DALLASFED

Occasional Paper

The Chen-Tindall System and the Lasso Operator: Improving Automatic Model Performance

Jiaqi Chen and Michael L. Tindall

Federal Reserve Bank of Dallas
Financial Industry Studies Department
Occasional Paper 16-01
May 2016

The Chen-Tindall System and the Lasso Operator: Improving Automatic Model Performance

by

Jiaqi Chen and Michael L. Tindall
Financial Industry Studies Department
Federal Reserve Bank of Dallas*
May 2016

Abstract

Using U.S. monthly macroeconomic data, the automatic model system presented in Chen and Tindall [2016] outperforms the lasso automatic system, but the lasso is improved where Bayesian model averaging is employed to combine its forecasts with those from autoregressive schemes. The best performance is obtained using Bayesian model averaging to combine the Chen–Tindall system, the lasso, and autoregressive schemes. Performance is virtually the same using this combined approach where the elastic-net operator is substituted for the lasso. Similar overall outcomes are found for France and Germany treated as a single economic system and for Canada.

Keywords: Automatic model building, Bayesian model averaging, the lasso, the elastic net.

Previously, Chen and Tindall [2016], C–T hereafter, presented an automatic model-building system, the CTA system hereafter, and examined its out-of-sample macroeconomic forecasting performance over the course of the last recession and subsequent recovery across multiple national economies. C–T focused on the construction of the CTA system with emphasis on 1) the set of rules under which it operates, and 2) the use of Bayesian model averaging, BMA hereafter, in its operation.

In this paper, the CTA system is applied to a larger U.S. database than that used in C–T, the database is extended through the end of 2014, and BMA is employed in alternative ways to combine forecasts from the CTA system, the lasso system (Tibshirani [1996]), and other schemes. We apply the same processes to a combined database for France and Germany and a Canada database. We find that 1) the CTA system and the lasso generate forecasting equations that are very different in structure, 2) the CTA system, which is designed to incorporate BMA, generally outperforms the lasso, which does not

*Please direct correspondence to jiaqi.chen@dal.frb.org or michael.tindall@dal.frb.org. The authors thank SungJe Byun for his comments on the paper. The views expressed herein are those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Dallas or the Federal Reserve System.

benefit from BMA, 3) combining forecasts from the CTA and lasso systems with forecasts from autoregressive schemes using a two-step BMA approach produces superior performance results, and 4) superior performance can also be achieved substituting the elastic-net operator, a variation on the lasso proposed by Zou and Hastie [2005], for the lasso in the two-step BMA approach.

Automatic model-building systems address key problems which arise where forecasting models are based on economic theory. First, the economy is a complex, dynamic system where relationships are constantly evolving because of changes in economic policy, technology, etc. complicating the formulation of theory upon which forecasting models can be based. Second, economic theory is a simplified representation of reality which may not capture information that could be useful in making forecasts. Third, available time series used in building forecasting models are often poor measures of the concepts embodied in economic theory. Given these problems, the best one can do may be to build forecasting models based on statistical properties in the data. This is what automatic forecasting systems are meant to do. They appeal directly to the data, bypassing economic theory.

Automatic systems operate without the need for human judgment, and they can build forecasting models even where the dataset is so large that its size effectively prohibits construction by a limited number of human model builders. Automatic systems are unconstrained by preconceived ideas, i.e., economic theories, and so can provide new ideas about model construction that can be used by analysts to improve existing models built by human developers.

In the area of traditional econometric forecasting, where econometric model-building is performed without automatic systems, the Federal Reserve has, over a period of about five decades, committed substantial resources to the development and implementation of econometric models as an aid to policymaking. In recent years, automatic systems have become a topic of research at the Federal Reserve. At the Federal Reserve Board of Governors, Ericsson has published extensively, as author and co-author, on the application of automatic model-building systems to forecasting and policy analysis (see Campos, Ericsson, and Hendry [2005], Ericsson [2016a], Ericsson [2016b], and Ericsson, Hendry, and Hood [2016]). Another banking regulator, the Federal Deposit Insurance Corporation, has conducted research regarding their use as tools for performing key bank monitoring functions (see Kapinos and Mitnik [2015]). In addition, researchers at the European Central Bank have examined automatic methods as forecasting tools (see De Mol, Giannone, and Reichlin [2006]). Properly constructed automatic systems offer policymakers the prospect of advances in the kinds of economic analysis that underlie policy formulation.

In the following section, we present the data used in the analysis. In the next section, we present an overview of the operation of the CTA system and the lasso. Next, we present out-of-sample forecast results and examine them comparatively. Then, we present conclusions.

DATA

All data used in the analysis are actual, rather than artificial, data. All data are monthly in frequency. There is a forecasting equation for every series in each of the economic databases used in the analysis. We employ a U.S. database consisting of 434 variables, a database for France and Germany, treated as a single economy, consisting of 105 series, and a Canada database consisting of 61 series. The appendix shows the series in the three databases broken down by types of data. The databases are not intended to be comprehensive in scope or detail; in practice, an automatic system may be presented with datasets that are not comprehensive.

The U.S. economy is an obvious choice for automatic model-building work. Following C–T, we also apply automatic model building to a combined database of time series from France and Germany. C–T argue that France and Germany can be combined into a single economic system because 1) the two countries use the euro as their common currency so that their markets are based on the same numéraire, 2) they employ harmonized regulatory environments, 3) they operate within the common legal framework of the European Union, 4) they trade with each other across a shared border with a high degree of specialization and economic integration, and 5) financial capital flows freely in large amounts between them. Thus, we treat France and Germany as a single economy, which means that equations for time series in each country may contain right-hand variables from both countries. Combining France and Germany into a single economic system has the effect of subjecting automatic model-building systems to a rather unconventional forecasting problem. Canada's unique economic structure presents another challenge to automatic model-building systems. Its economy is technologically advanced, but Canada also has a large agricultural sector and is a major producer and exporter of timber and petroleum products.

