicability and (modest) effectiveness of the shrinkage techniques proposed here.

It is assumed, then, that the column vector of dependent variable observations, $Y = (y_1 \ldots y_T)^t$, is generated by the usual multiple regression model:

$$Y = X\beta + \epsilon \qquad \epsilon \sim N(0, \sigma_\epsilon^2 I_T)$$

where X is a given T×k matrix of full column rank with typical element x_{ij}, containing the sample data on the k explanatory variables, and I_T is a T×T identity matrix. It is assumed that this same model also holds for the forecast period, where t equals T+1.

The 1×k vector of explanatory variable values for period T+1 – i.e., $x_{T+1}^t = (x_{T+1,1} \ldots$

$x_{T+1,k}$) – is assumed to be known but, of course, y_{T+1} is not.[4] Since the forecasts are conditional on X, it and $\mathbf{x}_{T+1}^{t}$ are treated as fixed.

In the context of this model it is well known that least squares estimation yields the estimator

$$\hat{\beta}^{ols} = (X^{t}X)^{-1} X^{t}Y = \beta + (X^{t}X)^{-1} X^{t}\epsilon \sim N(\beta, \sigma_{\epsilon}^{2}(X^{t}X)^{-1})$$

which is unbiased and efficient for β and that the estimator

$$s^{2} = \frac{1}{T-k} \left(Y - X\hat{\beta}^{ols}\right)^{t} \left(Y - X\hat{\beta}^{ols}\right)$$

is unbiased and consistent for σ_{ϵ}^{2}.

Recalling that $\mathbf{x}_{T+1}^{t}$ denotes the given and fixed k-dimensional row vector ($x_{T+1,1}$ ... $x_{T+1,k}$), this model implies that the conditional mean of y_{T+1} is just

$$\mu_{y} \equiv E\left\{y_{T+1} \mid X\right\} = x_{T+1}^{t}\beta$$

and the unbiased forecast of y_{T+1} from this model is

$$\hat{y}_{T+1} = x_{T+1}^{t}\hat{\beta}^{ols} \sim N\left[x_{T+1}^{t}\beta\ ,\ \sigma_{\epsilon}^{2}\left(1 + x_{T+1}^{t}(X^{t}X)^{-1}x_{T+1}\right)\right].$$

[4]Note that the superscript "t" denotes the transpose here and throughout the rest of this paper – it is **not** indexing time period t.

Since ν_{T+1} is $\hat{y}_{T+1} - \mu_y$, it follows that

$$\begin{aligned}
\nu_{T+1} &= x_{T+1}^t \hat{\beta}^{ols} - x_{T+1}^t \beta \\
&= x_{T+1}^t (\hat{\beta}^{ols} - \beta) \\
&= x_{T+1}^t (X^t X)^{-1} X^t \epsilon \\
&\sim N\left[x_{T+1}^t \beta \ , \ \sigma_\epsilon^2 \, x_{T+1}^t (X^t X)^{-1} x_{T+1} \right]
\end{aligned}$$

which shows explicitly that ν_{T+1} in this model arises entirely from the sampling errors in $\hat{\beta}^{ols}$ and in fact is just a weighted sum of $\epsilon_1 \ldots \epsilon_T$, the model errors during the sample. The assumption that var(ϵ) = $\sigma_\epsilon^2 I_T$ implies that $\epsilon_1 \ldots \epsilon_T$ are uncorrelated with each other; since the model is assumed to also hold for period (T+1), this implies that ϵ_{T+1} is uncorrelated with all previous errors and hence with ν_{T+1}.[5]

From expressions in Section 2, $cv_\nu^2 = var(\nu_{T+1}) / \mu_y^2$ can thus be consistently estimated by

$$\hat{cv}_\nu^2 \equiv \frac{s^2 \, x_{T+1}^t (X^t X)^{-1} x_{T+1}}{\left(x_{T+1}^t \hat{\beta}^{ols} \right)^2}$$

and the optimal shrinkage factor $\hat{\lambda}^*$ obtained from $1 / \left(1 + \hat{cv}_\nu^2 \right)$.

By construction, $MSE(\hat{\lambda}^* \hat{y}_{T+1})$ is strictly less than $MSE(\hat{y}_{T+1})$. However, due to the sampling errors contaminating $\hat{cv}_\nu^2$, it is not so clear that $MSE(\hat{\lambda}^* \hat{y}_{T+1})$ is strictly less than $MSE(\hat{y}_{T+1})$. To examine this issue, the multiple regression model described above was simulated N_{rep} = 10,000 times for each of a variety of values of T, k, and σ_ϵ^2. The model

[5] Recall that the derivations of the expressions for the $MSE\left(\hat{y}_{T+1}\right)$ and $MSE\left(\lambda \hat{y}_{T+1}\right)$ in Section 2 require that ν_{T+1} and ϵ_{T+1} are uncorrelated.

