INTERNATIONAL ECONOMIC REVIEW
Vol. 31, No. 2, May 1990

SHRINKAGE ESTIMATION WITH GENERAL LOSS FUNCTIONS: AN APPLICATION OF STOCHASTIC DOMINANCE THEORY

BY RICHARD ASHLEY[1]

Shrinkage estimation is analyzed using stochastic dominance theory over a broad class of loss functions. (Neither symmetry nor boundedness is imposed.) A recommended shrinkage factor interval is calculated for gaussian, unbiased estimators based on this analysis. Since the minimum MSE estimator is generally found to lie within this interval for $t \geq 1$, these results justify the minimum MSE criterion as a desideratum over a wide class of loss functions. Also, the unbiased estimator is found to be dominated by shrunken (biased) estimators over a number of loss function classes. This implies that the unbiased linear projections used to model expectations formation in neoclassical macroeconomic models are stochastically dominated by biased expectations. Finally, practical shrinkage factors are given which are shown to provide modest improvements in expected losses over a wide range of symmetric and asymmetric loss functions.

1. INTRODUCTION

i. *Context.* Most econometricians would agree that the ideal estimator is that which minimizes the expected loss. Of course, this ideal is difficult or impossible to achieve in practice but, in fact, the ideal itself is defective because it is almost always predicated on an arbitrarily chosen (usually quadratic) loss function. Put another way, optimality per se is not all that meaningful if the loss function involved represents the analyst's convenience rather than the end user's preferences.

Remarkably enough, this observation does not necessarily lead to a dead end. After all, we do know *something* about the loss function. For example, it is ordinarily reasonable to assume that a larger estimation error will never lead to a smaller loss. One might additionally be willing to assume something about the symmetry of the loss function.

Such restrictions are used below, in conjunction with standard results from stochastic dominance theory, to obtain quite explicit results on the particular problem of how much one should shrink an unbiased parameter estimate toward zero. For example, the rightmost column of Table 1 below gives an interval of "best" shrinkage factors to apply to an unbiased gaussian parameter estimate, assuming little more than that the loss function is symmetric in the estimation errors; similar results are given for asymmetric loss functions in Table 2.

Clearly, the concept used here to define "best" must transcend optimality, since the concept of optimality is meaningful only in the context of a particular loss function. Since our ignorance of the precise forms of loss functions and utility

[1] The author wishes to thank Jacques Cremer, Douglas McManus, Douglas Patterson, Eric Smith, Richard Steinberg, Warren Weber, and an anonymous referee for helpful comments.

functions is profound throughout economics, some variation on the notion of "best" defined below may prove useful in areas well beyond the confines of shrinkage estimation.

In particular, the calculations reported below indicate that unbiased gaussian estimators are stochastically dominated by shrunken (biased) estimators for large classes of reasonable loss functions. Thus, these results have something to say about the reasonableness of the expectations formation mechanism characteristically employed in rational expectations models. This issue is discussed further at the end of the paper.

ii. *Overview.* Suppose that $\hat{\beta}$ is an unbiased estimator of β, where β is one component of a vector of unknown parameters. Alternatively, β could equally well be any given linear combination of the components of this unknown vector, as when $\hat{\beta}$ is a point forecast from a linear regression model.

Letting t denote the population t ratio of $\hat{\beta}$ (i.e. $t \equiv \beta/\{\mathrm{Var}\ (\hat{\beta})\}^{.5}$) it is easily shown that the shrinkage estimator,

$$\tilde{\beta}(k) \equiv k\hat{\beta}, \tag{1}$$

with

$$k = t^2/(1 + t^2) \equiv k_{MSE} \tag{2}$$

always has smaller mean squared error (MSE) than does $\hat{\beta}$. In fact, k_{MSE} minimizes $\mathrm{MSE}\{\tilde{\beta}(k)\}$ over all k for given t.[2] It is thus always optimal to shrink the unbiased estimator toward zero under a squared error loss function.

This observation is hardly new—rather it has motivated the development of a number of estimation methods at least in part designed to exploit this potential for improving on the MSE of the unbiased estimator by shrinking it toward zero. Examples include the James/Stein estimators (James and Stein, 1961) and the ridge regression estimators.[3]

Note, however, that MSE is of particular interest only when the loss function is assumed to be squared error. But this assumption is, in general, quite arbitrary.

In this paper stochastic dominance theory is used to analyze shrinkage estimation assuming only that the loss function is a continuous, non-decreasing function of $M(\varepsilon; \lambda)$, the generalized magnitude of the estimation error, ε:

$$M(\varepsilon; \lambda) \equiv \varepsilon \text{ for } \varepsilon > 0 \tag{3a}$$

$$\equiv -\lambda\,\varepsilon \text{ otherwise} \tag{3b}$$

[2] This follows from $\mathrm{MSE}\{\tilde{\beta}(k)\} = \beta^2\{(k/t)^2 + (k - 1)^2\}$.