AN OVERVIEW OF THE CTA SYSTEM AND THE LASSO

The CTA System

C–T present a full description of the CTA system, but in summary its key distinguishing feature is that it is comprised of a set of rules developed experimentally, i.e., through empirical tests of alternative combinations of computational methods. C–T say that this feature sets the system apart from those automatic systems based on theories of model selection. They note that there is a history of controversy in the literature over the issue of whether it is possible to posit a theory of automatic model selection, and they present an approach where the automatic system is developed and constructed using empirical means applied to a U.S. database and then tested without further alteration using various other countries' databases. C–T note that the computational methods employed in the CTA system are not new taken by themselves except that they make small modifications to some of the methods. Rather, they propose that the CTA system may be novel in that it is constructed using this empirical approach.

The CTA system begins computation by converting each time series in the database to stationary form. C–T test alternative methods of conversion, and a step-wise method is chosen based on empirical tests of model performance. A modified form of Granger causality tests is then applied to create a list of time series for possible inclusion as independent variables in the forecasting equation for a given dependent variable. That is, each dependent variable has its own list of candidate independent variables generated by the modified Granger causality tests. The Granger causality tests are modified so that bidirectional causality, where x Granger causes y, and y Granger causes x, is not allowed. Where bidirectional Granger relationships arise between two variables, the weaker relationship based on p-values of *F* statistics is ignored. C–T find that this modification, which disallows tight feedback loops between two variables, increases the stability of the forecasting system improving model performance in empirical tests.

Using the list of variables created by the Granger tests, a modified form of the Bayesian information criterion is employed to construct the forecasting equation in a process that switches back and forth automatically between forward selection and backward deletion. The modified Bayesian information criterion is selected from a set of various information criteria based on empirical tests of model performance. Forward selection and backward deletion are computed using cross validation. C–T argue that the seasonal-adjustment methods used by reporting agencies do not always effectively remove seasonality from time series and that this may leave substantial residual seasonal variation in equations. So, in the next step, seasonal binary variables are appended to the equation even where all variables in the equation are ostensibly seasonally adjusted, and the equation is again subjected to backward deletion. This improves overall model performance in empirical tests. Following that, a time-trend variable is appended, and the equation is subjected to backward deletion. This improves overall model performance marginally. We refer to this combined set of processes as the initial phase of model construction.

Next, BMA is employed to combine this initial model and two autoregressive modeling schemes: 1) an AR(12) model and 2) an AR(1) with appended seasonal binary variables as in $y_t = \beta_1 + \beta_2 y_{t-1} + \sum_{j=1}^{11} \beta_{j+2} s_{t-j+1} + \varepsilon_t$ where s denotes the seasonal binary variable. C–T find that this application of BMA substantially improves forecasting performance in empirical tests.

They argue that the improvement is a result of the fact that the three approaches represent the data in very different ways. While the model generated in the initial phase of construction allows feedback effects, the AR(12) model only contains information from the lagged dependent variable. This eliminates forecast instability arising from feedback effects. Feedback effects are also eliminated in the equations for the AR(1) process with appended seasonal binary variables, but C–T argue that there is another reason for incorporating forecasts from this model. Seasonal binary variables are appended to the equations produced in the initial phase of construction, but these binary variables are then subjected to backward deletion. Other variables may mimic the effects of seasonality and provide additional information other than seasonality and, thus, at times, displace seasonal binary variables in the backward-deletion phase that would otherwise survive the process. So, the forecasts from the AR(1) equations with appended seasonal binary variables may serve to restore lost information about seasonality to the forecasts using BMA. In C–T, using U.S. data, the average weights computed from BMA are 0.456, 0.277, and 0.267 for, respectively, 1) the equations constructed in the initial phase, 2) the AR(12) model, and 3) the AR(1) with appended seasonal binary variables. The weighting patterns for the other country databases examined in C–T are very similar.

The CTA system employs the Gauss-Seidel method (Jeffreys and Jeffreys [1988]), GS hereafter, used in various mathematics software packages to solve systems of equations. Under GS, in each forecast period, each contemporaneous right-hand variable is set initially at its value in the previous period, and the forecasting equations are computed using that value in the forecasting equations. This generates updates of the contemporaneous right-hand variables, and the forecasting equations are computed again. The process is repeated until the values of variables stabilize. With well-behaved systems of equations, GS produces results identical to Cramer. In the empirical results presented below, where various models are combined using BMA, the Gauss-Seidel method is used to solve the system of equations.

The Lasso System

The lasso (least absolute shrinkage and selection operator) is well known, but in summary it is a shrinkage method, which, for a linear regression $y_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} + \varepsilon_i, i = 1,2,3, \dots, n,$ can be written in the form of a Lagrangian (Hastie, Tibshirani, and Friedman [2009]):

$$\hat{\beta}^{lasso} = \underset{\beta}{\operatorname{argmin}} \left\{ \sum_{i=1}^{n} (y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij})^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right\}$$

The term λ is a tuning parameter, and $\lambda \sum_{j=1}^{p} |\beta_j|$ penalizes nonzero $\beta_j; j = 1,2,3, \dots, p$. The effect of the penalty is that some of the β_j are shrunk to exactly zero eliminating variables, making the lasso an automatic system. The solutions to the Lagrangian are nonlinear, and there is no closed-form solution. The underlying idea of the lasso is that overall forecast error can be reduced by introducing some error due to bias. Tibshirani's seminal 1996 paper on the lasso has been cited more than 14,000 times to date, and the lasso method has gained many adherents in the fields of economics and finance.