coefficients were held fixed ($\beta_j = 1$ for j = 1 ... k), but on each repetition a new set of explanatory variables (X and x_{T+1}) were used. So as to generate a wide variety of different multiple regression models, the components of X and x_{T+1} used on each repetition were generated as independent draws from the unit normal distribution.[6]

Each repetition yielded an observed forecast error from the shrinkage forecast $\left(\hat{\lambda}^* \hat{y}_{T+1} - y_{T+1}\right)$ and from the unbiased forecast $\left(\hat{y}_{T+1} - y_{T+1}\right)$, from which the MSE was calculated for each forecast by averaging appropriately over the N_{rep} simulations. The resulting MSE ratios are tabulated in Table 1; Table 2 displays analogous results where only half as much shrinkage was done – e.g., if $\hat{\lambda}^*$ was .70 for a particular simulation, then a value of .85 was actually used. This option was explored so as to reduce the risk of shrinking too much due to errors in estimating cv_v. The ratios are tabulated only for selected combinations of T and k; more extensive tabulations are available in an unpublished appendix.[7] Also, the reader will note that these tables display the MSE ratios as a function of the model's fit to the sample data, measured by the average observed value of R^2 adjusted for degrees of freedom $\left(\mathbf{R}_c^2\right)$ rather than as a function of σ_ϵ^2; this is more easily interpretable.[8]

---

[6]Except for the first column of X and the first component of x_{T+1}; these were set to one so that β_1 provided an intercept for the regression model. Note that X and x_{T+1} were nevertheless treated as fixed in the estimation and forecasting – there is no contradiction here. To clarify this point, however, it is worth noting that one should expect that the sample variance of $\hat{\beta}^{ols}$ observed across the N_{rep} simulations will exceed $\sigma_\epsilon^2 (X^t X)^{-1}$ due to the fact that variation in $\hat{\beta}^{ols}$ arises both from variation in the model errors and from the fact that each of the N_{rep} models is different because it has a different set of explanatory variables.

[7]Generally speaking, shrinkage is less effective with values of k smaller than those reported here and more effective with values of k larger than those reported here.

[8]The MSE ratio figures for an evenly-spaced grid of $\mathbf{R}_c^2$ values were obtained by interpolation using a regression equation in which the ratio observations over a grid of twenty σ_ϵ^2 values (from 0.5 to 10.0) were fit to a cubic polynomial in the corresponding average $\mathbf{R}_c^2$ values.

The results in Table 1 indicate that the minimum-MSE shrinkage forecast $\left(\hat{\lambda}^{*}\hat{y}_{T+1}\right)$ obtained using the estimated value of cv_v provides notably more accurate forecasts where σ_v^2, the variance in the errors in estimating the conditional mean of y_{T+1} is substantial – i.e. for ill-fitting models and small samples. For models with larger values of T and R_c^2, however, using the full amount of shrinkage indicated by the estimated value of cv_v may not always improve on the MSE of the unbiased forecast. This is because the available MSE reduction is smaller in such cases and hence more easily overwhelmed by over-shrinkage due to an overestimate of cv_v. Consequently, the empirical performance of a less ambitious shrinkage forecast, which shrinks the unbiased estimator by only half as much as would be optimal if the estimate of cv_v were error-free, is examined in Table 2. This shrinkage forecast is less effective for very weak models, but more effective – and not very risky – for stronger models.

Overall, the forecast efficiency gains from shrinkage are – as indicated at the outset – modest, except in instances where the conditional mean is quite noisy due to poor fit and/or small estimation samples. On the other hand, the shrinkage estimator is very easy to implement and these simulations indicate that there **are** gains to be had from this source.

Table 1

MSE Reduction Using Full Shrinkage Based on Estimated cv_v

$$\text{Ratio } \text{MSE}\left[\hat{\lambda}^{*}\hat{y}_{T+1}\right] / \text{MSE}\left[\hat{y}_{T+1}\right] \text{ versus } R_c^2$$

R_c^2	T = 10	T = 20	T = 40	T = 60	T = 80
	k = 3	k = 5	k = 8	k = 10	k = 10
0.05	0.803	0.856	0.897	0.921	0.947
0.10	0.819	0.870	0.910	0.933	0.956
0.15	0.835	0.885	0.923	0.943	0.964
0.20	0.850	0.898	0.934	0.953	0.972
0.25	0.864	0.911	0.944	0.961	0.978
0.30	0.878	0.924	0.954	0.969	0.984
0.35	0.892	0.935	0.962	0.975	0.988
0.40	0.905	0.946	0.970	0.981	0.992
0.45	0.917	0.956	0.977	0.986	0.995
0.50	0.930	0.966	0.983	0.991	0.998
0.55	0.941	0.975	0.988	0.994	1.000
0.60	0.953	0.983	0.992	0.998	1.001
0.65	0.964	0.990	0.996	1.000	1.003
0.70	0.975	0.996	1.000	1.002	1.003
0.75	0.986	1.002	1.002	1.004	1.004
0.80	0.996	1.007	1.005	1.005	1.004
0.85	1.006	1.011	1.007	1.006	1.004
0.90	1.017	1.014	1.008	1.006	1.004
0.95	1.027	1.016	1.010	1.006	1.004