[3] The Bayesian point estimator is not included in this list because it is not always a shrinkage estimator in the sense used here. With gaussian conditional distributions and symmetric, unimodal prior distributions, Andrews, et al. (1972) show that the posterior mean shrinks toward the prior mean, but the prior mean is not necessarily zero; also, the Bayesian estimator is not necessarily the posterior mean with other loss functions.

where $\varepsilon \equiv \tilde{\beta}(k) - \beta$. Thus, setting λ to one corresponds to restricting the analysis to symmetric loss functions.[4] The loss function need not be bounded; it is sufficient to assume that the expected loss of $\tilde{\beta}(k)$ is bounded. The class of loss functions satisfying these restrictions is denoted $L(\lambda)$ below.

Stochastic dominance is defined in the present context as follows:

DEFINITION. *Stochastic Dominance in Loss* (SDL) $\tilde{\beta}(k_a)$ SDL $\tilde{\beta}(k_b)$ *with respect to* $L(\lambda)$ *if and only if*

$$E \text{ loss } \{\tilde{\beta}(k_a) - \beta\} \leq E \text{ loss } \{\tilde{\beta}(k_b) - \beta\}$$

for all loss functions in $L(\lambda)$ *with strict inequality holding for at least one loss function in* $L(\lambda)$.

Stochastic dominance theory is reviewed in Section 2. There a theorem due to Tesfatsion (1976) is used to derive a necessary and sufficient condition under which $\tilde{\beta}(k_a)$ SDL $\tilde{\beta}(k_b)$ with respect to $L(\lambda)$. This condition is independent of $E(\hat{\beta})$ and Var $(\hat{\beta})$; it depends only on t, on the standardized distribution of $\hat{\beta}$ (denoted $\Psi\{\cdot\}$), and on the asymmetry parameter, λ.[5]

Given $\Psi\{\cdot\}$ and λ, this dominance condition is used to define and calculate three intervals for each value of t:

(1) the *inadmissible shrinkage interval*, or the set of all k such that there exists a k' for which $\tilde{\beta}(k')$ SDL $\tilde{\beta}(k)$ with respect to $L(\lambda)$,
(2) the *unbiasedness dominating interval*, or the set of all k such that $\tilde{\beta}(k)$ SDL $\hat{\beta}$ with respect to $L(\lambda)$,

and

(3) the *safe admissible shrinkage interval*, or the set of all k which are in the unbiasedness dominating interval but not in the inadmissible shrinkage interval.

Any shrinkage estimator in the inadmissible shrinkage interval is clearly unacceptable; another (typically more shrunken) estimator exists which dominates it over the loss function class. The inadmissible shrinkage interval excludes all $\tilde{\beta}(k)$ which are uniquely optimal for some particular loss function in $L(\lambda)$. There is no guarantee, however, that every $\tilde{\beta}(k)$ which is optimal for some particular loss function in the class is in the unbiasedness dominating interval—a shrinkage

[4] This particular form for $M(\varepsilon)$ is used below because it is the simplest possible asymmetric generalization of the ordinary absolute value function. A more complex quasiconvex piecewise linear form could be used instead; the loss function class L would then depend on a vector of parameters. Sales forecasting cum inventories provides an illustrative example. A sales overestimate leads to additional inventory holding costs; an underestimate smaller than the amount of inventory on hand leads to inventory replacement costs; a larger underestimate leads to stockout costs. Assuming constant marginal costs in all three cases leads to a generalized magnitude function analogous to equation (3), but characterized by two parameters.

[5] In fact, it is because this dominance condition depends explicitly on the value of λ that the loss function classes are defined so as to be indexed by it.

estimator that is optimal under one particular loss function could fail to dominate $\hat{\beta}$ over the entire class of loss functions. Nevertheless, if all one is certain of is that the loss function is in the specified loss function class, then the only way to ensure an expected loss which is no worse than that of $\hat{\beta}$ for all loss functions in the class and which is better than that of $\hat{\beta}$ for at least one loss function in the class is to restrict one's choice of shrinkage estimators to the unbiasedness dominating interval.

In the calculations reported below, the safe admissible shrinkage interval contains all of the k which are sufficiently small as to lie outside the inadmissible shrinkage interval, yet sufficiently large as to remain inside the unbiasedness dominating interval. The safe admissible shrinkage interval is in no sense optimal—optimality is a meaningful concept only in contexts where one can plausibly specify a particular loss function. But where all that one can be sure of about the loss function is that it is in $L(\lambda)$, it would be difficult to justify using a shrinkage estimator which is not in the safe admissible shrinkage interval.[6]

On the other hand, the safe admissible shrinkage interval is uninformative if it is too large. All three intervals are calculated in Section 3 for the particular case of gaussian estimators. In that case, the safe admissible shrinkage intervals are fairly small for $|t| > 2$. It is also comforting to find that the minimum MSE shrinkage estimator is inside these calculated safe admissible shrinkage intervals for $\lambda \leq 1$ and $|t| \geq 1$.

In ordinary usage, an estimator is termed "inadmissible" if it is dominated by some other estimator with respect to a single, given loss function but for all values of distributional parameters like t. Here the word "inadmissible" is used rather differently in the newly defined term, "inadmissible shrinkage interval." This interval contains the shrinkage estimators which are dominated by some other shrinkage estimator with respect to a very wide class of loss functions, but for a single, given value of t.

In practice, we know neither the loss function nor t. The great strength of the approach used here is that, while we rarely or never estimate the loss function, we routinely estimate t. In fact, the most commonly used tool for statistical inference in regression analysis—the ordinary t statistic, $\hat{t}$—is an estimator of t.