The paper focuses on two automatic model-building systems: the CTA system and the lasso. However, in the empirical results presented below, we also examine the forecasting performance of the elastic-net operator. The elastic-net operator is predicated on the idea that neither the lasso nor the ridge-regression operator has a clear advantage over the other, and the elastic net is proposed as an operator that combines properties of both. The elastic-net operator can be written in the form of a Lagrangian:

$$\hat{\beta}^{elastic\ net} = \underset{\beta}{\operatorname{argmin}} \left\{ \sum_{i=1}^{n} (y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij})^2 + \lambda \sum_{j=1}^{p} \left[(1-\alpha)|\beta_j| + \alpha\beta^2 \right] \right\}$$

The term α, like λ, is a tuning parameter. For $\alpha = 0$, the expression above is the lasso. For $\alpha = 1$, it is the ridge-regression operator. The idea underlying the elastic net is that α can be tuned between 0 and 1 to take advantage of the specific strengths of the two operators.

Differences in Forecasting Models Produced by the CTA System and the Lasso

The CTA system and the lasso obviously are based on very different concepts of automatic construction, and they produce forecasting models that are very different in structure. The forecasting equations produced by the CTA system in the initial phase of construction often have a characteristic appearance. The CTA system converts most time series to differences of logs. The regression equations may contain short distributed lags of the dependent variable and certain independent variables. Sometimes there are one or two independent variables in an equation with large *t* statistics in absolute value, which make immediate intuitive economic sense as determinants of the dependent variable. These key variables may appear as single variables or in the form of distributed lags. In addition, there are sometimes other independent variables with smaller *t* statistics that seem to serve as modifier variables, i.e., independent variables that serve to modify the relationship between the dependent variable and the key independent variables. One or more seasonal dummy variables may be present in some equations. Less frequently, the equation contains the time-trend variable. Some equations contain independent variables that are difficult to rationalize in economic terms. Infrequently, equations contain only the intercept term. Where the dependent variable is in the form of differences of logarithms, such an equation means of course that the series is forecast to grow at a constant rate. Estimation performed by the CTA system is OLS. However, residual autocorrelation and heteroskedasticity are generally absent in the equations. Conversion of time series to stationary form contributes to this effect.

Exhibit 1 presents a selected example of an equation created by the CTA system. The dependent variable is single-unit housing starts. The sample period is January 2001 to December 2011, which is the initial sample period used in the out-of-sample forecasting results presented below. The equation has 18 parameters, and it contains a short distributed lag of the dependent variable and a distributed lag of single-unit building permits. A housing start is defined as the excavation of the land for the construction of a new house. In most localities in the country, the right to excavate requires a building permit. A building permit can be used for months, varying with the locality, after it is obtained, which explains the distributed lag of building permits with positive coefficients. Once excavation takes place, the building permit cannot be used again, which explains the negative coefficients of the distributed lag of housing starts. The two sets of estimated coefficients sum to approximately 1.0 which is in agreement with our interpretation of their meaning. Thus, building permits seem to serve as the kind of key independent variable we described above, and the other independent variables in the equation seem to play the role of modifier variables to the key relationship between starts and permits.

Exhibit 1
Housing Starts Equation Generated by the CTA System

Dependent variable:	$\Delta\ln(HS_t)$
BIC:	-6.058
Sample period:	January 2001 to December 2011
Observations/parameters:	132/18

Variable	Coefficient	Standard Error
Intercept = 1	0.00698	0.0041
1. $\Delta\ln(HS_{t-1})$	-0.62956	0.0716
2. $\Delta\ln(HS_{t-2})$	-0.48641	0.0818
3. $\Delta\ln(HS_{t-3})$	-0.22045	0.0720
4. $\Delta\ln(BP_t)$	0.58655	0.0824
5. $\Delta\ln(BP_{t-1})$	0.74247	0.0962
6. $\Delta\ln(BP_{t-2})$	0.40127	0.1105
7. $\Delta\ln(BP_{t-3})$	0.34972	0.1044
8. $\Delta\ln(BP_{t-4})$	0.21425	0.0908
9. $\Delta\ln(IPFAB_t)$	1.47439	0.3870
10. $\Delta\ln(IPFAB_{t-4})$	-1.36260	0.4527
11. $\Delta\ln(IPFAB_{t-6})$	-1.09913	0.4779
12. $\Delta\ln(IPFAB_{t-7})$	1.63958	0.4204
13. $\Delta\ln(PPICF_{t-10})$	0.56761	0.1113
14. $\Delta\ln(PCEON_t)$	1.45003	0.4770
15. $\Delta\ln(PCEON_{t-8})$	-1.69359	0.4771
16. $\Delta\ln(PCEON_{t-10})$	-1.19497	0.5021
17. $\Delta\ln(PCEON_{t-12})$	-1.02638	0.4331
18. $SEASON_{t-8}$	-0.03215	0.0117

where:
HS = Housing starts: single unit
BP = Building permits: single unit
IPFAB = Industrial production: fabricated metal products
PPICF = Producer price index: crude foodstuffs and feedstuffs
PCEON = Personal consumption expenditures: other nondurable, chained dollars
SEASON = Seasonal binary variable = 1 in January, 0 otherwise

Contrasted with equations created by the CTA system, lasso-generated equations using our data typically contain many more independent variables, and the independent variables often appear with relatively small estimated coefficients. Distributed lags are infrequent in lasso-generated equations using our data. The large number of independent variables in a lasso-generated equation can complicate economic interpretation.

Exhibit 2 shows the lasso-generated equation for single-unit housing starts. The dependent variable appears on the right-hand side of the equation only once and with a one-period lag, and building permits appear only once as a contemporaneous independent variable. The coefficients of these two variables sum to about 0.43 rather than 1.0 as in the corresponding CTA equation. The equation contains 39 parameters, which means that the lasso selects many more of the kind of independent variables that we refer to as modifier variables.