Table 2

MSE Reduction Using Half Shrinkage Based on Estimated cv_v

$$\text{Ratio MSE}\left[\left(\frac{\hat{\lambda}^{*}+1}{2}\right)\hat{y}_{T+1}\right] / \text{MSE}\left[\hat{y}_{T+1}\right] \text{ versus } R_c^2$$

R_c^2	T = 10	T = 20	T = 40	T = 60	T = 80
	k = 3	k = 5	k = 8	k = 10	k = 10
0.05	0.885	0.916	0.939	0.953	0.968
0.10	0.893	0.923	0.946	0.959	0.973
0.15	0.901	0.930	0.952	0.964	0.977
0.20	0.908	0.937	0.958	0.969	0.981
0.25	0.916	0.943	0.963	0.973	0.984
0.30	0.922	0.950	0.968	0.977	0.987
0.35	0.929	0.955	0.972	0.981	0.989
0.40	0.936	0.961	0.976	0.984	0.991
0.45	0.942	0.966	0.980	0.987	0.993
0.50	0.949	0.971	0.983	0.989	0.994
0.55	0.955	0.976	0.986	0.991	0.996
0.60	0.961	0.980	0.988	0.993	0.997
0.65	0.967	0.984	0.991	0.995	0.997
0.70	0.973	0.988	0.993	0.996	0.998
0.75	0.979	0.991	0.995	0.997	0.999
0.80	0.985	0.994	0.997	0.998	0.999
0.85	0.991	0.997	0.998	0.999	0.999
0.90	0.998	1.000	1.000	1.000	1.000
0.95	1.004	1.002	1.001	1.001	1.000

4. The Shrinkage Forecast Which Stochastically Dominates the Unbiased Forecast

Ashley (1990) uses a standard result from stochastic dominance theory to obtain an explicit condition under which $\lambda\hat{\beta}^{unbiased}$, the shrinkage of an unbiased estimator of a parameter β, stochastically dominates $\hat{\beta}^{unbiased}$ itself. If this necessary and sufficient condition is satisfied for a particular value of the shrinkage parameter (λ), then the expected loss from $\lambda\hat{\beta}^{unbiased}$ is no larger than that of $\hat{\beta}^{unbiased}$ for any loss function in the class of all loss functions which are continuous nondecreasing functions of $M(v; \omega_{asy})$, the generalized magnitude of v, the estimation error; and this expected loss is strictly less than that of $\hat{\beta}^{unbiased}$ for at least one loss function in the class.[9] The generalized error magnitude is defined as:

$$M[x; \omega_{asy}] = x \quad \text{for } x > 0$$
$$= -\omega_{asy} x \quad \text{otherwise}$$

so that ω_{asy} quantifies the degree of asymmetry in the loss function.

Theorem 2, derived in Appendix 2 below, provides an analogous necessary and sufficient condition under which $\lambda\hat{y}_{T+1}$, the shrinkage of an unbiased forecast of y_{T+1}, stochastically dominates $\hat{y}_{T+1}$ itself. If this condition is satisfied for a particular value of the shrinkage parameter (λ), then the expected loss from $\lambda\hat{y}_{T+1}$ is no larger than that of $\hat{y}_{T+1}$ for any loss function in the class of all loss functions which are continuous nondecreasing functions of $M[\lambda\hat{y}_{T+1} - y_{T+1}; \omega_{asy}]$, the generalized magnitude of the forecast error; and this expected loss is strictly less than that of $\hat{y}_{T+1}$ for at least one loss function in the class. In Section 2 this forecast

[9] It should be noted that – largely to accommodate the standard notation used in multiple regression modeling – there are several notational differences between the present paper and Ashley (1990). In particular, the variables here called ω_{asy}, v, λ, and cv_v are there called λ, ϵ, k, and t^{-1}.

error was shown to be

$$\lambda \hat{y}_{T+1} - y_{T+1} = (\lambda - 1)\mu_y + \lambda \nu_{T+1} - \epsilon_{T+1}$$

where μ_y is the conditional mean of y_{T+1}, ν_{T+1} is what was called the "sampling error" in $\hat{y}_{T+1}$ (i.e., $\hat{y}_{T+1} - \mu_y$), and ϵ_{T+1} is what was called the "intrinsic forecast error" in $\hat{y}_{T+1}$ (i.e., $y_{T+1} - \mu_y$). Theorem 2 is a non-trivial extension of the derivation in Ashley (1990) due the presence here of a second random term (ϵ_{T+1}) in the expression for the forecast error.

