

HYPOTHESIS TESTING USING NUMEROUS APPROXIMATING FUNCTIONAL FORMS

Bailey Norwood

Jayson Lusk

Peyton Ferrier

*Paper presented at the NCR-134 Conference on Applied Commodity Price Analysis,
Forecasting, and Market Risk Management
St. Louis, Missouri, April 23-24, 2001*

Copyright 2001 by Bailey Norwood, Jayson Lusk and Peyton Ferrier. All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.

* Doctoral candidate (fbnorwoo@unity.ncsu.edu), Department of Agricultural and Resource Economics, North Carolina State University, assistant professor (JLusk@agecon.msstate.edu), Department of Agricultural Economics, Mississippi State University, and graduate research assistant (pmferrie@unity.ncsu.edu), Department of Agricultural and Resource Economics, North Carolina State University. Helpful comments from Cliff Mansfield and Mike Cantore are gratefully acknowledged.

HYPOTHESIS TESTING USING NUMEROUS APPROXIMATING FUNCTIONAL FORMS

While the combination of several or more models is often found to improve forecasts (Brandt and Bessler, Min and Zellner, Norwood and Schroeder), hypothesis tests are typically conducted using a single model approach¹. Hypothesis tests and forecasts have similar goals; they seek to define a range over which a parameter should lie within a degree of confidence. If it is true that, on average, composite forecasts are more accurate than a single model's forecast, it might also be true that hypothesis tests using information from numerous models are, on average, more accurate in the sense of lower Type I and Type II errors than hypothesis tests using a single model.

Researchers often employ J-Tests to identify the best functional form for the mean of a variable. The J-Test requires testing the null hypothesis that Model A is correct versus the alternative that Model B is correct. Data employed in such tests often prefer a combination of Model A and Model B though, thereby rejecting both models (Anderson et al. and McAleer et al.) In such cases, hypothesis tests may be best conducted using a combination of the two models, rather than just one.

This paper is an evaluation of hypothesis tests from numerous versus a single model. In the construction of hypothesis tests, seldom is it mentioned that part of the null hypothesis is that the assumed model is true. Most economic data are non-experimental though and the true functional form governing the economic process, if there is a true functional form, is unknown. Unfortunately, estimates are often sensitive to the choice of functional form and the science of model identification is in its early stages². For any given data there exists several or more appropriate functional forms with potentially different implications. Shumway and Lim provide excellent examples of how elasticities vary under alternative functional forms. They summarize the robustness problem by stating (page 275) "Attempting to narrowly bound estimates of output supply and input demand elasticity for a given category remains an exceedingly difficult task. Even using the same data, holding the point of evaluation constant, and using alternative functional forms with the same number of free parameters to be estimated, the implied elasticities can vary widely."

Though econometricians have not reached a consensus on the best method of model selection, many claim the best models are those that perform well out-of-sample. This is because in-sample model performance criteria are often arbitrary and unreliable. In-sample fit can be manipulated by the addition of parameters. Measure like the Akaike Information Criteria and adjusted coefficient-of-determination are constructed to correct for this manipulation, but are still unpopular. Hypothesis tests are sometimes useful, but at other times are sensitive to the type of test conducted and not all models are nested. Non-nested tests are available but sometimes yield ambiguous conclusions (Anderson et al. and McAleer et al.) The likelihood dominance procedure provides an unambiguous model ranking, but requires all models to have an identical number of variables (Anderson et al). It is taken as given in this paper that a model is best judged by its out-of-sample performance.

However, the fact that Model A performs outperforms Model B out-of-sample does not imply that only Model A should be incorporated in the construction of hypothesis tests. The forecasting literature often finds incorporating the results of two or more models improves accuracy (Brandt and Bessler, Min and Zellner, and Norwood and Schroeder). In a summary of how ERS conducts retail price forecasts, the forecasters state (Page 15) “It appears that the competing forecasts can be profitably combined to yield a composite forecast which is superior to each of the individual forecasts.” If this holds true for forecasts, it may hold true for hypothesis tests as well. The purpose of this paper is to evaluate a method of hypothesis testing which employs numerous models, none of which are assumed true, but each a potentially appropriate approximation. Each model may contribute information regarding the true value of a statistic, where each model’s contribution depends on its out-of-sample performance.

This analysis assumes the true process governing economic variables is complex and unknown to the researcher. It must be approximated by one or more simpler forms. Hence, the usual properties regarding hypothesis tests do not apply. For instance, if a 90% confidence interval is constructed for repeated samples, it will not necessarily contain the true value 90% of the time because models are likely specified with error. This has been noted in the literature as x% confidence intervals often do not contain the true value x% of the time out-of-sample (Makridakis et al. and Norwood and Schroeder). Hence, two often ignored issues are focused upon 1) the fact that hypothesis tests are conducted using approximate, as opposed to true, functional forms and 2) numerous models may provide better tests than a single model.

THE WEIGHTED STATISTIC APPROACH

If there were a true model governing economic processes, it is likely the true form is more complex than the models researcher typically specify. Additionally, any economic variable is likely influenced by a myriad of variables. Consequently, econometricians cannot precisely measure all relevant variables, and even if they could, the number of variables would likely outnumber the observations of most datasets. Hence, any specified functional form will likely be a biased estimator even if it appears unbiased in-sample³. Without knowing the true functional form, and without knowing if a form is unbiased, determining the most efficient method of estimation becomes increasingly difficult. Nevertheless, this is the state of economic data and must be considered as part of one’s analysis.

Suppose a researcher is interested in the value of an elasticity, denoted η . Let this elasticity be the percent change of variable y with respect to a one percent change in variable x_1 ⁴. Also, let four other variables, x_2 , x_3 , x_4 , and x_5 influence y . The researcher wants to estimate η observed at the mean of $X = [x_1 \ x_2 \ x_3 \ x_4 \ x_5]'$. Assume this researcher has K available functional forms mapping X into y . Denote the i^{th} model as $f_i(X, \epsilon)$ where ϵ is a stochastic error and its corresponding elasticity estimate as \hat{h}_i . The researcher has the option of either estimating all K models and choosing one \hat{h}_i or combining each model’s elasticity estimate into a single elasticity estimate; a composite estimate. One

method of obtaining a composite estimate is a simple average; $\hat{\mathbf{h}} = 1/K \sum_{i=1}^K \hat{\mathbf{h}}_i$. Brandt and Bessler found when forecasting hog prices that a weighted average performed better, where each model's weight increases as it performed better relative to the other models.