Exhibit 2

Housing Starts Equation Generated by the Lasso

Dependent variable:	$\Delta\ln(HS_t)$
BIC:	-6.612
Sample period:	January 2001 to December 2011
Observations/parameters:	132/39

Variable	Coefficient	Standard Error
1. Intercept = 1	-0.00767	0.0126
2. $\Delta\ln(HS_{t-1})$	-0.01998	0.0282
3. $\Delta\ln(BP_t)$	0.44830	0.1729
4. $\Delta\ln(IPCON_t)$	0.04572	0.2088
5. $\Delta\ln(IPMET_{t-10})$	0.06107	0.0727
6. $\Delta\ln(IPMIS_t)$	0.04405	0.1883
7. $\Delta\ln(UCMAN_t)$	0.42591	0.1996
8. $\Delta\ln(PCMOT_{t-11})$	0.02908	0.0414
9. $\Delta\ln(PCODUR_{t-8})$	-0.65370	0.3050
10. $\Delta\ln(PCFIN_{t-12})$	-0.00420	0.1077
11. $\Delta\ln(PCGAS_{t-7})$	0.10079	0.0705
12. $\Delta\ln(PCONON_{t-9})$	0.56139	0.2905
13. $\Delta\ln(RIFOOD_{t-5})$	-0.10519	0.3636
14. $\Delta\ln(NUTIL_{t-2})$	-0.66224	1.1165
15. $\Delta\ln(NRE_{t-3})$	1.92980	1.3655
16. $\Delta\ln(EARNCOM_{t-11})$	-0.30676	0.3630
17. $\Delta\ln(DOLLAR_{t-6})$	-0.27922	0.2301
18. $\Delta\ln(HOMENE_{t-5})$	-0.02092	0.0180
19. $\Delta\ln(HOMEPRI_{t-1})$	0.01627	0.0536
20. $\Delta(RSMOT_{t-11})$	4.2284×10^{-7}	8.110×10^{-7}
21. $\Delta\ln(RSFURN_t)$	0.25011	0.2094
22. $\Delta\ln(RSGEN_{t-3})$	0.05796	0.1742
23. $\Delta\ln(WMACH_t)$	0.01709	0.0863
24. $\Delta\ln(WMISDUR_{t-8})$	-0.14253	0.0710
25. $\Delta\ln(WDRUG_{t-12})$	0.49080	0.2478
26. $\Delta\ln(WALC_{t-3})$	0.00121	0.0611
27. $\Delta\ln(WALC_{t-11})$	-0.08150	0.1228
28. $\Delta\ln(WIALC_{t-7})$	-0.03995	0.1231
29. $\Delta\ln(MSMISDUR_{t-3})$	-0.13634	0.0898
30. $\Delta\ln(MSFOOD_{t-11})$	-0.11275	0.1541
31. $\Delta\ln(MSTEXT_{t-9})$	0.03166	0.0798
32. $\Delta\ln(MNFAB_{t-9})$	0.02231	0.0375
33. $\Delta\ln(MICHEM_{t-3})$	-0.12628	0.1354
34. $\Delta\ln(IMAUTO_{t-4})$	0.12048	0.0662
35. $\Delta\ln(IMOMER_{t-10})$	0.02793	0.0447
36. $\Delta\ln(IMOMER_{t-11})$	-0.02097	0.0779
37. $\Delta\ln(ISCHARGE_{t-1})$	-0.00742	0.0171
38. $\Delta(PHILCUR_{t-7})$	0.00051	0.0004
39. $\Delta(PHILEMP_{t-11})$	-0.00025	0.0003

where:

HS = Housing starts: single unit

BP = Building permits: single unit

IPCON = Industrial production: construction supplies

IPMET = Industrial production: primary metals

IPMIS = Industrial production: miscellaneous durable goods

UCMAN = Capacity utilization: manufacturing

PCMOT = Personal consumption: motor vehicles, current dollars
PCODUR = Personal consumption: other durable goods, current dollars
PCFIN = Personal consumption: financial services, current dollars
PCGAS = Personal consumption: gasoline and other energy, chained dollars
PCONON = Personal consumption: other nondurable good, chained dollars
RIFOOD = Retails inventories: food and beverage
NUTIL = Nonfarm payroll employment: utilities
NRE = Nonfarm payroll employment, real estate, rental, leasing
EARNCOM = Average hourly earnings: computer and electronic
DOLLAR = Nominal broad trade-weighted value of the dollar
HOMENE = New single-family houses sold: northeast
HOMEPRI = New single-family houses: mean sales price
RSMOT = Retail sales: motor vehicles
RSFURN = Retail sales: furniture and home furnishings
RSGEN = Retail sales: general merchandise
WMACH = Wholesale sales: machinery
WMISDUR = Wholesale sales: miscellaneous durable goods
WDRUG = Wholesale sales: drugs
WALC = Wholesale sales: alcohol
WIALC = Wholesale inventories: alcohol
MSMISDUR = Manufacturing shipments: miscellaneous durable goods
MSFOOD = Manufacturing shipments: food
MSTEXT = Manufacturing shipments: textile mills
MNFAB = Manufacturing new orders: fabricated metals products
MICHEM = Manufacturing inventories: basic chemicals
IMAUTO = Imports: retail auto
IMOMER = Imports: other merchandise
ISCHARGE = Imports services: charges of use of intellectual property
PHILCUR = Philadelphia Fed manufacturing business outlook: current delivery
PHILEMP = Philadelphia Fed manufacturing business outlook: current employment

We noted above that C–T report that the performance of their system is substantially improved using BMA to combine the forecasts from the initial model with those from the two autoregressive schemes and that they argue that the success of applying BMA to their system in this way arises from the fact that the initial model and the two autoregressive schemes represent the data in very different ways. This raises two key issues. First, given that that application of BMA substantially improves the performance of the CTA system, can the performance of the lasso be improved using BMA to combine its forecasts with those of the two autoregressive schemes? Second, and perhaps more important, given the marked structural differences in the equations produced by the CTA system and lasso-generated equations, as exemplified in the exhibits above, can we use BMA to combine forecasts from these systems, i.e., the CTA system and the lasso, improving forecasting performance even further? We return to these issues in the section below where we examine and present forecasting results.