Theorem 2:

Given μ_y, ν_{T+1}, and ϵ_{T+1} as defined in equations 1 and 2 of Section 2, the shrinkage forecast $\lambda \hat{y}_{T+1}$ stochastically dominates the unbiased forecast $\hat{y}_{T+1}$ over the class of loss functions which are nondecreasing functions of the generalized forecast error magnitude $M[\lambda \hat{y}_{T+1} - y_{T+1}; \omega_{asy}]$ if and only if

$$\int_{-\infty}^{\infty} \left\{ \tilde{F}_\nu\left(\frac{\tau}{\lambda} - \frac{(\lambda - 1)}{\lambda cv_\nu} + \frac{cv_\epsilon}{\lambda cv_\nu} z \right) - \tilde{F}_\nu\left(-\frac{\tau \omega_{asy}}{\lambda} - \frac{(\lambda - 1)}{\lambda cv_\nu} + \frac{cv_\epsilon}{\lambda cv_\nu} z \right) \right\} \tilde{g}(z)\, dz$$

is greater than or equal to

$$\int_{-\infty}^{\infty} \left\{ \tilde{F}_\nu\left(\tau + \frac{cv_\epsilon}{cv_\nu} z \right) - \tilde{F}_\nu\left(-\tau \omega_{asy} + \frac{cv_\epsilon}{cv_\nu} z \right) \right\} \tilde{g}(z)\, dz$$

for all non-negative τ, with strict inequality holding for at least one value of τ and where

(1) $\tilde{F}_\nu(\nu_{T+1}/\sigma_\nu)$ is the standardized cumulative distribution function of ν_{T+1} conditional on the value of ϵ_{T+1},

(2) $\tilde{g}_{\epsilon}(\epsilon_{T+1}/\sigma_{\epsilon})$ is the standardized density function for ϵ_{T+1},

(3) $cv_{\nu} \equiv \sigma_{\nu}/|\mu_{y}|$, or "$\hat{y}_{T+1}$'s coefficient of variation due to the sampling error ν_{T+1},"

and

(4) $cv_{\epsilon} \equiv \sigma_{\epsilon}/|\mu_{y}|$, or "$\hat{y}_{T+1}$'s coefficient of variation due to the intrinsic forecast error ϵ_{T+1}."

Proof: See Appendix 2.

This condition for stochastic dominance is readily checked for the special case where ν_{T+1} and ϵ_{T+1} are uncorrelated gaussian variates: good numerical approximations are available for the cumulative distribution function of a unit normal variate (the function $\tilde{F}_{\nu}$) and the remaining integration over $z = \epsilon_{T+1}/\sigma_{\epsilon}$ is not troublesome. Results of such calculations are given in Table 3 for the symmetric case, where ω_{asy} is set to one. What is displayed there is λ_{sdom}, the lower limit of an "unbiasedness dominating interval" – an interval containing all of the values of λ for which the shrinkage forecast $\lambda\hat{y}_{T+1}$ stochastically dominates the unbiased forecast $\hat{y}_{T+1}$. These results were obtained by decrementing the value of λ from 1.00 in steps of .01 until the condition of Theorem 2 is no longer satisfied.[10]

Note that Table 3 tabulates λ_{sdom} for different values of $cv_{\epsilon} \equiv \sigma_{\epsilon}/|\mu_{y}|$, as well as for different values of cv_{ν}. In contrast, recall that the minimum-MSE shrinkage forecast does not depend on σ_{ϵ}^2, the variance of the intrinsic forecast error. It is also worth noting that the shrinkage factor which is optimal for the squared error loss function is generally similar to λ_{sdom} when σ_{ϵ}^2 is relatively small.

[10] Similar calculations could be done for other values of ω_{asy}.

The minimum-MSE shrinkage factor can be either above or below the lower limit of the unbiasedness dominating interval. This is not an error in the calculations. It is possible for the minimum-MSE shrinkage factor to lie outside the unbiasedness dominating interval since what is optimal for the squared error loss function could, for some other loss function, yield a forecast with higher expected loss than the unbiased forecast. And the minimum-MSE shrinkage ratio can lie well inside the unbiasedness dominating interval since shrinkage factors well below the optimal value can still have MSE less than that of the unbiased forecast.

It is also noteworthy that the value of λ_{sdom} fairly suddenly approaches one (the unbiased forecast) when cv_ϵ exceeds a threshold value. Still, the main thing to be learned from these results is that fairly substantial amounts of shrinkage stochastically dominate the unbiased forecast when the unbiased forecast is itself noisy (so that cv_y is substantial) and the intrinsic forecast error $(\hat{y}_{T+1} - y_{T+1})$ is not too large, so that cv_ϵ is small.