In Section 4 this estimator is used to make the analysis operational. There a shrinkage estimator which is a simple function of $\hat{t}$ is analyzed using simulation methods. This estimator is found to yield smaller average losses than the unbiased estimator over a wide variety of symmetric and asymmetric loss functions.

2. STOCHASTIC DOMINANCE THEORY AND SHRINKAGE ESTIMATION

Stochastic dominance theory has been developed and applied by a number of authors, including Blackwell (1951, 1953), Blackwell and Girschick (1954), Lehmann (1955), and Hardy et al. (1959) in the statistics literature; Kolm (1966, 1969),

[6] Since $\hat{\beta}$ cannot possibly dominate itself, the safe admissible shrinkage interval excludes it by construction. In practice $\hat{\beta}$ is dominated by shrunken estimators for most values of λ and t (and hence inadmissible) anyway, so this property is not inconvenient.

Rothschild and Stiglitz (1970), and Tesfatsion (1976) in the economics literature; and Hadar and Russell (1969) and Hanoch and Levy (1969) in the finance literature. Levy and Ben-Horim (1982) and Ben-Horim and Levy (1982, 1984) apply stochastic dominance theory to the comparison of the expected losses from two different estimators. Their analysis is restricted to loss functions which are symmetric and bounded, however, and they do not consider shrinkage estimation. The following theorem provides a useful necessary and sufficient condition for whether or not one shrinkage estimator stochastically dominates another. This condition is utilized below to calculate safe admissible shrinkage intervals.

SHRINKAGE ESTIMATION DOMINANCE THEOREM. $\{\tilde{\beta}(k_a)$ SDL $\tilde{\beta}(k_b)$ w.r.t. $L(\lambda)\}$ *if and only if*

$$\left[\begin{array}{c} \Omega(z, k_a) \le \Omega(z, k_b) \textit{ for all } z \le 0, \textit{ with strict} \\ \textit{inequality holding for at least one such } z \textit{ and} \\ \Omega(z, k) \equiv 1 - \Psi\{[(1-k)t - z]/k\} + \Psi\{[(1-k)t + z/\lambda]/k\} \end{array}\right]$$

where

(a) $\tilde{\beta}(k) \equiv k\hat{\beta}$, $\hat{\beta}$ *is an unbiased estimator of* β *with standardized, right continuous distribution* $\Psi\{\cdot\}$, *and given* $t \equiv \beta/\{\mathrm{Var}\,(\hat{\beta})\}^{.5}$,

(b) $L(\lambda)$ *is the class of all continuous, non-decreasing functions of* $M\{\tilde{\beta}(k) - \beta;\ \lambda\}$ *(defined in equation* (3)) *such that the expected loss from* $\tilde{\beta}(k)$ *is unbounded for at most one non-negative value of* k,

and

(c) $\{\tilde{\beta}(k_a)$ SDL $\tilde{\beta}(k_b)$ w.r.t. $L(\lambda)\}$ *means that the expected loss from* $\tilde{\beta}(k_a)$ *is less than or equal to that from* $\tilde{\beta}(k_b)$, *with strict inequality holding for at least one loss function in* $L(\lambda)$.

PROOF. Since $\hat{\beta}$ has mean β and variance $\sigma^2 = \beta^2/t^2$, $\tilde{\beta}(k)$'s estimation error (ε) has mean $(k - 1)\beta$ and variance $k^2\sigma^2$. Consequently, the standardized estimation error is

$$\{\varepsilon - (k-1)\beta\}/(k\sigma) = \{(1-k)t + (\varepsilon/\sigma)\}/k, \tag{4}$$

so that

$$1 - \Psi\{[(1-k)t + (-z/\sigma)]/k\} \tag{5}$$

is the probability that ε exceeds the positive value $-z$ and

$$\Psi\{[(1-k)t + (z/\lambda\sigma)]/k\} \tag{6}$$

is the probability that ε is less than the negative number z/λ. Thus,

$$1 - \Psi\{[(1-k)t + (-z/\sigma)]/k\} + \Psi\{[(1-k)t + (z/\lambda\sigma)]/k\} \tag{7}$$

is the probability that $M(\varepsilon;\ \lambda) \ge -z$ or, equivalently, that $-M(\varepsilon;\ \lambda) \le z$. Hence, $\Omega(z/\sigma, k)$ is the c.d.f. of $-M(\varepsilon;\ \lambda)$.

It then follows from Theorem 1* of Tesfatsion (1976) that $\tilde{\beta}(k_a)$ SDL $\tilde{\beta}(k_b)$ with respect to $L(\lambda)$ whenever $\Omega(z/\sigma, k_a)$ is less than or equal to $\Omega(z/\sigma, k_b)$ for all z/σ, with strict inequality holding for at least one value of z/σ. {Tesfatsion's theorem is stated in terms of the utility generated by the returns from a risky asset. Letting the "return" on the estimator be minus the generalized magnitude of the estimation error, ε, and hence

$$\text{Utility}\,\{-M(\tilde{\beta}(k) - \beta);\,\lambda)\} \equiv -\text{Loss}\,\{M(\tilde{\beta}(k)) - \beta;\,\lambda\}, \tag{8}$$

it follows that each loss function in $L(\lambda)$ corresponds to a utility function in what Tesfatsion calls $U^*(F, G)$.} Since $-M(\varepsilon; \lambda)$ is inherently non-positive, $\Omega(z/\sigma, k)$ is always one for positive z/σ regardless of k; consequently, only non-positive values of z/σ need to be examined. Since the condition $\Omega(z/\sigma, k_a) \leq \Omega(z/\sigma, k_b)$ must hold for all $z/\sigma \leq 0$ and since z and σ appear only as the quotient z/σ, there is no loss of generality in replacing z/σ by z.