Denote a weighted average elasticity estimate as $\hat{\mathbf{h}} = \sum_{i=1}^K w_i \hat{\mathbf{h}}_i$ where $\sum_{i=1}^K w_i = 1$.

The difficult part of constructing a composite estimate is determining the weights. The weights should have some objective in mind. Suppose the objective is to minimize the mean-squared error of $\hat{\mathbf{h}} - \mathbf{h}$. The mean-squared error can be decomposed into its variance and bias. Suppose a researcher was considering combining K estimators into a single composite estimator in a manner which minimizes mean-squared error, i.e., minimizes

$$(1) \text{MSE} \left(\hat{\mathbf{h}} = \sum_{i=1}^K w_i \hat{\mathbf{h}}_i \right) = \sum_{i=1}^K w_i^2 \mathbf{s}_{ii}^2 + \sum_{i \neq j} \sum_{j \neq i} w_i w_j \mathbf{r}_{ij} \mathbf{s}_i \mathbf{s}_j + \left\{ \sum_{i=1}^K w_i B_i \right\}^2$$

where \mathbf{s}_{ij}^2 is the covariance between $\hat{\mathbf{h}}_i$ and $\hat{\mathbf{h}}_j$ and B_i is the bias of $\hat{\mathbf{h}}_i$ and is equal to $E(\hat{\mathbf{h}}_i) - \mathbf{h}$. Brandt and Bessler consider weighting two unbiased estimates. By setting $B_i = 0$ for all i, taking the derivative of (1) with respect to w_i , imposing the constraint that $w_2 = 1 - w_1$, and setting this relationship equal to zero yields

$$(2) w_1 = \frac{\mathbf{s}_{22}^2 - \mathbf{r}_{12} \mathbf{s}_1 \mathbf{s}_2}{\mathbf{s}_{11}^2 + \mathbf{s}_{22}^2 - 2 \mathbf{r}_{12} \mathbf{s}_1 \mathbf{s}_2}.$$

For this to be the solution of a minimum variance, the term

$$(3) \mathbf{s}_{11}^2 + \mathbf{s}_{22}^2 - 2 \mathbf{r}_{12} \mathbf{s}_1 \mathbf{s}_2 > 0$$

must hold, and it will if the correlation between the estimators is equal to or less than zero. If the correlation is positive, each σ_i must be estimated to determine if (3) holds. Note the intuition behind (2). As the variance of η_2 increases more weight is given to η_1 . If the variance of η_1 and η_2 are equal and their correlation is one, then the mean-squared error is the same regardless of the assigned weights, and they should be, as they would be the exact same estimate.

Other weighting schemes have been developed within the Bayesian framework. Such weights are assigned almost exclusively to two models due to the mathematical complexities of calculating posterior distributions. Min and Zellner consider two competing models, $f_1(X, \epsilon)$ and $f_2(X, \epsilon)$ and define the expected value of the dependent variable y as: $E\{y\} = P_1 E\{f_1(X, \epsilon)\} + P_2 E\{f_2(X, \epsilon)\}$ where P_1/P_2 is the odds ratio of each model being correct. The prior for this odds ratio is set to one (an uninformative prior)

and can then be updated based on data to obtain a posterior distribution. Overall, this weighting scheme did not provide much, if any, improvement in forecasts. Regardless, this scheme either improved the performance of out-of-sample predictions or did not affect it and therefore is a useful tool. Though sound fundamentally, this technique is difficult to extend to numerous and different classes of models.

Dorfman considers the case of forecasting the sign of the change in y . He uses three models; a reduced form econometric model, a state-space model, and expert forecasts. Let $z_{it} = 1$ if Model i forecasts $y_t - y_{t-1}$ to be greater than zero and $z_{it} = 0$ otherwise. Based on past predictions, one may use a logit model to predict the probability z_{it} predicts the correct change. Denote this probability as P_{it} . Dorfman then normalizes each P_{it} as $p_{it} = P_{it} / \sum_{j=1}^3 P_{jt}$. Then, denote the loss incurred from predicting $z_{it} = 1$ when $y_t - y_{t-1} < 0$, relative to predicting $z_{it} = 0$ when $y_t - y_{t-1} > 0$ as L . The composite forecast is then

$$(4) \quad \hat{z}_t = \begin{cases} 1 & \text{if } \sum_{i=1}^3 p_{it} z_{it} > L \\ 0 & \text{if } \sum_{i=1}^3 p_{it} z_{it} < L \end{cases} .$$

Dorfman found the composite model outperformed each of the three individual models.

This study seeks a simple method of combining models that allows hypothesis testing extendable to any number and type of models. This study uses the mean-squared error criterion, as Brandt and Bessler, but allows more than 2 models. Keeping the statistic of interest as an elasticity of y with respect to x_1 , denoted η , the objective is to calculate a set of weights, w_i , which minimize

$$(5) \quad MSE\left(\hat{\mathbf{h}} = \sum_{i=1}^K w_i \mathbf{h}_i\right) = \sum_{i=1}^K w_i^2 \sigma_{ii}^2 + \sum_{i \neq j} \sum_{j \neq i} w_i w_j \rho_{ij} \sigma_i \sigma_j + \left\{ \sum_{i=1}^K w_i B_i \right\}^2 + I \left(1 - \sum_{i=1}^K w_i \right)$$