OUT-OF-SAMPLE FORECASTS

Forecasts are computed over the three-year period from January 2012 to December 2014. The U.S. and Canada sample periods start in January 2001. The France–Germany sample period starts in January 2002, a year later because of data availability issues. To compute forecasts, the stationary forms of the variables and the equations of each macroeconomic model are created using data for the sample periods from the respective start periods to December 2011. Out-of-sample forecasts of one to six months ahead are computed for the six-month period January 2012 to June 2012, and the resulting root mean square errors (RMSEs) one to six months ahead are stored. Then, the sample period is lengthened one month to end in January 2012, and new stationary forms and equations are created with that sample. Out-of-sample forecasts of one to six months are computed for February 2012 to July 2012, and the RMSEs one to six months ahead are stored. The process is continued, extending the end month by one month and computing out-of-sample results.

Alternative Yardsticks and Performance Measurement

Alternative forecasting approaches are also computed as yardsticks for comparison with the system's forecasts. In one of these, the forecast is set equal to the actual value in the last period before the forecast period begins. We refer to this as the naïve model. The naïve approach provides a very useful diagnostic in that it is a "flat" forecast, which means that, if the automatic system outperforms the naïve yardstick, the system is in a sense "getting right" the direction, up or down, of the time series in the forecast period. We also compute an AR(12) model for each variable in the database and use it as a benchmark, comparing its forecasts to those from the automatic system. AR models are often used as a benchmark comparison for tests of forecasting systems (Stock and Watson [2003], Diebold and Li [2006], Stock and Watson [2008], and Elliott and Timmermann [2013]).

Let RS_{ij} denote the RMSE of variable i for the forecast j months ahead produced by an automatic system and RB_{ij} denote the corresponding RMSE of a benchmark model such as the naïve model or the AR(12). In the empirical results presented below, we report the mean D_j in forecast period j of the ratios RB_{ij}/RS_{ij}.

$$D_j = \sum_{i=1}^{n} \frac{RB_{ij}/RS_{ij}}{n};\ j = 1,2,3,\ldots,6$$

where $n =$ number of variables in the model. We use D_j as a measure of forecasting performance. Using ratios of RMSEs in D_j normalizes results so that RMSEs of individual series do not dominate performance measurement. $D_j > 1$ indicates that the system is outperforming in an overall sense the benchmark model in forecast period j. We report D_j separately for both the naïve and AR(12) benchmarks. In addition, in tests of forecasts generated by variations of the lasso system reported below, we report D_j using the forecasts of the CTA system as the benchmark.

Empirical Results

Exhibit 3 presents the D_j where we compare the forecasting performance of the CTA system to both the naïve and AR(12) benchmark models for forecasts of one to six months for the three databases. With the exception of the AR(12) benchmark for Canada in forecast period 1, all values of D_j in the exhibit are greater than 1, which indicates that the CTA system outperforms both the naïve and AR benchmarks.

Exhibit 3
The CTA System

Months ahead	1	2	3	4	5	6
U.S. database:						
Naïve benchmark D_j	1.0838	1.1036	1.1216	1.1398	1.1504	1.1611
AR(12) benchmark D_j	1.0104	1.0157	1.0206	1.0245	1.0288	1.0315
France–Germany database:						
Naïve benchmark D_j	1.2520	1.2862	1.2889	1.2857	1.2430	1.1585
AR(12) benchmark D_j	1.0492	1.1375	1.0901	1.0515	1.0534	1.0534
Canada database:						
Naïve benchmark D_j	1.2412	1.2843	1.3405	1.3601	1.3453	1.3697
AR(12) benchmark D_j	0.9948	1.0023	1.0104	1.0042	1.0017	1.0022

Exhibit 4 shows the performance results for the lasso system. Here, the performance of the lasso is presented without the benefit of BMA which makes comparisons between its performance and that of the CTA system, which does employ BMA, ambiguous. Rather, we present the results to establish a starting point for the later application of BMA to the lasso. In this context, comparing the D_j in Exhibit 4 to those in Exhibit 3, we see that the CTA system, which benefits from BMA, outperforms the lasso, which in Exhibit 4 does not benefit from BMA, in each forecast period 1 to 6 for both the naïve and AR(12) benchmarks for all three databases with the single exception of the naïve benchmark in forecast period 1 for the Canada database. Exhibit 4 also shows the performance of the lasso using the CTA system itself as the benchmark, and the D_j for that comparison are all less than 1, another indication that the lasso without BMA underperforms the CTA system.

Exhibit 4 also shows the p-values for the Wilcoxon signed rank test with continuity correction that compares the performance results of the two systems statistically. Briefly, the test is nonparametric and compares the ratios RB_{ij}/RS_{ij} series by series in each forecasting period j for the lasso and the CTA system. For instance, the p-values listed under the heading "naïve benchmark D_j" in the table are computed from the RB_{ij}/RS_{ij}^{lasso} and RB_{ij}/RS_{ij}^{CTA} compared series by series where RB_{ij} are the RMSEs of the naïve benchmarks for series i in period j, RS_{ij}^{lasso} are the lasso-generated RMSEs, and RS_{ij}^{CTA} are the CTA RMSEs. The p-values listed under the heading "CTA benchmark D_j" compare the RB_{ij}/RS_{ij}^{lasso} from the lasso system series by series to $RB_{ij}/RS_{ij}^{CTA} = 1$ for the CTA system since the CTA RMSEs appear in both the numerator of the ratio as the benchmark RB_{ij} and in the denominator as the RS_{ij}^{CTA}, i.e., the RMSEs of the CTA system.

The Wilcoxon test is an appropriate test because 1) the distributions of the RB_{ij}/RS_{ij} are bounded on the left at zero, which argues for a nonparametric test and 2) the proper comparison is between forecasts from pairs of equations for a given series. The p-values are interpreted in the usual sense that a p-value less than a critical value indicates a significant difference in performance between the two systems.