Table 3

λ_{sdom}, the Lower Limit of the Unbiasedness Dominating Interval for Gaussian Errors and Symmetric Loss Functions[11]

cv_v	min MSE	cv_ϵ =0.00	cv_ϵ= 0.25	cv_ϵ= 0.50	cv_ϵ =0.75	cv_ϵ =1.00	cv_ϵ =1.25	cv_ϵ =1.50	cv_ϵ =1.75	cv_ϵ =2.00
0.25	0.94	0.90	0.89	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.50	0.80	0.71	0.69	0.67	1.00	1.00	1.00	1.00	1.00	1.00
0.75	0.64	0.55	0.52	0.44	0.49	1.00	1.00	1.00	1.00	1.00
1.00	0.50	0.44	0.41	0.30	0.32	0.38	0.88	1.00	1.00	1.00
1.25	0.39	0.36	0.33	0.24	0.19	0.26	0.30	0.81	1.00	1.00
1.50	0.31	0.31	0.28	0.20	0.13	0.17	0.22	0.25	0.77	1.00
1.75	0.25	0.26	0.23	0.17	0.11	0.09	0.16	0.18	0.20	0.74
2.00	0.20	0.23	0.20	0.15	0.10	0.05	0.11	0.13	0.15	0.54

[11]The "min MSE" column gives the optimal shrinkage ratio for a squared error loss function, $[1 + (cv_v)^2]^{-1}$.

The effectiveness of choosing a forecast shrinkage factor either at the lower limit of the unbiasedness dominating interval or in the middle of this interval is investigated here using the same simulated multiple regression forecasts as in Section 3.

In these simulations cv_v is estimated in the same way as before. The stochastic dominance condition provided by Theorem 2 also requires a value for cv_ϵ, however. But note that there is no need to estimate cv_ϵ since it enters the stochastic dominance condition only as the ratio cv_ϵ/cv_v, which is completely determined by the observed data on the explanatory variables:

$$\frac{cv_\epsilon}{cv_v} = \frac{\sigma_\epsilon}{\sigma_v} = 1 / \sqrt{x_{T+1} (X^t X)^{-1} x_{T+1}^t}$$

Again, as in the Table 3, the results are presented for the special case of symmetric loss functions (ω_{asy} equal to one) and for both the fully shrunken value of $\hat{\lambda}_{sdom}$ – i.e., the smallest estimate of λ_{sdom} for which $\hat{\lambda}_{sdom} \hat{y}_{T+1}$ dominates $\hat{y}_{T+1}$ – and for a less ambitious shrinkage forecast which is shrunk only half as much away from the unbiased forecast. These results are displayed in Tables 4 and 5.

Comparing the results in Tables 4 and 5 to the analogous results in Tables 1 and 2, where the minimum-MSE shrinkage factor was used, it is immediately evident that shrinkage forecasts based on $\hat{\lambda}_{sdom}$ are less effective in terms of MSE reduction. This is as one might expect since the minimum-MSE shrinkage factors are obviously optimized for this particular criterion. This effect is minor for small samples (T = 10 and T = 20), but quite marked for sample sizes much larger than this. What is going on is that the average amount of shrinkage being applied is declining quickly to zero (i.e., $\hat{\lambda}_{sdom}$ is quickly increasing toward one) as T increases because

this increase in sample length increases the precision with which the model parameters are estimated, decreasing cv_v. This causes the ratio cv_ϵ/cv_v to often lie beyond the threshold value observed in Table 3, where λ_{sdom} suddenly increases to one.

It seems reasonable to conclude that $(\hat{\lambda}_{sdom} + 1)/2$ provides reasonably useful shrinkage forecasts, which are almost as effective in MSE terms as the minimum-MSE shrinkage forecasts, for T = 10 or T = 20, but that the price exacted for the additional generality (in terms of lower expected loss over the entire class of loss functions which are nondecreasing functions of the error magnitude) is too high for sample sizes much larger than this.

Table 4

MSE Reduction Using Full Stochastic Dominance Shrinkage Based on Estimated cv_v

$$\text{Ratio } \text{MSE}\left[\hat{\lambda}_{sdom}\,\hat{y}_{T+1}\right] / \text{MSE}\left[\hat{y}_{T+1}\right] \text{ versus } R_c^2$$

R_c^2	T = 10	T = 20	T = 40	T = 60
	k = 3	k = 5	k = 8	k = 10
0.05	0.816	0.921	0.988	0.998
0.1	0.827	0.926	0.987	0.998
0.15	0.838	0.931	0.987	0.999
0.2	0.850	0.937	0.987	0.999
0.25	0.862	0.942	0.988	0.999
0.3	0.874	0.948	0.989	0.999
0.35	0.887	0.955	0.990	0.999
0.4	0.899	0.961	0.991	0.999
0.45	0.912	0.967	0.993	0.999
0.5	0.925	0.973	0.994	0.999
0.55	0.939	0.980	0.996	0.999
0.6	0.952	0.986	0.997	0.999
0.65	0.966	0.991	0.999	0.999
0.7	0.980	0.997	1.001	0.999
0.75	0.993	1.002	1.002	1.000
0.8	1.007	1.006	1.003	1.000
0.85	1.021	1.010	1.004	1.000
0.9	1.035	1.014	1.004	1.000
0.95	1.049	1.016	1.005	1.001