subject to

$$0 \leq w_i \leq 1$$

The terms σ_{ij}^2 and ρ_{ij} for any two models can be calculated analytically for simple forms or numerically through bootstrapping. The terms B_i are unidentifiable though as the researcher does not know the true elasticity. Suppose, instead of choosing the w_i 's to minimize the mean-squared error of $\hat{\mathbf{h}} = \sum_{i=1}^K w_i \mathbf{h}_i$, the weights were chosen to minimize the mean-squared error of the dependent variable y over a set of out-of-sample forecasts. Let $\hat{y}_i = f_i(X, \mathbf{e})$ be the prediction of y from Model i . Previously, the objective was to minimize the mean-squared error of a constant η . The fact that η was a constant allowed decomposing the mean-squared error into variance and bias components. Now, the objective is to minimize the mean-squared error of the variable y which changes at each

observation, so the decomposition does not follow. The mean-squared error for a set of out-of-sample forecasts is then stated as

$$(6) \sum_{t=1}^T \left\{ \frac{\left(\sum_{i=1}^K w_i \hat{y}_{i,t} - y_t \right)^2}{T} \right\} + I \left(1 - \sum_{i=1}^K w_i \right) \text{ s.t. } 0 \leq w_i \leq 1 \forall i$$

which is minimized by a constrained OLS regression⁵. The subscript t denotes an out-of-sample forecast.

A word of caution: The mean-squared error may be set to zero arbitrarily by incorporating a number of models equal to the number of out-of-sample forecasts. Note the solution may not be interior, especially if individual model's forecasts are positively correlated. Once the weights are chosen they may be employed in hypothesis tests. Suppose a researcher wants to test whether the elasticity of y with respect to x_i (denoted η as before) observed at the mean vector of $X = [x_1, x_2, x_3, x_4, x_5]$, is greater than or equal to a value γ .

The elasticity of interest is

$$(7) \frac{E\{y | \bar{x}_1(1.01), \dots, \bar{x}_5\} - E\{y | \bar{x}_1, \dots, \bar{x}_5\}}{E\{y | \bar{x}_1, \dots, \bar{x}_5\}} \Big/ .01 \quad \text{where } \bar{x}_i \text{ is the sample mean of } x_i.$$

Equation (7) suggests the elasticity is only calculated at the mean of x_1 . Actually, in simulations it was calculated and hypothesis tests were conducted for this elasticity at the sample mean and four other values of x_1 . All other variables are held at their means for elasticity calculations though. The composite estimate of $E\{y | \bar{x}_1, \dots, \bar{x}_5\}$ is

$\sum_{i=1}^K w_i \hat{y}_i(\bar{x}_1, \dots, \bar{x}_5) - B$ where B is the estimated bias. This bias may not be a constant.

$$(8) E\left\{ \sum_{i=1}^K w_i \hat{y}_i(\bar{x}_1, \dots, \bar{x}_5) - B \right\} = E\left\{ \sum_{i=1}^K w_i \hat{y}_i(\bar{x}_1, \dots, \bar{x}_5) \right\} - \left\{ E\left\{ \sum_{i=1}^K w_i \hat{y}_i(\bar{x}_1, \dots, \bar{x}_5) \right\} - y \right\} = y$$

After calculating the weights, if one measured the bias by

$$(9) B = \sum_{t=1}^T \frac{\sum_{i=1}^K \{w_i \hat{y}_{i,t} - y_t\}}{T} \quad \text{the estimated bias will be zero by OLS construction. This}$$

bias is then assumed zero. An estimate of η is then

$$(10) \hat{\eta} = \left\{ \frac{\sum_{i=1}^K w_i \hat{y}_i(\bar{x}_1(1.01), \dots, \bar{x}_5) - \sum_{i=1}^K w_i \hat{y}_i(\bar{x}_1, \dots, \bar{x}_5)}{\sum_{i=1}^K w_i \hat{y}_i(\bar{x}_1, \dots, \bar{x}_5)} \right\} \Big/ .01.$$

The probability density function of this statistic may be obtained by bootstrapping. The strongest assumption made in this analysis is that the bias is zero, because the zero estimated bias is just a construction of OLS estimation. However, it seems plausible that the bias using the weighted average is less than the bias from any one individual model, and the assumption is no stronger than assuming one specified model is correct/unbiased. While the bias can be estimated using out-of-sample observations, this was not attempted in the present analysis. The approach of using numerous models to obtain a probability density function for a statistic in the fashion described above is referred to as the Weighted Statistic Approach (WSA).

The WSA has several appealing features. Results will be robust since hypothesis tests using the WSA are based on numerous models. This decreases the probability that results of a statistical test are due solely to the functional form chosen. Researchers often try to avoid this fallacy by estimating several functional forms and reporting each form's estimates and test results. Space considerations only allow a limited number of estimations to be reported though, and it is often difficult for readers to interpret numerous estimations. The WSA can take numerous models and report then in a single point estimate, probability density function, and hypothesis test.

Kastens and Brester revived Wold's claim that "Forecasting is to nonexperimental model building as replications are to controlled experiments." Estimating a different model is like conducting an experiment, and science is the business of performing numerous experiments and reporting the results in a succinct and informative fashion. In the physical sciences, many experiments are performed before results are confirmed. Useful experiments are included in the results and uninformative experiments are not. Econometricians, however, often base results on one or a few models when the cost of estimating more models is small.

The WSA has a formal method of determining which models are informative and which are not by assigning positive and zero values for each weight. Models are given greater weight they better they explain movements in the dependent variable out-of-sample. Models which better explain a variable y as a function of variables in X seem more likely to explain functions of y using X , such as elasticities. If a composite forecast is superior to a single model's forecast, then a subset of weights will be less than one. Therefore, by extending weights assigned to each individual model to other statistics like elasticities, the weighted statistic may very well be superior to any single statistic.

While the WSA has some intuitive merits, whether it is indeed superior to a single model approach is an empirical question—a difficult empirical question. If one does not know the true functional form mapping X into y there is no way to prove analytically whether the WSA is appropriate. Therefore, simulations are performed. If the WSA is found to be superior, subsequent research may formalize the method and improve estimates. If it is not, then hypothesis tests are simpler and can be conducted using the one superior model.