For the U.S. and the France–Germany databases, the p-values in Exhibit 4 are each less than a critical value of 0.05—in fact, in the exhibit, they are each zero to the four decimal places shown—indicating that the CTA system is statistically significantly better than the lasso system. Using the Canada database, forecast performance results are significantly different in most forecast periods with some exceptions in forecast periods 2, 3, and 6.

Exhibit 4
The Lasso

Months ahead	1	2	3	4	5	6
U.S. database:						
Naïve benchmark D_j	1.0361	1.0599	1.0722	1.0843	1.0950	1.1048
p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
AR(12) benchmark D_j	0.9755	0.9854	0.9891	0.9916	0.9970	0.9999
p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
CTA benchmark D_j	0.9654	0.9695	0.9687	0.9674	0.9691	0.9694
p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
France–Germany database:						
Naïve benchmark D_j	1.1500	1.1643	1.1463	1.1507	1.1177	1.0615
p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
AR(12) benchmark D_j	0.9721	1.0466	0.9748	0.9522	0.9536	0.9640
p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
CTA benchmark D_j	0.9321	0.9161	0.8986	0.9098	0.9077	0.9173
p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Canada database:						
Naïve benchmark D_j	1.2566	1.2795	1.3035	1.3105	1.2877	1.2911
p-value	0.0003	0.0489	0.0551	0.1177	0.1565	0.0940
AR(12) benchmark D_j	0.9620	0.9763	0.9796	0.9775	0.9771	0.9655
p-value	0.0007	0.0900	0.0985	0.1679	0.3010	0.0784
CTA benchmark D_j	0.9720	0.9774	0.9724	0.9745	0.9747	0.9619
p-value	0.0007	0.0956	0.0985	0.1816	0.2910	0.0628

The performance of the lasso can be improved by employing BMA to combine its forecasts with those of the two autoregressive schemes. Exhibit 5 shows the results. According to the D_j reported in the exhibit, using U.S. data, this system outperforms the CTA system in absolute terms for the naïve and AR(12) benchmarks in each forecast period 1 to 6. With U.S. data, the p-values in the exhibit indicate that the performance of the lasso with BMA is statistically significantly better than that of the CTA system in forecast periods 1 and 2 for all three benchmarks, but performance results are not significantly different in periods 3 to 6. Using the France–Germany database, the performance of the lasso with BMA is worse in absolute terms than that of the CTA system in all periods for the naïve and AR(12) benchmarks. Compared to the CTA benchmark, the comparisons are mixed. The p-values are greater than a critical value of 0.05 in all forecast periods for all benchmarks, indicating that the lasso with BMA does not significantly outperform any benchmark. Using the Canada database, the performance of the lasso with BMA is better in each forecast period for all benchmarks, but the only significant result is for the AR(12) and CTA benchmarks in forecast period 5.

Overall, then, comparing the performance of the CTA system, which is designed implicitly to use BMA, with that of the lasso combined using BMA with the two autoregressive schemes, the outcome is mixed across the three databases and three benchmarks.

Exhibit 5
BMA of the Lasso, AR(12), and AR(1) with appended seasonal binary variables

Months ahead	1	2	3	4	5	6
U.S. database:						
Naïve benchmark D_j	1.0934	1.1127	1.1262	1.1394	1.1495	1.1600
p-value	0.0006	0.0321	0.5797	0.5910	0.5271	0.5563
AR(12) benchmark D_j	1.0197	1.0228	1.0235	1.0232	1.0268	1.0300
p-value	0.0002	0.0230	0.4916	0.7143	0.6016	0.6432
CTA benchmark D_j	1.0118	1.0092	1.0052	1.0013	1.0011	1.0023
p-value	0.0001	0.0105	0.3241	0.9105	0.7451	0.8146
France–Germany database:						
Naïve benchmark D_j	1.2437	1.2629	1.2577	1.2618	1.2296	1.1555
p-value	0.6684	0.0559	0.2914	0.3969	0.5824	0.8806
AR(12) benchmark D_j	1.0474	1.1190	1.0747	1.0417	1.0451	1.0489
p-value	0.9745	0.0866	0.3506	0.4410	0.8254	0.7541
CTA benchmark D_j	1.0012	0.9901	0.9875	0.9942	0.9966	1.0000
p-value	0.8629	0.1330	0.3951	0.5846	0.9974	0.6225
Canada database:						
Naïve benchmark D_j	1.2991	1.3269	1.3620	1.3773	1.3613	1.3788
p-value	0.2776	0.4374	0.9442	0.2976	0.1460	0.2585
AR(12) benchmark D_j	1.0034	1.0109	1.0139	1.0124	1.0170	1.0101
p-value	0.1839	0.1936	0.5655	0.1448	0.0379	0.2114
CTA benchmark D_j	1.0132	1.0136	1.0095	1.0125	1.0184	1.0102
p-value	0.1315	0.1792	0.5273	0.1315	0.0399	0.1701

We can achieve better performance. Exhibit 6 shows the performance results where we use BMA in a two-step process. First, we use BMA to combine the forecasts from the CTA equations from the initial phase of construction with those of the lasso. Then, we use BMA to combine those forecasts with those of the AR(12) equations and the equations of the AR(1) with appended seasonal binaries. The exhibit shows the D_j for this two-step BMA system and the p-values for the Wilcoxon test where again we compare the performance of the two-step BMA system with that of the CTA system. Using U.S. data, the performance of the two-step BMA system is better than that of the CTA system both in absolute terms and statistically for all benchmarks in each forecast period 1 to 6. Using the France–Germany database, the performance results of the forecasting system produced by the two-step BMA procedure and the CTA system are remarkably similar in absolute terms across all forecast periods for all benchmarks. Statistically, the p-values are all greater than a critical value of 0.05 for all benchmarks in all periods except for the naïve and CTA benchmarks in period 6, indicating that the difference in performance in the CTA system and the two-step system is in general not statistically significant. Using the Canada database, the two-step BMA system outperforms the CTA system in absolute terms in each forecast period 1 to 6 for all benchmarks. The difference in performance is statistically significant in each forecast period for all benchmarks except in period 3.