Table 5

MSE Reduction Using Half Stochastic Dominance Shrinkage Based on Estimated cv_v

$$\text{Ratio MSE}\left[\left(\frac{\hat{\lambda}_{sdom}+1}{2}\right)\hat{y}_{T+1}\right] / \text{MSE}\left[\hat{y}_{T+1}\right] \text{ versus } R_c^2$$

R_c^2	T = 10	T = 20	T = 40	T = 60
	k = 3	k = 5	k = 8	k = 10
0.05	0.890	0.953	0.993	0.999
0.1	0.895	0.956	0.993	0.999
0.15	0.900	0.958	0.992	0.999
0.2	0.906	0.961	0.992	0.999
0.25	0.911	0.963	0.993	0.999
0.3	0.917	0.966	0.993	0.999
0.35	0.922	0.969	0.993	0.999
0.4	0.928	0.971	0.994	0.999
0.45	0.934	0.974	0.995	0.999
0.5	0.940	0.977	0.995	0.999
0.55	0.946	0.980	0.996	1.000
0.6	0.953	0.983	0.997	1.000
0.65	0.960	0.986	0.998	1.000
0.7	0.967	0.989	0.999	1.000
0.75	0.974	0.992	0.999	1.000
0.8	0.982	0.994	1.000	1.000
0.85	0.990	0.997	1.001	1.000
0.9	0.998	1.000	1.001	1.000
0.95	1.007	1.002	1.002	1.000

5. Conclusions

From a conceptual point of view, this paper makes two contributions:

1. Where only a noisy estimate of the conditional mean is available – due, for example, to errors in estimating model parameters or to errors in measuring/forecasting explanatory variables needed for the forecast – it is shown that the optimal forecast on a squared-error criterion is always a shrinkage of the estimated conditional mean (the unbiased forecast) toward zero. The optimal degree of shrinkage depends in a simple way on the amount of noise in the conditional mean estimate, but not at all on the dispersion of the value to be forecast around its conditional mean.

And:

2. In an analysis transcending the concept of optimality altogether, a fairly simple condition is derived under which a given shrinkage of the conditional mean estimate provides a forecast which stochastically dominates the unbiased forecast. When this condition is satisfied, this shrinkage forecast has expected loss no larger than that of the estimated conditional mean for all loss functions in the class of loss functions which are nondecreasing functions of the generalized magnitude of the forecast error, and it yields expected loss strictly less than that of the estimated conditional mean for at least one loss function in this class. This stochastic dominance condition again depends on the amount of noise in the conditional mean estimate, but it also depends on the dispersion of the value to be forecast around its conditional mean. Calculations based on a restriction to gaussian errors and symmetric loss functions indicate that the maximum amount of shrinkage for which the forecast still dominates the unbiased forecast is similar to that which minimizes the MSE so long as the dispersion of the value to be forecast

around the unbiased forecast is substantially smaller than the sampling variance of the conditional mean estimator.

These results are consistent with recent work in the empirical macroeconomics literature, such as Dave (2004), which finds pervasive evidence for biased expectations in the investment spending behavior of individual Canadian manufacturing firms.

From an applications point of view, the extensive simulations obtained here using a variety of multiple regression forecasting models indicate that useful empirical approximations to the minimum-MSE shrinkage forecast can be readily obtained, but that the estimates of the forecast stochastically dominating the unbiased forecast are of practical use only for very small samples. In particular, the results reported in Tables 1 and 2 indicate that the empirical approximations to the minimum-MSE shrinkage forecast can provide modest but non-negligible MSE reductions in a variety of circumstances where the noise in the estimated conditional mean forecast is substantial – i.e., where the sample is fairly small and adjusted R^2 is not too large.

References

Armstrong, J. Scott (1978) *Long Range Forecasting* New York: Wiley-Interscience.

Ashley, Richard (1983) "On the Usefulness of Macroeconomic Forecasts as Inputs to Forecasting Models" *Journal of Forecasting 2*, pp. 211 - 223.

Ashley , Richard (1990) "Shrinkage Estimation with General Loss Functions: An Application of Stochastic Dominance Theory" *International Economic Review 31(2)*, pp. 301-313.

Blackwell, D. (1951) "Comparison of Experiments" in *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability* Berkeley: University of California Press, pp. 93 -102.

Dave, C. (2004) "Are Investment Expectations Adaptive, Rational or Neither?" (Unpublished manuscript).

Granger, C. W. J. and Newbold, P. (1977) *Forecasting Economic Time Series* New York: Academic Press.

Hadar, J. and W. R. Russell (1969) "Rules for Ordering Uncertain Prospects" *American Economic Review 54*, pp. 25-34.

Tesfatsion, L. (1976) "Stochastic Dominance and the Maximization of Expected Utility" *Review of Economic Studies 43*, pp. 301 - 315.