This completes the proof of the theorem. Note that the dominance condition depends only on k_a, k_b, $\Psi\{\cdot\}$, t, and λ.[7]

3. SAFE ADMISSIBLE SHRINKAGE INTERVALS

The Shrinkage Estimation Dominance Theorem proven above specifies how to test for $\tilde{\beta}(k_a)$ SDL $\tilde{\beta}(k_b)$ with respect to $L(\lambda)$ given t, λ, and $\Psi\{\cdot\}$. This result makes feasible the construction of inadmissible shrinkage intervals, unbiasedness dominating intervals, and safe admissible shrinkage intervals. These intervals are calculated below as a function of t for the particular case of gaussian $\hat{\beta}$ (i.e. $\Psi\{\cdot\}$ is the unit normal distribution function). The theorem is used to explicitly determine whether or not $\tilde{\beta}(\mathrm{k_a})$ SDL $\tilde{\beta}(\mathrm{k_b})$ for each (k_a, k_b) pair over a grid of values ranging from .01 to 1.00 in increments of .01.[8] Dominance is checked for each pair by numerically comparing $\Omega(z, k_a)$ to $\Omega(z, k_b)$ over a grid of negative values of z; the density and extent of this grid are increased until the results stabilize.

Table 1 summarizes the results for the special case of symmetric loss functions—i.e. the case $\lambda = 1$. Note that $\hat{\beta}$ itself (i.e. $\tilde{\beta}(1)$} is always in the inadmissible shrinkage interval, along with an interval of shrinkage estimators which are insufficiently shrunk. Thus, since $\hat{\beta}$ is stochastically dominated, *some* degree of shrinkage is always called for. Note also that it is possible to shrink too much—the unbiasedness dominating intervals do not extend down to zero.

For significant $\hat{\beta}$'s (i.e. for $\hat{\beta}$ such that $|t| > 2$), the lengths of the safe admissible intervals are only .10 or less. Thus, for gaussian estimators and symmetric loss functions, one can say quite a bit about the amount of shrinkage which is reasonable.

[7] This theorem uses first degree stochastic dominance. A similar theorem can be proven, using second degree stochastic dominance, by further restricting the loss function class to convex functions of the generalized error magnitude. No doubt because $L(\lambda)$ is already quasiconvex in the estimation error itself (ε), initial calculations showed that this extension does not lead to interestingly different results.

[8] Values of k exceeding one are also examined. They are always outside the unbiasedness dominating interval.

TABLE 1
SHRINKAGE INTERVALS {GAUSSIAN $\hat{\beta}$ AND SYMMETRIC LOSS FUNCTIONS}

$\lvert t\rvert$[a]	k^b MSE	Inadmissible Interval	Unbiasedness Dominanting Interval	Safe Admissible Interval
0.50	.20	[.40, 1.00]	[.23, 1.00)	[.23, .39]
1.00	.50	[.63, 1.00]	[.44, 1.00)	[.44, .62]
1.50	.61	[.76, 1.00]	[.60, 1.00)	[.60, .75]
2.00	.80	[.84, 1.00]	[.71, 1.00)	[.71, .83]
2.50	.86	[.89, 1.00]	[.79, 1.00)	[.79, .88]
3.00	.90	[.92, 1.00]	[.84, 1.00)	[.84, .91]
4.00	.94	[.95, 1.00]	[.90, 1.00)	[.90, .94]
5.00	.96	[.99, 1.00]	[.93, 1.00)	[.93, .98]

[a] Due to the symmetry of both M(ε;1) and $\Psi\{\cdot\}$, the sign of t is irrelevant.
[b] This is the minimum MSE shrinkage estimator, $t^2/(1+t^2)$.

Table 2 extends these results to asymmetric loss function classes—i.e. $L(\lambda)$ with $\lambda \neq 1$. Here a remarkable pattern emerges: the shrinkage intervals with respect to $L(\lambda)$ with $\lambda < 1$, where the asymmetry favors shrinkage, are virtually identical to those with respect to $L(1)$. Indeed, where the value of t is not too high, the shrinkage intervals are very similar for $\lambda > 1$ as well.

Where λ and $|t|$ are both sufficiently large, however, the shrinkage estimators no longer dominate $\hat{\beta}$ so the safe admissible shrinkage interval disappears. This happens because, with $|t|$ large, $\hat{\beta}$ is so accurate that the variance drop from shrinkage is unable to compensate for any degree of bias in a direction which is penalized by loss functions with $\lambda > 1$.