Exhibit 6
Two-step BMA of the CTA System (initial phase), the Lasso, AR(12), and AR(1) with appended seasonal binary variables

Months ahead	1	2	3	4	5	6
U.S. database:						
Naïve benchmark D_j	1.0985	1.1175	1.1340	1.1497	1.1607	1.1715
p-value	0.0000	0.0000	0.0000	0.0080	0.0189	0.0235
AR(12) benchmark D_j	1.0229	1.0261	1.0293	1.0308	1.0355	1.0385
p-value	0.0000	0.0000	0.0000	0.0077	0.0180	0.0258
CTA benchmark D_j	1.0149	1.0127	1.0111	1.0088	1.0093	1.0096
p-value	0.0000	0.0000	0.0000	0.0033	0.0078	0.0134
France–Germany database:						
Naïve benchmark D_j	1.2572	1.2857	1.2839	1.2867	1.2475	1.1692
p-value	0.2220	0.6754	0.9872	0.5268	0.4564	0.0497
AR(12) benchmark D_j	1.0550	1.1316	1.0894	1.0557	1.0559	1.0593
p-value	0.1114	0.9109	0.8254	0.6180	0.3707	0.0512
CTA benchmark D_j	1.0071	1.0016	1.0003	1.0057	1.0044	1.0077
p-value	0.0670	0.9694	0.8279	0.5184	0.2828	0.0440
Canada database:						
Naïve benchmark D_j	1.2789	1.3182	1.3635	1.3802	1.3622	1.3818
p-value	0.0103	0.0093	0.2176	0.0296	0.0078	0.0425
AR(12) benchmark D_j	1.0051	1.0136	1.0185	1.0150	1.0168	1.0130
p-value	0.0029	0.0017	0.0942	0.0074	0.0030	0.0269
CTA benchmark D_j	1.0130	1.0143	1.0116	1.0134	1.0171	1.0124
p-value	0.0020	0.0015	0.0900	0.0074	0.0027	0.0203

Superior performance can also be achieved by substituting the forecasts of the elastic-net operator described above for those of the lasso in the two-step BMA scheme. Exhibit 7 shows the performance results where we do this, i.e., where we use BMA to combine the forecasts from the CTA equations from the initial phase of construction with those of the elastic net and then use BMA to combine those forecasts with those of the two autoregressive schemes. The D_j in Exhibit 7, where we use the elastic net, are virtually the same in value item for item across databases, benchmarks, and forecast periods as those in Exhibit 6, where we use the lasso in the two-step BMA process. The p-values differ somewhat item by item between the two exhibits because they take into account the dispersion of the ratios RB_{ij}/RS_{ij} series by series. Still, in terms of their significance relative to a critical value of 0.05, there is virtually no difference in the results in the two exhibits. So, with our data, it makes no practical difference whether we employ the lasso or the elastic net in the two-step BMA scheme.

Exhibit 7
Two-step BMA of the CTA System (initial phase), the Elastic Net, AR(12), and AR(1) with appended seasonal binary variables

Months ahead	1	2	3	4	5	6
U.S. database:						
Naïve benchmark D_j	1.0980	1.1170	1.1335	1.1490	1.1600	1.1709
p-value	0.0000	0.0000	0.0000	0.0144	0.0305	0.0350
AR(12) benchmark D_j	1.0225	1.0256	1.0288	1.0302	1.0349	1.0380
p-value	0.0000	0.0000	0.0000	0.0133	0.0280	0.0390
CTA benchmark D_j	1.0146	1.0123	1.0107	1.0083	1.0089	1.0093
p-value	0.0000	0.0000	0.0000	0.0062	0.0133	0.0208
France–Germany database:						
Naïve benchmark D_j	1.2554	1.2821	1.2795	1.2821	1.2441	1.1666
p-value	0.3707	0.3934	0.6661	0.7711	0.6453	0.0689
AR(12) benchmark D_j	1.0537	1.1278	1.0869	1.0530	1.0539	1.0573
p-value	0.1988	0.5737	0.7420	0.9109	0.6001	0.0665
CTA benchmark D_j	1.0055	0.9987	0.9974	1.0025	1.0020	1.0052
p-value	0.1673	0.6430	0.7565	0.8254	0.5102	0.0559
Canada database:						
Naïve benchmark D_j	1.2786	1.3184	1.3638	1.3780	1.3621	1.3820
p-value	0.0095	0.0047	0.2403	0.0291	0.0075	0.0396
AR(12) benchmark D_j	1.0049	1.0137	1.0186	1.0150	1.0168	1.0133
p-value	0.0031	0.0012	0.1092	0.0080	0.0027	0.0224
CTA benchmark D_j	1.0128	1.0144	1.0117	1.0134	1.0171	1.0125
p-value	0.0021	0.0011	0.0942	0.0075	0.0028	0.0181

SUMMARY AND CONCLUSIONS

Absent the need for human expertise, automatic systems can build forecasting models for very large datasets, and they can be employed to generate new ideas to improve existing models. Automatic systems offer policymakers the potential for improved policy analysis. Where forecasting models are based on economic theory, forecasting is complicated by the fact that 1) the structure of the economy changes over time complicating the formulation of economic theory, 2) economic theory is a highly simplified representation of reality which may ignore information useful in making forecasts, and 3) available data series used in building forecasting models are often poor representations of theoretical variables. Automatic systems seek to address these issues by eliminating economic theory, appealing instead directly to the data. The CTA system examined here takes that idea one step further, dispensing with the idea of basing an automatic model-building system on a theory of model selection and instead employing an empirical approach to constructing the automatic system.

The CTA system outperforms the lasso in absolute terms, i.e., as measured by the D_j, for all benchmarks in each forecast period in each of the three datasets. Using both the U.S. database and the France–Germany database, performance results are significantly different, as measured by the p-values for the Wilcoxon signed rank test, in each forecast period for all benchmarks. Using the Canada database, forecast performance results are significantly different in most forecast periods.