SIMULATION DESCRIPTION

A series of simulations is performed to determine how the WSA performs compared to a single model under a setting where the true functional form is unknown. This simulation is briefly described here but is given more detail in Appendix A. The true functional form of a dependent variable is assumed to be a function of five independent variables; x_{1n} , x_{2n} , x_{3n} , x_{4n} , and x_{5n} where n denotes an observation. If there is a true functional form for an economic variable, it is probably complex. Therefore, the true functional form in the simulation is chosen randomly from a set of complex models. The true form will differ for each simulation, making results applicable to a wider array of models.

Let $X = [I_T, x_1, \dots, x_5, x_1^2, \dots, x_5^2, x_1x_2, \dots, x_1x_5, x_2x_3, \dots, x_4x_5]$ where x_i is a vector of x_{in} for $n = 1$ to N and I_N is a vector of ones with N columns. Then, let $X(1:j)$ be a matrix equal to columns 1 through j of X . The parameter vector mapping $X(1:j)$ into the expected value of y is $\beta(1:j)$. Let $\beta = [\beta_0 \beta_1 \beta_2 \beta_3 \beta_4 \beta_5 \beta_{12} \beta_{22} \beta_{32} \beta_{42} \beta_{52} \beta_{i_{12}} \beta_{i_{13}} \beta_{i_{14}} \beta_{i_{15}} \beta_{i_{23}} \beta_{i_{24}} \beta_{i_{25}} \beta_{i_{34}} \beta_{i_{35}} \beta_{i_{45}}]'$ where $\beta_{i_{25}}$ denotes the parameter corresponding to the interaction term x_1x_5 . The vector $\beta(1:j)$ then contains the first j rows of β . By letting j vary across simulations, a wide array of true models is simulated.

The true model is allowed to become even more complex by allowing a Box-Cox transformation on each x_i . Each x_i is transformed to $x_i^{(\lambda_i)}$ where $x_i^{(\lambda_i)} = (x_i^{\lambda_i} - 1)/\lambda_i$ and the value of λ_i differs for each x_i within a simulation and across simulations. Redefine the mean equation of y then as $E\{y\} = X(1:j, \lambda)\beta(1:j)$ where λ is a vector of λ_i 's. The values of j , X , λ , and β are chosen randomly from distributions defined in Appendix A.

The value of y is defined as $y = X(1:j, \lambda)\beta(1:j) + \varepsilon$ where ε is a normally distributed term with a zero mean. The variance of ε is allowed to vary randomly across simulations. Define $mX(1:j, \lambda)$ as the mean vector of $X(1:j, \lambda)$. The variance of ε is then defined as $(gmX(1:j, \lambda)\beta(1:j))^2$ where g is a uniformly distributed variable between .05 and .5. Values of y are then simulated to form a set of data on y and X .

It is assumed the researcher has no idea what the true functional form is, or is even aware a separate Box-Cox transformation is made to each variable. The simulated researcher proceeds by estimating the elasticity of y with respect to x_1 , evaluated at the mean of the untransformed variables in X . The researcher uses five parsimonious models where the Box-Cox transformation values are .01, .25, .5, .75, and 1 and five flexible functional forms with the same transformation values. The parsimonious models are

$$(11) \quad E\{y\} = \mathbf{b}_o + \sum_{i=1}^5 x_i^{(\lambda)} \hat{\mathbf{b}}_i$$

where λ is the same for all variables within a model, and vary from Model 1 to Model 5 by $\lambda = .01, .25, .5, .75, \text{ and } 1$. The flexible functional forms take the form

$$(12) \quad E\{y\} = \mathbf{b}_o + \sum_{i=1}^5 x_i^{(I)} \hat{\mathbf{b}}_i + \sum_{i=1}^5 \sum_{j=1}^5 x_i^{(I)} x_j^{(I)} \hat{\mathbf{b}}_{ij}$$

where, again, λ is the same for all independent variables and equal to 0, .25, .5, .75, and 1 in Models 6, 7, 8, 9, and 10, respectively. Models 6 through 10 are similar to the flexible functional forms used frequently except that the dependent variable is not transformed⁶.

During each simulation all 10 models are estimated. The parameters from these estimations are then employed in a set of out-of-sample forecasts. The weights assigned to each model are then calculated by minimizing the out-of-sample squared errors as described in (6). These weights are treated as constants. The in- and out-of-sample observations are then combined for a series of bootstraps. The estimated parameters from each bootstrap are used to calculate the elasticity as in (10), and when all bootstraps are combined, yield a probability distribution for the elasticity to use in hypothesis tests.

For each simulation, the null hypothesis is whether the elasticity is equal to $\tau\eta$ where τ equals the true elasticity one half the time and the other half is an uniformly distributed random variable over the (-2,2) interval. The confidence level is set to be 10%. Using the collection of estimated elasticities from the bootstrap, a confidence interval is constructed by selecting two values separating the lowest 5% and highest 5% of elasticities in the bootstrap series. Let η_L be the value for which η_L is greater than the lowest 5% of elasticities and less than the rest. Let η_H be the value for which η_H is less than the highest 5% of elasticities and greater than the rest. The two values η_L and η_H then form an empirical confidence interval. If $\tau\eta$ lies outside the (η_L, η_H) interval the null hypothesis that $\eta = \tau\eta$ is rejected.

The purpose of this study is to compare hypothesis tests using numerous models to tests using a single model. Therefore, the hypothesis test described previously is compared to the hypothesis of a single model. For each simulation the model with the lowest out-of-sample squared error (OSRMSE) is chosen as superior. This model is then used for hypothesis tests. Using the same methodology an empirical confidence interval is constructed for this single model and the same hypothesis test is conducted. This hypothesis test was conducted during each simulation for an elasticity observed at five different values of x_1 .

The hypothesis tests using the WSA versus the single model approach is compared by the frequencies of a Type I and Type II errors. The method which results in a probability of Type I error closest to 10% and the model with the greatest power will be deemed the better approach. Since the approaches are evaluated on two criteria, the WSA may fair better by one criteria but not another.