It is also worth noting that, for $|t| \geq 1$ and $\lambda \leq 1$, the minimum MSE shrinkage estimator always lies inside the safe admissible shrinkage interval, toward the upper end. This result provides a rationale for minimum MSE estimation methods (James/Stein, ridge regression, etc.) which substantially transcends the squared error loss function assumption.

TABLE 2
SAFE ADMISSIBLE SHRINKAGE INTERVALS[a] {GAUSSIAN $\hat{\beta}$ AND ASYMMETRIC LOSS FUNCTIONS}

$\lvert t\rvert$[b]	k^c_{MSE}	$\lambda = .50$	$\lambda = .75$	$\lambda = .90$	$\lambda = 1.00$	$\lambda = 1/.90$	$\lambda = 1/.75$	$\lambda = 1/.50$
0.50	.20	[.22, .38]	[.23, .38]	[.23, .38]	[.23, .39]	[.23, .39]	[.24, .40]	[.26, .37]
1.00	.50	[.44, .61]	[.44, .61]	[.44, .61]	[.44, .62]	[.45, .62]	[.46, .65]	[.57, .77]
1.50	.61	[.59, .74]	[.60, .74]	[.60, .75]	[.60, .75]	[.61, .76]	[.64, .80]	[.93, .96]
2.00	.80	[.70, .82]	[.71, .82]	[.71, .82]	[.71, .83]	[.72, .84]	[.78, .89]	
2.50	.86	[.78, .87]	[.79, .87]	[.78, .87]	[.79, .88]	[.80, .89]	[.91, .95]	
3.00	.90	[.83, .90]	[.83, .90]	[.84, .91]	[.84, .91]	[.86, .93]	[.99, .99]	
4.00	.94	[.89, .94]	[.90, .94]	[.90, .94]	[.90, .94]	[.94, .97]		
5.00	.96	[.93, .96]	[.93, .96]	[.93, .96]	[.93, .98]	[.98, .99]		

[a] If the safe admissible shrinkage interval is given as [x,y], then the unbiasedness dominating interval is [x,1) and the inadmissible shrinkage interval is (y,1].
[b] Since M(ε;λ) is asymmetric for $\lambda \neq 1$, the sign of t is relevant. The safe admissible shrinkage interval with respect to L(λ) for -$\hat{\beta}$ (i.e. for $t < 0$) is equivalent to the interval with respect to L($1/\lambda$) for $\hat{\beta}$.
[c] This is the minimum MSE shrinkage estimator, $t^2/(1+t^2)$.

Finally, note that, for $\lambda \leq 1$, $\hat{\beta}$ itself is always dominated by shrunken, biased estimators. The implications of this result for rational expectations modelling are discussed in Section 5.

The practical application of these results is limited, however, by the fact that t is not known. This drawback is remedied in the next section.

4. IMMEDIATELY APPLICABLE RESULTS

In theory, it makes sense to require that a shrinkage estimator not only dominate the unbiased estimator (i.e. is "safe") but that it is also sufficiently shrunken that no other shrinkage estimator dominates it (i.e. it is "admissible"). As shown above, this can be done when t, the population t ratio, is known; the ensuing theoretical results yielded useful insights.

In practice, however, the population t ratio is not known. It can, of course, be estimated using $\hat{t}$, the usual sample t ratio, but this estimator has two major defects. First, although $\hat{t}$'s sampling distribution has a well known form (noncentral t), its density depends on a noncentrality parameter whose value is t itself and hence is unknown. This parameter can be approximated for numerical purposes by $\hat{t}$,[9] but this expedient is rendered less palatable by $\hat{t}$'s other major defect—it is quite noisy. Thus, one can calculate a $k^*(\hat{t})$ which approximately maximizes the probability (conditional on the observed $\hat{t}$) that the resulting shrinkage estimator $\{k^*(\hat{t})\hat{\beta}\}$ is in the safe admissible shrinkage interval with respect to $L(\lambda)$ for gaussian $\hat{\beta}$. However, the resulting estimator will typically have a larger expected loss than does $\hat{\beta}$ itself except when t is very small. This is because the additional sampling variation in $k^*(\hat{t})\hat{\beta}$ due to the sampling errors in $\hat{t}$ swamps the beneficial effect of the shrinkage.[10]

Nevertheless, it is still possible to obtain a shrinkage estimator which dominates the unbiased (gaussian) estimator with respect to $L(\lambda)$ over a substantial range of t values. For example, notice from Table 1 that $k = .95$ is in the unbiasedness dominating interval with respect to $L(1)$ over the entire range of t values considered. Reference to Table 2 shows that this is still the case for all values of $\lambda \leq 1$. In fact, a systematic calculation shows that $k = .95$ is in the unbiasedness dominating interval with respect to $L(1)$ for all $t \in [-6.085, 6.085]$.[11] A similar calculation shows that $k = .99$ dominates $\hat{\beta}$ with respect to $L(1)$ for all $t \in [-14.054, 14.054]$. (Table 3 gives the analogous t ratio intervals for additional values of k.) Thus, in many instances, a 5% shrinkage is clearly safe, while a 1% shrinkage would nearly always be safe—point estimates with t ratios in excess of fourteen are quite uncommon.