Appendix 1
Proof of Theorem 1

Theorem 1

Given μ_y, ν_{T+1}, and ϵ_{T+1} as defined above in equations 1 and 2, the shrinkage forecast which minimizes the mean square forecasting error is $\lambda^* \hat{y}_{T+1}$, where λ^* is

$$\lambda^* = \frac{\mu_y^2}{\mu_y^2 + \sigma_\nu^2} = \frac{1}{1 + cv_\nu^2}$$

and $cv_\nu \equiv \sigma_\nu / |\mu_y|$ is what might sensibly be called $\hat{y}_{T+1}$'s coefficient of variation due to its sampling error ν_{T+1}.

Proof:

The mean square error of the shrinkage forecast is thus:

$$\begin{aligned} MSE(\lambda \hat{y}_{T+1}) &= E\left\{ \left([\lambda - 1]\mu_y + \lambda \nu_{T+1} - \epsilon_{T+1} \right)^2 \right\} \\ &= [\lambda - 1]^2 \mu_y^2 + \lambda^2 \sigma_\nu^2 + \sigma_\epsilon^2 \end{aligned}$$

so that

$$\frac{dMSE(\lambda \hat{y}_{T+1})}{d\lambda} = 2[\lambda - 1]\mu_y^2 + 2\lambda \sigma_\nu^2$$

Consequently, setting

$$\left(\frac{dMSE(\lambda \hat{y}_{T+1})}{d\lambda} \right)_{\lambda = \lambda^*} = 0$$

implies that

$$\lambda^* = \frac{\mu_y^2}{\mu_y^2 + \sigma_v^2} = \frac{1}{1 + cv_v^2}$$

where $cv_v \equiv \sigma_v / |\mu_y|$. This value of λ is clearly a global minimum since

$$d^2 MSE(\lambda \hat{y}_{T+1}) / d\lambda^2 = 2\left(\sigma_v^2 + \mu_y^2\right)$$

is positive.

Appendix 2
Proof of Theorem 2

Theorem 2

Given μ_y, ν_{T+1}, and ϵ_{T+1} as defined in equations 1 and 2 of Section 2, the shrinkage forecast $\lambda\hat{y}_{T+1}$ stochastically dominates the unbiased forecast $\hat{y}_{T+1}$ over the class of loss functions which are nondecreasing functions of the generalized forecast error magnitude $M[\lambda\hat{y}_{T+1} - y_{T+1}; \omega_{asy}]$ if and only if

$$\int_{-\infty}^{\infty} \left\{ \tilde{F}_\nu\left(\frac{\tau}{\lambda} - \frac{(\lambda-1)}{\lambda cv_\nu} + \frac{cv_\epsilon}{\lambda cv_\nu} z \right) - \tilde{F}_\nu\left(-\frac{\tau\omega_{asy}}{\lambda} - \frac{(\lambda-1)}{\lambda cv_\nu} + \frac{cv_\epsilon}{\lambda cv_\nu} z \right) \right\} \tilde{g}(z)\, dz$$

is greater than or equal to

$$\int_{-\infty}^{\infty} \left\{ \tilde{F}_\nu\left(\tau + \frac{cv_\epsilon}{cv_\nu} z \right) - \tilde{F}_\nu\left(-\tau\omega_{asy} + \frac{cv_\epsilon}{cv_\nu} z \right) \right\} \tilde{g}(z)\, dz$$

for all non-negative τ, with strict inequality holding for at least one value of τ and where

(1) $\tilde{F}_\nu(\nu_{T+1}/\sigma_\nu)$ is the standardized cumulative distribution function of ν_{T+1} conditional on the value of ϵ_{T+1},

(2) $\tilde{g}_\epsilon(\epsilon_{T+1}/\sigma_\epsilon)$ is the standardized density function for ϵ_{T+1},

(3) $cv_\nu \equiv \sigma_\nu/|\mu_y|$, or "$\hat{y}_{T+1}$'s coefficient of variation due to the sampling error ν_{T+1},"

and

(4) $cv_\epsilon \equiv \sigma_\epsilon/|\mu_y|$, or "$\hat{y}_{T+1}$'s coefficient of variation due to the intrinsic forecast error ϵ_{T+1}."

Proof:

Stochastic dominance has been defined many times, going back to Blackwell (1951), Hadar and Russell (1969), and Tesfatsion (1976). The basic idea is always the same: a random variable x stochastically dominates another random variable y in size if the cumulative distribution function of x lies entirely above (or, equivalently, to the left of) the cumulative distribution function of y. This amounts to requiring that the probability that x exceeds a given value τ exceeds the probability that y exceeds τ for all values of τ. This characterizes the notion that the random variable x "is larger than" the random variable y.