SIMULATION RESULTS

Table 1. provides descriptive statistics regarding the weights assigned in the composite model for the 1,716 simulations. The composite model prefers parsimonious models with a λ value of .01 and 1, although each model is used frequently. On average, the composite model employ 4 individual models. The frequency of Type I and Type II errors are shown in Table 2. Regarding Type I Errors, the frequency is lowest at the mean of x_1 , which is expected. An extremely large number of Type II errors are found, and surprisingly, is highest at the when the elasticity is observed at the mean of x_1 .

It seems counterintuitive that the percent of Type II errors were largest at the mean of x_1 , as the confidence interval surrounding the dependent variables is usually smallest at the means of the variables, which should increase the frequency of rejections. A possible, but not confirmed, explanation is provided. Due to the simulation code, elasticities tended to become larger as x_1 becomes larger. Hence, as x_1 increases from 60 to 100, the range of elasticities is larger implying less rejections. However, for models using a Box-Cox transformation on the independent variables, an increase in x_1 has a smaller impact on the dependent variable as it increases implying less variability in elasticities. As x_1 increases there exists two forces, one which tends to increase the range of elasticities and one which tends to decrease. The only explanation is that the mean of x_1 separates the regions where one forces dominates the other.

It is difficult to discern whether the Weighted Statistic Approach or the Single Model Approach dominated. The number of Type I errors are higher using the Weighted Approach, but the number of Type II errors are higher using the Single Model Approach. The percent of Type I and Type II errors are only significantly different for the range $x_1 = 80$ and 100. We do not feel these simulation results allow any general statements to be made regarding how well composite models perform relative to single models in hypothesis tests, so long as the composite and single models are chosen to minimize out-of-sample forecast error.

FOOTNOTES

- 1) The terms “model” and “functional form” are used interchangeably. They both mean a researcher’s description of how economic values are determined. They include both the functional form, the error distribution, and estimated parameter values.
- 2) This claim is made based on the fact there seems little consensus on the one best method of model selection.
- 3) Most estimators, like OLS regressions, appear unbiased in-sample because the residuals sum to zero. These same estimators are often found to be biased in out-of-sample forecasts though.
- 4) Usually elasticities are measured using the instantaneous percent change in y due to an instantaneous percent change in x_1 . This study uses the percent change in y due to a 1% change in x_1 to simplify simulation programming.
- 5) This is identical to the weights chosen by the ERS for retail food price forecasts.
- 6) The dependent variable was not transformed for coding reasons. Performing simulations when the dependent variables was transformed proved difficult because, since the parameters are random variables, the Box-Cox transformation resulting in complex numbers that were difficult to deal with.

REFERENCES

- Anderson, D.P., T. Chaisantikulawat, A.T.K. Guan, M. Kebbeh, N. Lin, and C.R. Shumway. "Choice of Functional Form for Agricultural Production Analysis." *Review of Agricultural Economics*. 18(1996):223-231.
- Brandt, Jon A. and David A. Bessler. "Composite Forecasting: An Application With US Hog Prices." *American Journal of Agricultural Economics*. February. 1981. 63:135-140.
- Dorfman, Jeffrey H. "Bayesian Composite Qualitative Forecasting: Hog Prices Again." *American Journal of Agricultural Economics*. August. 1998. 80:543-551.
- Economic Research Service. "Retail Food Price Forecasting at ERS: The Process, Methodology, and Performance from 1984 to 1997." Technical Bulletin Number 885. May. 2000.
- Makridakis, Spyros, Michele Hibson, El Lusk, and Moncef Belhadjali. "Confidence Intervals: An Empirical Investigation of the Series of the M-Competition." *International Journal of Forecasting* 3(1987):489-508.
- McAleer, Michael, Gordon Fischer, and Paul Volker. "Separate Misspecified Regressions and the U.S. Long-Run Demand for Money Function." *The Review of Economics and Statistics*.
- Min, Chung-ki and Arnold Zellner. "Bayesian and Non-Bayesian Methods for Combining Models and Forecasts with Applications to Forecasting International Growth Rates." *Journal of Econometrics*. 56(1993):89-118.
- Norwood, Bailey and Ted C. Schroeder. "A Preview of the Usefulness of Placement Weight Data." Proceedings of the NCR-134 Conference of Applied Commodity Price Analysis, Forecasting, and Market Risk Management, 1998.
- Shumway, Richard C. and Hongil Lim. "Functional Form and U.S. Agricultural Production Elasticities." *Journal of Agricultural and Resource Economics*. 18:2(1993):266-276.
- Wold, H.O. "Econometrics as Pioneering in Non-experimental Model Building." *Econometrica*. July. 1969. 37:369-81.

TABLE 1.
DESCRIPTIVE STATISTICS OF WEIGHTS USED IN COMPOSITIVE MODELS

Composite Model: $\hat{y} = \sum_{i=1}^{10} w_i \hat{y}_i$; \hat{y}_i is prediction from Model i

Models 1 through 5 are $E\{y\} = \hat{\mathbf{b}}_o + \sum_{i=1}^5 x_i^{(1)} \hat{\mathbf{b}}_i$ where $\lambda = .01, .25, .5, .75$, and 1

Models 5 through 10 are $E\{y\} = \hat{\mathbf{b}}_o + \sum_{i=1}^5 x_i^{(1)} \hat{\mathbf{b}}_i + \sum_{i=1}^5 \sum_{j=1}^5 x_i^{(1)} x_j^{(1)} \hat{\mathbf{b}}_{ij}$

where $\lambda = .01, .25, .5, .75$, and 1

Number of Simulations = 1,716

	Mean	Median	Maximum	Minimum	Standard Deviation	Percent of Time > 0
w ₁	0.26	0.06	1.00	0.00	0.35	0.72
w ₂	0.05	0.00	1.00	0.00	0.19	0.44
w ₃	0.04	0.00	1.00	0.00	0.13	0.32
w ₄	0.05	0.00	1.00	0.00	0.14	0.34
w ₅	0.34	0.23	1.00	0.00	0.35	0.73
w ₆	0.05	0.00	0.66	0.00	0.13	0.34
w ₇	0.06	0.00	0.99	0.00	0.19	0.39
w ₈	0.05	0.00	0.95	0.00	0.19	0.38
w ₉	0.05	0.00	0.85	0.00	0.16	0.39
w ₁₀	0.06	0.00	0.73	0.00	0.15	0.43
Number of Weights Within Simulation That Were Greater Than Zero	4.74	4	10 ^a	0 ^b	1.46	

a) Positive weights were assigned to all models only 2% of the time.

b) Only once were all weights assigned to zero. This may have happened because the dependent variable could take positive and negative values, and if all parameters in the true model are close to zero (which is possible as parameter values were randomly chosen) a value of zero may be the best prediction.