[9] Actually, one would multiply $\hat{t}$ by a factor, depending on the number of degrees of freedom, which makes the resulting estimate unbiased for t. (See Patel and Read 1982, p. 117 for details.)

[10] One can also calculate the minimum value of k (conditional on the observed value of t) such that the probability of being outside the unbiasedness dominating interval does not exceed some given value, p. However, the resulting k values depend too strongly on t for the resulting estimator, to work well; again, this is due to the sampling errors in t swamping the beneficial effect of the shrinkage.

[11] This calculation uses the Shrinkage Estimation Dominance Theorem of Section 2 to explicitly check $\tilde{\beta}(.95)$ SDL $\hat{\beta}$ over a grid of t values.

TABLE 3
UNBIASEDNESS DOMINATING T RATIO INTERVALS[a]

k	t ratio interval
.80	[−2.665, 2.665]
.85	[−3.225, 3.225]
.90	[−4.125, 4.125]
.95	[−6.085, 6.085]
.96	[−6.865, 6.865]
.97	[−7.984, 7.984]
.98	[−9.854, 9.854]
.99	[−14.054, 14.054]

[a] Each entry gives the interval of t ratio values for which a given value of the shrinkage factor k is in the unbiasedness dominating interval with respect to $L(1)$. Gaussianity is assumed.

The foregoing results raise the following two questions:

1. Does $k = .95$ or .99 provide a nontrivial improvement in expected losses over the appropriate range of t values?

and

2. Can one do better (in terms of expected loss reduction over a reasonable range of t values) by allowing k to depend weakly on $\hat{t}$?

These questions are addressed below using simulation methods.

For a given value of t, $\hat{\beta}$ is generated as

$$\hat{\beta} \sim N(1, 1/t^2)$$

and $\hat{t}$ is calculated from

$$\hat{t} = \hat{\beta}t/(x/30)^{.5}$$

where x is generated as $\chi^2(30)$ independently of $\hat{\beta}$. Thus, each value of $\hat{t}$ is a pick from the usual t distribution with 30 degrees of freedom. (Similar results were obtained using 60 degrees of freedom.)

Since $\hat{\beta}$ is unbiased by definition, the true value of β assumed here is one. The average squared values of $\hat{\beta} - 1$, $.99\hat{\beta} - 1$, and $.95\hat{\beta} - 1$ are then calculated over 40,000 realizations of $\hat{\beta}$. Table 4 gives the percentage MSE reduction (compared to the unbiased estimator, $\hat{\beta}$) provided by each of these two shrinkage estimators.[12]

Note that the 5% shrinkage yields a substantial (3% to 9%) MSE reduction so long as $t \leq 5$ and that the 1% shrinkage yields a 1% to 2% MSE reduction so long as $t \leq 10$. (Indeed, we know from the results in Table 3 that $k = .99$ dominates $\hat{\beta}$ with respect to *all* loss functions in $L(1)$ for all $t \in [-14.054, 14.054]$.) These improvements are admittedly not dramatic, but they are not negligible either and they are obtainable at essentially zero cost.

Table 4 also gives the MSE reduction results for a composite estimator which is a compromise between these two shrinkage estimators. This composite estimator is

[12] Of course, this is a relevant measure only for squared error loss functions; estimator performance with respect to other loss functions is explicitly considered later in this section.

TABLE 4
SHRINKAGE ESTIMATORS PERCENTAGE MSE REDUCTION[a] (λ = 1.00)

t ratio	$.95\hat{\beta}$	Composite[b]	$.99\hat{\beta}$
1.0	9.5	6.0	2.0
2.0	8.9	3.8	2.0
2.5	8.1	2.6	1.9
3.0	7.5	1.9	1.9
4.0	5.4	1.5	1.8
5.0	3.4	2.3	1.7
6.0	0.8	2.3	1.7
7.0	−2	1.8	1.5
8.0	−7	1.8	1.2
9.0	−10	1	1.3
10.0	−15	0	1

[a] Figures quoted to .1 are accurate to ±.1; figures quoted as integers are accurate to ±.5.
[b] This estimator is $k\hat{\beta}$ where $k = .96$ for $|\hat{t}| \leq 1$ and $k = \min\{.96 + (.02/3)(|\hat{t}| - 1), .98\}$ for $|\hat{t}| > 1$.

$k\hat{\beta}$ with $k = .96$ for $|\hat{t}| \leq 1$ and $k = \min\{.96 + (.02/3)(|\hat{t}| - 1), .98\}$ for $|\hat{t}| > 1$. This particular compromise is the product of a good deal of experimentation. Estimators which depend more heavily on $\hat{t}$ than does this estimator perform less well because of the additional sampling error introduced by that heavier $\hat{t}$ dependence. Estimators which depend still less on $\hat{t}$ perform well only over a smaller range of t values. As the results quoted in Table 4 show, this composite shrinkage estimator works noticeably better than $k = .99$ for $t \leq 2.5$ and substantially better than $k = .95$ for $t \geq 6$. While $k = .99$ is clearly preferable to the composite estimator for $t \geq 9$, it must be recognized that the expected loss improvements available from shrinkage are minor for estimators this accurate anyway.[13]

Table 5 examines the performance of the composite estimator for a number of other symmetric loss functions. These results indicate that this shrinkage estimator is a noticeable, if not always substantial, improvement over the unbiased estimator over a wide variety of loss functions in $L(1)$.