Using BMA to combine the lasso with the two autoregressive schemes, the combined model generally performs better than the lasso in outright form. Using both the U.S. and Canada databases, the lasso combined with the autoregressive schemes outperforms the CTA system in absolute terms in each forecast period for all benchmarks, although performance is generally not statistically significantly different in terms of the p-values for either database. Using the France–Germany database, the lasso combined with the two autoregressive schemes underperforms the CTA system, but again the difference in performance is generally not statistically significant.

Better results are obtained using BMA in a two-step process to combine the CTA system, the lasso, and the two autoregressive schemes. Using U.S. data, this system outperforms the CTA system both absolutely and statistically in each forecast period for all benchmarks. Using Canada data, the outcome is the same except that performance is not significantly different in forecast period 3 for any benchmark. Using the France–Germany database, this system produces performance results which are similar to that of the CTA system in absolute terms, and p-values generally exceed a critical value of 0.05. Substituting forecasts from the elastic net for those of the lasso in the two-step BMA scheme produces forecasting performance results which are practically the same.

APPENDIX

Series in the Databases

Exhibit A
Types of Series in the Databases

	U.S.	France–Germany	Canada
Total series	434	105	61
Stock indexes and stock market indicators	12	1	1
Short- and long-term interest rates	9	9	7
Currency and money supply	3	5	3
Exchange rate indexes	1	1	1
Personal consumption expenditures current dollars	16		
Personal consumption expenditures constant dollars	16	1	
Retail sales	14	5	3
Retail inventories	6		
Wholesale trade	18	2	1
Wholesale inventories	18		
Manufacturers' shipments/sales	21	3	3
Manufacturers' orders	7	1	1
Manufacturers' inventories	21		1
Personal income	7		
Consumer price indexes	11	26	7
Producer and industrial price indexes	10	7	2
Commodity and metals prices	4		
Oil prices	2		
Industrial production indexes	41	23	4
Capacity utilization indexes	3		
Nonfarm employment	49		15
Agricultural employment			1
Workweek hours	32		
Hourly earnings	31	1	1
Unemployment and labor force	10	2	2
Building permits, housing starts, and completions	10	3	2
New home sales	4		
New home prices	3		1
Construction spending	3		
Foreign trade: goods and services	15	3	3
Motor vehicle unit sales and production	6	3	2
Consumer sentiment surveys	4	4	
Business sentiment surveys	24	2	
Government outlays	3		
Business and personal loans		3	

REFERENCES

Campos, Julia, Neil R. Ericsson, and David Hendry (2005), “General-to-Specific Modeling: An Overview and Selected Bibliography,” International Finance Discussion Paper No. 838 (Washington, D.C., Federal Reserve Board, August).

Chen, Jiaqi, and Michael L. Tindall (2016), “The Structure of a Machine-Built Global Macro Forecasting System,” Occasional Paper no. 13-01, (Federal Reserve Bank of Dallas, April 2013, rev. January 2016), and *Journal of Alternative Investments* (forthcoming).

De Mol, Christine, Domenico Giannone, and Lucrezia Reichlin (2006), “Forecasting Using a Large Number of Predictors: Is Bayesian Regression a Valid Alternative to Principal Components?” European Central Bank Working Paper Series No. 700 (Frankfurt, Germany, December), and *Journal of Econometrics* (2008), 146 (2): 318–28.

Diebold, Francis X., and Canlin Li (2006), “Forecasting the Term Structure of Government Bond Yields,” *Journal of Econometrics,* 130 (2): 337–64.

Elliott, Graham, and Allan Timmermann (2013), *Handbook of Economic Forecasting, Volume 2* (Amsterdam: North Holland).

Ericsson, Neil R. (2016a), “Eliciting GDP Forecasts from the FOMC’s Minutes Around the Financial Crisis,” *International Journal of Forecasting*, 32 (forthcoming).

——— (2016b), “Testing for and Estimating Structural Breaks and Other Nonlinearities in a Dynamic Monetary Sector," International Finance Discussion Paper (Washington, D.C., Federal Reserve Board), in preparation, and *Studies in Nonlinear Dynamics & Econometrics* (forthcoming).

Ericsson, Neil R., David F. Hendry, and Stedman B. Hood (2016), “Milton Friedman as an Empirical Modeler,” International Finance Discussion Paper (Washington, D.C., Federal Reserve Board), in preparation, and Cord, Robert A., and J. Daniel Hammond, eds., (March 2016), *Milton Friedman: Contributions to Economics and Public Policy* (Oxford: Oxford University Press)

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman (2009), *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (New York: Springer).

Jeffreys, Harold, and Bertha S. Jeffreys (1988), *Methods of Mathematical Physics, Third Edition* (Cambridge: Cambridge University Press).

Kapinos, Pavel S., and Oscar A. Mitnik (2015), “A Top-down Approach to Stress-testing Banks,” Federal Deposit Insurance Corporation Working Paper Series, FDIC Center for Financial Research Working Paper No. 2015-02 (Washington, D.C., FDIC, March), and *Journal of Financial Services Research* (forthcoming).

Stock, James H., and Mark W. Watson (2003), “Forecasting Output and Inflation: The Role of Asset Prices,” *Journal of Economic Literature*, 41 (3): 788–829.

——— (2008), “Phillips Curve Inflation Forecasts” (paper presented at the Federal Reserve Bank of Boston’s 53rd Annual Economic Conference, “Understanding Inflation and the Implications for Monetary Policy: A Phillips Curve Retrospective,” Chatham, Massachusetts, June 9–11).

Tibshirani, Robert (1996), “Regression Shrinkage and Selection Via the Lasso,” *Journal of the Royal Statistical Society, Series B*, 58 (1): 267–88.

Zou, Hui and Trevor Hastie (2005), "Regularization and Variable Selection via the Elastic Net," *Journal of the Royal Statistical Society, Series B*, 67 (2): 301–320.