BUSINESS SCHOOL
Te Kura Pakihi

ISSN 1178-2293 (Online)

University of Otago
Economics Discussion Papers
No. 1405

March 2014

Information Processing, Pattern Transmission and Aggregate Consumption Patterns in New Zealand

Daniel Farhat
University of Otago – Department of Economics

Address for correspondence:

Dan Farhat
Department of Economics
University of Otago
PO Box 56
Dunedin
NEW ZEALAND
Email: dan.farhat@otago.ac.nz
Telephone: 64 3 479 8645

Information Processing, Pattern Transmission and Aggregate Consumption Patterns in New Zealand

[DRAFT: 21 MARCH 2014]

Dr. Daniel Farhat
Department of Economics
University of Otago
PO Box 56, Dunedin, 9054
Dan.farhat@otago.ac.nz
(Tel) 64 3 479 8645

ABSTRACT. This study explores the value of information transmission in training heterogeneous Artificial Neural Network (ANN) models to identify patterns in the growth rate of aggregate per-capita consumption spending in New Zealand. A tier structure is used to model how information passes from one ANN to another. A group of 'tier 1' ANNs are first trained to identify consumption patterns using economic data. ANNs in subsequent tiers are also trained to identify consumption patterns, but they use the patterns constructed by ANNs trained in the preceding tier (secondary information) as inputs. The model's results suggest that it is possible for ANNs downstream to outperform ANNs trained using empirical data directly on average. This result, however, varies from time period to time period. Increasing access to secondary information is shown to increase the similarity of heterogeneous predictions by ANNs in lower tiers, but not substantially affect average accuracy.

KEYWORDS. Artificial neural networks, aggregate consumption patterns, information transmission.

JEL. C45, C63, E27.

1. INTRODUCTION

Artificial neural networks (ANNs) are mathematical algorithms that transform input data into output data. Because they are flexible in structure and can incorporate a high degree of non-linearity, they can be trained to identify very complex relationships in data. As such, they prove to be useful for solving a variety of pattern recognition, optimization and forecasting problems in economics and finance.[1]

When it comes to reproducing aggregate consumption patterns, the value of ANN models has been studied by Church and Curram (1996) and Farhat (2012). Church and Curram show that although ANNs are flexible in nature and do not require a large number of data points to produce accurate patterns, they do not outperform standard econometric methods significantly. Farhat (2012) shows that pattern reproduction can be improved substantially by training heterogeneous ANNs (done by using different combinations of in-sample data in the training process), then forming a weighted average of the outputs. Several methods for computing this weighted average are compared with the intention of informing future systemic simulation models[2] as to the right ‘social rules’ for reproducing consumption patterns in an artificial economy.

This study expands on Farhat (2012) by exploring the value of adding information transmission from some heterogeneous ANN models to other heterogeneous ANN models. In the grand model presented below, a group of ANNs use economic data directly to reproduce patterns in aggregate per-capita consumption in New Zealand. These ‘tier 1’ ANNs can be thought of as ‘primary data processors’: the patterns they produce are based entirely on empirical properties of the economy. Next, another group of ANNs are trained to reproduce aggregate per-capita consumption patterns. Instead of using empirical data, however, they use the patterns generated by a sample of ANNs from tier 1. These ‘tier 2’ ANNs can be thought of as ‘secondary data processors’: they recycle the patterns produced by others to form their own predictions. This tier structure continues (tier 3, tier 4, …) with the ANNs in any tier d using a sample of estimates from the ANNs in the previous tier, $d-1$, to produce their own estimates.

Three questions arise. First, are the patterns produced by ‘lower’ tier ANNs less accurate on average than those produced by tier 1 ANNs? (In other words, does using secondary information lead to wrong-minded predictions?) Second, do lower tier ANNs produce a greater diversity of patterns than those produced by tier 1 ANNs? (In other words, does using secondary information lead to confusion or cohesion?) Third, can an ANN in any tier produce more accurate patterns by relying on a greater amount of information from the proceeding tier? (In other words, is having more secondary information better to having less?)