Finally, Table 6 examines the performance of the composite shrinkage estimator over loss functions with different values of the asymmetry parameter, λ. (This parameter is defined in equation 3.) As one might expect, the gains to shrinkage are enhanced when the asymmetry itself favors shrinkage toward zero ($\lambda < 1$ for $\beta > 0$) and are diminished otherwise. It is interesting to note, however, that the composite shrinkage estimator is still an improvement over the unbiased estimator for $\lambda = 1/.90$ when $t \leq 6$ and is still an improvement for $\lambda = 1/.75$ for $t \leq 2$.

This composite shrinkage estimator is not in itself useful for interval estimation or for statistical inference purposes since its sampling distribution is unknown. But that defect is irrelevant here—this estimator's raison d'etre is point estimation, where what is wanted is the best single estimate for β. In any case, $\hat{\beta}$ is still available for those other purposes. For example, if a measure of the estimation

[13] In particular, note that the seemingly large percentage increases in MSE from the 5% and the composite shrinkage estimators in Table 4 for large values of t actually represent modest *MSE* increases which are being divided by a small base value.

TABLE 5
COMPOSITE SHRINKAGE ESTIMATOR[a]
PERCENTAGE EXPECTED LOSS REDUCTION FOR VARIOUS SYMMETRIC LOSS FUNCTIONS[b]

t	M	M^2	M^4	log $(1+M)$	log $(1+M^2)$	log $(1+M^4)$	e^M	e^{M^2}	e^{M^4}
1.0	3.2	6.0	10.9	2.3	4.1	6.0	0.57	1.10	2.50
2.0	1.8	3.8	8.0	1.4	3.0	5.8	0.20	0.28	0.38
2.5	1.1	2.6	6.1	0.9	2.1	4.8	0.11	0.14	0.15
3.0	0.7	1.9	5.0	0.5	1.5	4.2	0.05	0.07	0.06
4.0	0.7	1.5	3.5	0.6	1.4	3.3	0.05	0.04	0.02
5.0	1.2	2.3	4.2	1.1	2.2	4.1	0.07	0.04	0.01
6.0	1.1	2.3	4.2	1.0	2.2	4.2	0.05	0.03	0.00
7.0	0.9	1.8	3.7	0.8	1.8	3.7	0.04	0.02	0.00
8.0	0.9	1.8	3.9	0.8	1.7	3.8	0.03	0.01	0.00
9.0	0.5	1.0	2.2	0.4	1.0	2.2	0.02	0.01	0.00
10.0	0.1	0.2	0.1	0.1	0.2	0.1	0.00	0.00	0.00

[a] This estimator is $k\hat{\beta}$ where $k = .96$ for $|\hat{t}| \leq 1$ and $k = \min\{.96 + (.02/3)(|\hat{t}| - 1), .98\}$ for $|\hat{t}| > 1$.

[b] M is the generalized magnitude function defined in equation (3). Figures are accurate to ± 1 in last digit quoted for $t \leq 4$, ± 2 for $t \leq 8$, and ± 3 for $t > 8$.

uncertainty is desired, one could still quote a 95% confidence interval for β alongside the composite point estimate.

In summary, shrinkage is most effective for estimators with moderate or low t values. Estimators with very high t ratios are already so accurate that the shrinkage does not lower the variance enough to compensate for any degree of bias. Where $t > 9$ is not a realistic possibility—and this, of course, includes most estimators found in applied econometric work—the composite shrinkage estimator analyzed here provides a modest (but non-negligible) expected loss improvement (compared

TABLE 6
COMPOSITE SHRINKAGE ESTIMATOR[a]
PERCENTAGE EXPECTED LOSS REDUCTION FOR VARIOUS ASYMMETRIC LOSS FUNCTIONS[b]

t	$\lambda = .50$	$\lambda = .75$	$\lambda = .90$	$\lambda = 1.00$	$\lambda = 1/.90$	$\lambda = 1/.75$	$\lambda = 1/.50$
1.0	8.7	7.2	6.5	6.0	5.5	4.7	3.3
2.0	8.8	6.1	4.6	3.8	2.9	1.4	−1.4
2.5	8.8	5.5	3.7	2.6	1.6	−.2	−3.4
3.0	8.9	5.2	3.1	1.9	0.6	−1.4	−5.1
4.0	10.2	5.6	3.0	1.5	0.1	−2.4	−7.0
5.0	12.1	6.9	4.1	2.3	0.6	−2.3	−7.5
6.0	13.7	7.6	4.3	2.3	0.3	−3.0	−9.1
7.0	15.1	8.0	4.2	1.8	−0.5	−4.3	−11.3
8.0	16.7	8.8	4.4	1.8	−0.9	−5.3	−13.5
9.0	17.8	8.9	4.0	1.0	−2.0	−7.0	−16.1
10.0	19.1	9.0	3.5	0.2	−3.1	−8.7	−18.8

[a] This estimator is $k\,\hat{\beta}$ where $k = .96$ for $|\hat{t}| \leq 1$ and $k = \min\{.96 + (.02/3)(|\hat{t}| - 1), .98\}$ for $|\hat{t}| > 1$.