Here the random variables at issue are the generalized magnitudes of the losses associated with each of the two forecasts, so the definition of dominance is amended in an obvious way to provide an "is smaller than" notion. Thus, the shrinkage forecast $\lambda \hat{y}_{T+1}$ stochastically dominates the unbiased forecast $\hat{y}_{T+1}$ if and only if

$$\text{probability}\left\{M[\lambda \hat{y}_{T+1} - y_{T+1}; \omega_{asy}] \leq \tau\right\} \geq \text{probability}\left\{M[\hat{y}_{T+1} - y_{T+1}; \omega_{asy}] \leq \tau\right\}$$

for all non-negative values of τ, with strict inequality holding for at least one value. (Negative values of τ need not be considered since the generalized magnitude function is inherently non-negative.)

Re-expressing the two forecast errors in terms of μ_y, ν, and ϵ (and dropping the time subscripts for simplicity) , this condition becomes:

$$\text{probability}\left\{M\left[(\lambda - 1)\mu + \lambda \nu - \epsilon\ ;\ \omega_{asy}\right] \leq \tau\right\}$$

$$\geq \text{probability}\left\{M\left[\nu - \epsilon\ ;\ \omega_{asy}\right] \leq \tau\right\}$$

for all non-negative values of τ, with strict inequality holding for at least one value.

Expressing the joint density of ν and ϵ as $f_\nu(\nu \mid \epsilon)g(\epsilon)$ using Bayes' Theorem,

$$\text{probability}\{M[(\lambda - 1)\mu + \lambda\nu - \epsilon\ ;\ \omega_{asy}] \le \tau\}$$

$$= \int_{-\infty}^{\infty} \int_{[-\tau\omega_{asy} - (\lambda-1)\mu + \epsilon]/\lambda}^{[\tau - (\lambda-1)\mu + \epsilon]/\lambda} f_\nu(\nu \mid \epsilon)\, g(\epsilon)\, d\nu\, d\epsilon$$

where the upper limit of the integral over ν is the value of ν just large enough that $(\lambda\text{-}1) + \lambda\nu - \epsilon$ equals τ and the lower limit is the value of ν just small enough that $-\{(\lambda\text{-}1) + \lambda\nu - \epsilon\}$ equals $\tau\omega_{asy}$. This probability can be expressed in terms of the cumulative distribution function of ν as:

$$\int_{-\infty}^{\infty} \left\{ F_\nu\left(\left[\tau - (\lambda-1)\mu + \epsilon\right]/\lambda\right) - F_\nu\left(\left[-\tau\omega_{asy} - (\lambda-1)\mu + \epsilon\right]/\lambda\right)\right\} g(\epsilon)\, d\epsilon$$

Noting that ν and ϵ have means of zero and variances of σ_ν^2 and σ_ϵ^2, respectively, it is useful to re-state this expression in terms of the standardized conditional distribution function of ν, $\tilde{F}_\nu(\nu)$, and the standardized density function of ϵ, $\tilde{g}_\epsilon(\epsilon)$:

$$\int_{-\infty}^{\infty} \left\{ \tilde{F}_\nu\left(\frac{\tau}{\lambda\sigma_\nu} - \frac{(\lambda-1)}{\lambda cv_\nu} + \frac{cv_\epsilon}{\lambda cv_\nu} z \right) - \tilde{F}_\nu\left(-\frac{\tau\omega_{asy}}{\lambda\sigma_\nu} - \frac{(\lambda-1)}{\lambda cv_\nu} + \frac{cv_\epsilon}{\lambda cv_\nu} z \right) \right\} \tilde{g}(z)\, dz$$

where cv_ϵ / cv_ν has been substituted for the ratio $\sigma_\epsilon / \sigma_\nu$.

Thus, the shrinkage forecast $\lambda\hat{y}_{T+1}$ stochastically dominates the unbiased forecast $\hat{y}_{T+1}$ if and only if

$$\int_{-\infty}^{\infty} \left\{ \tilde{F}_v\left(\frac{\tau}{\lambda \sigma_v} - \frac{(\lambda - 1)}{\lambda cv_v} + \frac{cv_\epsilon}{\lambda cv_v} z \right) - \tilde{F}_v\left(-\frac{\tau \omega_{asy}}{\lambda \sigma_v} - \frac{(\lambda - 1)}{\lambda cv_v} + \frac{cv_\epsilon}{\lambda cv_v} z \right) \right\} \tilde{g}(z)\, dz$$

is greater than or equal to

$$\int_{-\infty}^{\infty} \left\{ \tilde{F}_v\left(\frac{\tau}{\sigma_v} + \frac{cv_\epsilon}{cv_v} z \right) - \tilde{F}_v\left(-\frac{\tau \omega_{asy}}{\sigma_v} + \frac{cv_\epsilon}{cv_v} z \right) \right\} \tilde{g}(z)\, dz$$

for all non-negative τ, with strict inequality holding for at least one value of τ. Since this condition must hold for all non-negative values of τ, the ratio τ/σ_v in these expressions can be replaced by τ, yielding the result of the theorem.