TABLE 2.
SIMULATION RESULTS: FREQUENCY OF TYPE I AND TYPE II ERRORS

Null Hypothesis: Elasticity of y with respect to x_1 is equal to a particular value	Percent of Type I Errors: Null Hypothesis Rejected When Null Hypothesis is True	Percent of Type II Errors: Null Hypothesis Not Rejected When Null Hypothesis is False		
1,716 Simulations				
When x_1 is observed at	Using Weighted Statistic Approach	Using Single Model Approach	Using Weighted Statistic Approach	Using Single Model Approach
$x_1 = 60$	0.29	0.28	0.65	0.64
$x_1 = 80$	0.20	0.14	0.75	0.82
$x_1 = 100$	0.12	0.07	0.85	0.92
$x_1 = 120$	0.23	0.19	0.76	0.80
$x_1 = 140$	0.27	0.25	0.71	0.72

TABLE 3.
SIMULATION RESULTS: T-TEST FOR SIGNIFICANT DIFFERENCES IN
TYPE I AND TYPE II ERRORS

Null Hypothesis: Elasticity of y with respect to x_1 is equal to a particular value	T-Test for Percent of Type I Errors is Greater Using Weighted Statistic Approach than Single Model Approach	T-Test for Percent of Type II Errors is Greater Using Weighted Statistic Approach than Single Model Approach
1,716 Simulations		
When x_1 is observed at	Test Statistic ^a : $\frac{\bar{P}_W - \bar{P}_S}{\sqrt{\frac{\bar{P}_W(1-\bar{P}_W)}{T} + \frac{\bar{P}_S(1-\bar{P}_S)}{T} - \frac{2\text{cov}(D_W, D_S)}{T}}}$	
$x_1 = 60$ $x_1 = 80$ $x_1 = 100$ $x_1 = 120$ $x_1 = 140$	0.28 2.67 3.06 1.29 0.63	0.39 -2.52 -3.62 -1.36 -0.38

a) Where \bar{P}_W is the percent of Type I or Type II Errors using the Weighted Statistic

Approach and equals $\bar{P}_W = \sum_{t=1}^T D_W / T$ where $D_W = 1$ if null was not rejected and zero if null was rejected and T is the number of times the number of times the null was false. \bar{P}_S denotes the Single Model Approach.

APPENDIX A SIMULATION DESCRIPTION

Dependent Variable: y

Independent Variables Affecting y: $x_1, x_2, x_3, x_4, x_5 = X$

Step 1: Generating Data on x and Parameter Values

The variables $x_1, x_2, x_3,$ and x_5 were chosen randomly from a normal distribution with a mean of 100 and a standard deviation of 20. The unconditional distribution of x_4 is the same, however, it was set to have a correlation of .3 with x_3 .

$x_i \sim N(100, 20^2)$ for $i = 1, 2, 3, 5$.

$x_4 \sim N(100 + .3(x_3 - 100), \{20^2(1 - .3^2)\})$

Step 2: Generate Model and Sample Size

The process of y is given by $y = X^{(\lambda)}(1:j)\beta(1:j) + e$ and is described below. The matrix $X(1:j)$ is a matrix containing x_i 's. The superscript (λ) signifies that each x_i undergoes a Box-Cox transformation. The value of lambda used for the Box-Cox transformation is chosen randomly. It may take the values .01, .02, ..., 1 for each x_i with equal probability. Each x_i is transformed as $x_i^{(\lambda)} = (x_i^{\lambda} - 1)/I_i$. Let $X^{(\lambda)} = [I_c, x_1^{(\lambda)}, \dots, x_5^{(\lambda)}, [x_1^{(\lambda)}]^2, \dots, [x_5^{(\lambda)}]^2, x_1^{(\lambda)}x_2^{(\lambda)}, \dots, x_1^{(\lambda)}x_5^{(\lambda)}, x_2^{(\lambda)}x_3^{(\lambda)}, \dots, x_4^{(\lambda)}x_5^{(\lambda)}]$ where I_c denotes a column of ones. Then, $X^{(\lambda)}(1:j)$ contains the first j columns of $X^{(\lambda)}$. For instance, $X^{(\lambda)}(1:6) = [I_c, x_1^{(\lambda)}, x_2^{(\lambda)}, x_3^{(\lambda)}, x_4^{(\lambda)}, x_5^{(\lambda)}]$ and $X^{(\lambda)}(1:21) = [I_c, x_1^{(\lambda)}, \dots, x_5^{(\lambda)}, [x_1^{(\lambda)}]^2, \dots, [x_5^{(\lambda)}]^2, x_1^{(\lambda)}x_2^{(\lambda)}, \dots, x_1^{(\lambda)}x_5^{(\lambda)}, x_2^{(\lambda)}x_3^{(\lambda)}, \dots, x_4^{(\lambda)}x_5^{(\lambda)}]$. The value of j can take on 6, ..., 21; each with an equal probability. Next, the parameter vector mapping $X^{(\lambda)}(1:j)$ into the expected value of y is $\beta(1:j)$. Let $\beta = [\beta_0 \beta_1 \beta_2 \beta_3 \beta_4 \beta_5 \beta_{12} \beta_{22} \beta_{32} \beta_{42} \beta_{52} \beta_{i12} \beta_{i13} \beta_{i14} \beta_{i15} \beta_{i23} \beta_{i24} \beta_{i25} \beta_{i34} \beta_{i35} \beta_{i45}]'$ where β_{i25} denotes the parameter corresponding to the interaction term $x_1^{(\lambda)}x_5^{(\lambda)}$. The vector $\beta(1:j)$ then contains the first j rows of β .