[b] The loss functions considered in this Table are all of the form M^2. The generalized magnitude function (M) and the asymmetry parameter (λ) are defined in equation 3. Accuracy is as in Table 5.

to the unbiased estimator) over a wide range of loss functions, symmetric and asymmetric.

5. CONCLUSIONS

Three kinds of conclusions emerge from these results:

(1) Since the minimum MSE shrinkage factor

$$k_{\text{MSE}}(t) = t^2/(1 + t^2)$$

is in the safe admissible shrinkage interval with respect to $L(\lambda)$ for all gaussian estimators with $t \geq 1$ and $\lambda \leq 1$, minimum MSE methods can be justified from a perspective which dramatically transcends the squared error loss function.

and

(2) It is feasible—in fact, easy—to use the results of Section 4 to shrink unbiased gaussian parameter estimates enough to ensure that the resulting estimator dominates the unbiased estimator with respect to a wide class of symmetric and asymmetric loss functions over a substantial range of t values.

and

(3) The unbiased gaussian estimator is always in the inadmissible shrinkage interval with respect to $L(1)$, the class of (symmetric) loss functions which are continuous, nondecreasing functions of the error magnitude. This is true for many asymmetric loss function classes as well. Thus, unbiased gaussian parameter estimates (and point forecasts) are typically dominated by biased estimators which are shrunk to some degree toward zero.

This last result implies that the unbiased expectations generating mechanism characteristic of many modern macroeconomic models can only be optimal (i.e. rational) in the limit where the sampling variation in the estimated coefficients of the model available to the agents is negligible. We live in a world of distinctly finite samples due, in part, to the fact that most structural coefficients are clearly unstable over long time periods. Consequently, sampling variation in estimated coefficients is and will remain a fact of life.

The results developed above provide explicit guidance (not tied to an arbitrary squared error loss function assumption) as to how much shrinkage is reasonable when the relevant random variable is gaussian, as might well be the case for a linear neoclassical macroeconomic model.[14] These results thus make it possible to investigate the sensitivity of the policy conclusions from such a model to reason-

[14] Similar shrinkage results could be obtained using the techniques developed above with a (given) non-gaussian distribution, so long as it is symmetric.

able amounts of bias in the agents' expectations.[15]

REFERENCES

ANDREWS, R. L., J. C. ARNOLD AND R. G. KRUTCHKOFF, "Shrinkage of the Posterior Mean in the Normal Case," *Biometrika* 59 (1972), 693–694.

BEN-HORIM, M. AND H. LEVY, "Evaluating Estimators Using Stochastic Dominance Rules: The Variance of a Normal Distribution," *Communication in Statistics* 11 (1982), 1071–1076.

——— AND ———, "Stochastic Dominance and Parameter Estimation: The Case of Symmetric Stable Distributions," *Insurance, Mathematics and Economics* 3 (1984), 133–8.

BLACKWELL, D., "Comparison of Experiments," In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability* (Berkeley: University of California Press, 1951), 93–102.

———, "Equivalent Comparisons of Experiments," *Annals of Mathematical Statistics* 24 (1953), 265–72.

——— AND GIRSCHICK, M. A., *Theory of Games And Statistical Decisions* (New York: Wiley, 1954).

HADAR, J. AND W. R. RUSSELL, "Rules For Ordering Uncertain Prospects," *American Economic Review* 54 (1969), 25–34.

HANOCH, G. AND H. LEVY, "The Efficiency Analysis of Choices Involving Risk," *Review of Economic Studies* 36 (1969), 335–46.

HARDY, G. H., J. E. LITTLEWOOD AND G. POLYA, *Inequalities* (Cambridge: Cambridge University Press, 1959).

JAMES, W. AND C. STEIN, "Estimation With Quadratic Loss," In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* 1 (Berkeley: University of California Press, 1961).

KOLM, S. C., *Les Choix Financiers et Monetaires* (Paris: Dunod, 1966).

———, "The Optimal Production of Social Justice," In Margolis, J. and Guitton, H., eds., *Public Economics* (New York: Macmillan, 1969).

LEHMANN, E. L., "Ordered Families of Distributions," *Annals of Mathematical Statistics* 26 (1955), 399–419.

LEVY, H. AND M. BEN-HORIM, "Evaluating Estimators Using Stochastic Dominance Rules," Working Paper No. #83-12, School of Business, The Hebrew University, Jerusalem, Israel, 1982.

PATEL, J. K. AND C. B. READ, *Handbook of the Normal Distribution* (New York: Marcel Dekker, Inc., 1982).

ROTHSCHILD, M. AND J. E. STIGLITZ, "Increasing Risk: 1. A Definition," *Journal of Economic Theory* 2 (1970), 225–243.

STEGMAN, T., "On the Rationality of the Rational Expectations Hypothesis," *Australian Economic Papers* 24 (1985), 350–355.

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

[15] The distribution of the relevant random variables would doubtless be asymmetric as well as non-gaussian for a highly nonlinear model. In that case it is already well known (Stegman 1985 and others) that the optimal expectation is in general biased. The techniques developed above could still in many cases yield useful information as to how much bias is reasonable, although this bias might no longer represent shrinkage toward zero.