The moments of each parameter are:

$\beta_0 \sim N(10000, 100^2)$; $\beta_1 \sim N(10, 3^2) * 20^{(1-\lambda_1)}$; $\beta_2 \sim N(20, 6^2) * 20^{(1-\lambda_2)}$;
 $\beta_3 \sim N(15, 4^2) * 20^{(1-\lambda_3)}$; $\beta_4 \sim N(8, 2^2) * 20^{(1-\lambda_4)}$; $\beta_5 \sim N(18, 5^2) * 20^{(1-\lambda_5)}$;
 $\beta_{12} \sim N(.01, .005^2) * 20^{(1-\lambda_{12})}$; $\beta_{22} \sim N(.001, .005^2) * 20^{(1-\lambda_{22})}$; $\beta_{32} \sim N(.03, .008^2) * 20^{(1-\lambda_{32})}$;
 $\beta_{42} \sim N(.004, .0002^2) * 20^{(1-\lambda_{42})}$; $\beta_{52} \sim N(.0005, .00004^2) * 20^{(1-\lambda_{52})}$;
 $\beta_{i12} \sim N(.001, .005^2) * 5^{(1-\lambda_{i12})} * 5^{(1-\lambda_{i2})}$; $\beta_{i13} \sim N(.0001, .0005^2) * 5^{(1-\lambda_{i13})} * 5^{(1-\lambda_{i3})}$;
 $\beta_{i14} \sim N(.0005, .005^2) * 5^{(1-\lambda_{i14})} * 5^{(1-\lambda_{i4})}$; $\beta_{i15} \sim N(.0008, .0002^2) * 5^{(1-\lambda_{i15})} * 5^{(1-\lambda_{i5})}$;
 $\beta_{i23} \sim N(.00001, .00005^2) * 5^{(1-\lambda_{i23})} * 5^{(1-\lambda_{i3})}$; $\beta_{i24} \sim N(.001, .0005^2) * 5^{(1-\lambda_{i24})} * 5^{(1-\lambda_{i4})}$;
 $\beta_{i25} \sim N(.01, .005^2) * 5^{(1-\lambda_{i25})} * 5^{(1-\lambda_{i5})}$; $\beta_{i34} \sim N(.0003, .0001^2) * 5^{(1-\lambda_{i34})} * 5^{(1-\lambda_{i4})}$;
 $\beta_{i35} \sim N(.00025, .00005^2) * 5^{(1-\lambda_{i35})} * 5^{(1-\lambda_{i5})}$; $\beta_{i45} \sim N(.0025, .003^2) * 5^{(1-\lambda_{i45})} * 5^{(1-\lambda_{i5})}$

After these parameter values are simulated, each parameter is multiplied by one with a 50% chance and by -1 with a 50% chance. After the vector β is simulated, the expected value of y is then denoted as $X^{(\lambda)}(1:j)\beta(1:j)$.

The value of observations of y for each simulation consists of its expected value, $X^{(\lambda)}(1:j)\beta(1:j)$, and a stochastic error. The error distribution is normal with a constant variance, where the size of the variance varies among simulations. Within each simulation the value of g can equal .01, .02, ..., .5 with equal probability. Let $mX^{(\lambda)}(1:j)$ be the mean vector of $X^{(\lambda)}(1:j)$ for the sample. The error variance is then set to equal $\{gmX^{(\lambda)}(1:j)\beta(1:j)\}^2$. The process governing y can then be completely described by $y = X^{(\lambda)}(1:j)\beta(1:j) + e$ where $e \sim N(0, \{gmX^{(\lambda)}(1:j)\beta(1:j)\}^2)$.

Step 4: Generating the Data

Data on $X^{(\lambda)}$ and the value of $\beta(1:j)$ have already been generated. Data on y is then generating by simulating values of e from a normal distribution with a zero mean and variance as described above. The sample size may be any even number between, and including, 60 and 500 with equal probability. The “dataset” is then the collection of y 's and x_i 's.

Step 5: Estimation of Weights Assigned to Each Approximating Model

The true functional form and error variance is unknown to the researcher. The simulated researcher then proceeds to approximate the true functional form using 10 different models. These models are:

$$\text{Model 1} = M(i=5, \lambda=.01): E\{y\} = \mathbf{b}_o + \sum_{i=1}^5 x_i^{(I=.01)} \hat{\mathbf{b}}_i$$

$$\text{Model 2} = M(5, .25): E\{y\} = \mathbf{b}_o + \sum_{i=1}^5 x_i^{(I=.25)} \hat{\mathbf{b}}_i$$

$$\text{Model 3} = M(5, .5): E\{y\} = \mathbf{b}_o + \sum_{i=1}^5 x_i^{(I=.5)} \hat{\mathbf{b}}_i$$

$$\text{Model 4} = M(5, .75): E\{y\} = \mathbf{b}_o + \sum_{i=1}^5 x_i^{(I=.75)} \hat{\mathbf{b}}_i$$

$$\text{Model 5} = M(5, 1): E\{y\} = \mathbf{b}_o + \sum_{i=1}^5 x_i^{(I=1)} \hat{\mathbf{b}}_i$$

$$\text{Model 6} = M(i=21, \lambda=.01): E\{y\} = \mathbf{b}_o + \sum_{i=1}^5 x_i^{(I=.01)} \hat{\mathbf{b}}_i + \sum_{i=1}^5 \sum_{j=1}^5 x_i^{(I=.01)} x_j^{(I=.01)} \hat{\mathbf{b}}_{ij}$$

$$\text{Model 7} = M(i=21, \lambda=.25): E\{y\} = \mathbf{b}_o + \sum_{i=1}^5 x_i^{(I=.25)} \hat{\mathbf{b}}_i + \sum_{i=1}^5 \sum_{j=1}^5 x_i^{(I=.25)} x_j^{(I=.25)} \hat{\mathbf{b}}_{ij}$$

$$\text{Model 8} = M(i=21, \lambda=.5): E\{y\} = \mathbf{b}_o + \sum_{i=1}^5 x_i^{(I=.5)} \hat{\mathbf{b}}_i + \sum_{i=1}^5 \sum_{j=1}^5 x_i^{(I=.5)} x_j^{(I=.5)} \hat{\mathbf{b}}_{ij}$$

$$\text{Model 9} = M(i=21, \lambda=.75): E\{y\} = \mathbf{b}_o + \sum_{i=1}^5 x_i^{(I=.75)} \hat{\mathbf{b}}_i + \sum_{i=1}^5 \sum_{j=1}^5 x_i^{(I=.75)} x_j^{(I=.75)} \hat{\mathbf{b}}_{ij}$$

$$\text{Model 10} = M(i=21, \lambda=1): E\{y\} = \mathbf{b}_o + \sum_{i=1}^5 x_i^{(I=1)} \hat{\mathbf{b}}_i + \sum_{i=1}^5 \sum_{j=1}^5 x_i^{(I=1)} x_j^{(I=1)} \hat{\mathbf{b}}_{ij}$$

The first-half of the sample of y and X is used to estimate the parameters. These parameters are then used to conduct out-of-sample forecasts for the second-half of the sample. Let t denote an out-of-sample forecast and $\hat{y}_{i,t}$ be the t^{th} forecast from Model $i = 1, 2, \dots, 10$. The weight assigned to each model is calculated by estimating

$$\max_{w_i} \sum_{t=1}^T \left\{ \frac{\left(\sum_{i=1}^K w_i \hat{y}_{i,t} - y_t \right)^2}{T} \right\} + \mathbf{I} \left(1 - \sum_{i=1}^K w_i \right) \text{ s.t. } 0 \leq w_i \leq 1 \forall i$$

using a constrained maximization routine titled “constr” in Matlab.

Step 5: Calculating the Probability Distribution of the Elasticity Using the Weighted Statistic Approach:

The entire sample of observations on y and X are then sampled with replacement for a series of 200 bootstraps within each simulation. For each bootstrap, the new collection of y and X observations are used to re-estimate the ten Models $M(i, \lambda)$ for $i = 6, 21$ and $\lambda = .01, .25, .5, .75, \text{ and } 1$. The elasticity for each bootstrap is calculated as

$$\hat{\mathbf{h}}_{\text{WSA},b} = \left\{ \frac{\sum_{i=1}^K w_i \hat{y}_i(\bar{x}_1(1.01), \dots, \bar{x}_5) - \sum_{i=1}^K w_i \hat{y}_i(\bar{x}_1, \dots, \bar{x}_5)}{\sum_{i=1}^K w_i \hat{y}_i(\bar{x}_1, \dots, \bar{x}_5)} \right\} / .01 \text{ where the subscript } b \text{ denotes}$$

a bootstrap and the subscript WSA denotes the Weighed Statistic Approach.. The 200 $\hat{\mathbf{h}}_{\text{WSA},b}$ ’s then constitute an empirical probability distribution for the estimated elasticity.

Step 6: Calculating the Probability Distribution of the Elasticity Using the A Single Model

In Step 4 where the out-of-sample forecasts are conducted, the model with the smallest out-of-sample-root-mean-squared error is deemed the “superior model” and is used to estimate the elasticity. Let $\hat{y}_s(\bar{x}_1, \dots, \bar{x}_p)$ be the prediction using this single “superior” model. Another 200 bootstraps are conducted within each simulation where the estimated elasticity is calculated as

$$\hat{h}_{s,b} = \left\{ \frac{\hat{y}_s(\bar{x}_1(1.01), \dots, \bar{x}_5) - \hat{y}_s(\bar{x}_1, \dots, \bar{x}_5)}{\hat{y}_s(\bar{x}_1, \dots, \bar{x}_5)} \right\} / .01 \text{ where the subscript } b \text{ denotes a bootstrap.}$$

The bias from the superior model was assumed zero as is typically done by researchers. This is not necessarily because the bias is believed zero, but because the object was to simulate how research is conducted using the single model approach.

The 200 \hat{h}_b 's then constitute an empirical probability distribution for the estimated elasticity.

Step 7: Hypothesis Testing Using the Weighted Statistic Approach and the Single Model Approach

The probability density function for the estimated elasticity using the WSA is the collection of $\hat{h}_{WSa,b}$'s. Let $\hat{h}_{WSa,b}^{(O)}$ be the vector of $\hat{h}_{WSa,b}$'s in order of lowest to highest where $\hat{h}_{WSa,b}^{(1)}$ is the lowest elasticity and $\hat{h}_{WSa,b}^{(200)}$ is the highest. The vector $\hat{h}_{WSa,b}^{(O)}$ is then a collection of order statistics. The 90% empirical confidence interval for the WSA is then $(\hat{h}_{WSa,b}^{(11)}, \hat{h}_{WSa,b}^{(189)})$. The 90% empirical confidence interval for the single model is $(\hat{h}_{s,b}^{(11)}, \hat{h}_{s,b}^{(189)})$.

For each simulation, the hypothesis of interest is whether the elasticity is statistically different from $\tau\eta$ where τ is a uniformly distributed random variable, in .01 increments, in the (-2,2) interval. The null hypothesis is that η is equal to $\tau\eta$.

The null hypothesis using either the WSA and the single model approach is rejected if $\tau\eta$ lies outside the $(\hat{h}_{WSa,b}^{(11)}, \hat{h}_{WSa,b}^{(189)})$ and $(\hat{h}_{s,b}^{(11)}, \hat{h}_{s,b}^{(189)})$ interval, respectively.