Results suggest that lower tier ANNs are not always less accurate than tier 1 ANNs. This result, however, varies depending on the period being forecasted. Interestingly, the average performance of ANNs worsens as information is transmitted downstream, but then subsequently improves in many cases. The grand model also shows that increasing the number of secondary sources used by ANNs in lower tiers reduces the variability of patterns produced. However, using more secondary sources does not substantially improve accuracy.

In the remainder of this article, the structure of the artificial neural networks used in this project is first described. The tier structure defining how information is transmitted between ANNs is then presented. Simulation results are then shown and subsequently discussed. The article concludes with the model’s main implications and directions for future research.

2. METHOD

2.1. Artificial Neural Networks

Beltratti et al. (1996), Jain and Mao (1996), Warner and Misra (1996), Cooper (1999), Gonzalez (2000), and Detienne et al. (2003) provide easy-to-follow introductory guides on ANN algorithms. While these algorithms can become very complex, the ANN used in this study is a rather common and simple *feed-forward* neural network (see fig. 1). In this basic model, K different pieces of input data (X_k, $k = 1...K$) are weighted, summed together with a ‘bias term’[3] and then transformed by H separate functions (known as *activation functions*). The results from these transformations are themselves weighted and summed with another bias term to produce an output, $\widehat{C}$. In the ANN shown in fig. 1, ω_{kh} denotes the weight placed on input k in activation function h,

ω_{0h} is the bias term in activation function h, γ_h is the weight placed on the result from activation function h, and γ_0 is the bias term for the activation functions in computing the final output ($k = 1...K$ and $h = 1...H$). A compact, mathematical representation of the ANN in fig. 1 is:

$$n_h = \omega_{0h} + \sum_{k=1}^{K} \omega_{kh} X_k \quad (1)$$

$$N_h = A(n_h) \quad (2)$$

$$\widehat{C} = \gamma_0 + \sum_{h=1}^{2} \gamma_h N_h \quad (3)$$

[Figure 1 here]

We can create larger ANNs by adding additional hidden layers (where the output, $\widehat{C}$, is passed on as an input to other neurons) if we wish, however the simple structure described by fig. 1 will be quite sufficient (and computationally efficient) for this study.[4] Kuan and White (1994) note that passing weighted inputs directly to the construction of the output with no transformation makes the model above more relatable to many standard linear econometric analyses. This will be done in this study via the first activation function, A_1.[5] The other activation functions will be hyperbolic tangent (tanh) sigmoid functions:

$$A(n_h) = \frac{e^{n_h} - e^{-n_h}}{e^{n_h} + e^{-n_h}} \quad (4)$$

This sort of function is useful to use when the inputs (and outputs) can be either positive or negative. N_h will be bounded by $\pm$ 1.

'Training' the ANN to produce acceptable outputs involves determining appropriate values for the weights (ω's and γ's). To do this, an algorithm similar to that described by Aminian et al. (2006) is employed. First, all available data is divided into three sets: a training set, a validation set, and a forecasting set. The training set and the validation set together form the in-sample data used to derive the weights. The forecast set is used to test the out-of-sample performance of the ANN. Note that in the case of time-series data, these sets need not be time-ordered (i.e. we can randomly select time periods into each set). This is done for the training and validation sets: the in-sample data is randomly allotted with approximately 70% allocated to the training set and the remaining 30% to the validation set.[6] The forecast set will always consist of the final periods in the available data (here, the final 24 periods are used).

Next, a back-propagation (BP) algorithm is implemented. In this algorithm, data from the training set is fed into the ANN to produce outputs. A measure related to the sum of squared errors over the training set is computed. The model weights are then updated so as to minimize this measure. This process then repeats. If the algorithm runs ad infinitum, the ANN will start to memorize the training set patterns (known as 'overfitting') and out-of-sample forecasting will be poor. To prevent this, data from the validation set is also fed through the ANN to produce outputs and a measure related to the sum of squared errors of the validation set is computed as the BP algorithm runs. The BP algorithm is forced to quit updating weights when the sum of squared errors for the validation set starts to rise. Out-of-sample prediction is thus improved.[7] Once the weights are found, the forecast set data can be fed through the ANN to test the model's performance.

In this study, large batches of identically-structured ANNs are trained using BP. These ANNs produce heterogeneous outputs thanks to the random allocation of the in-sample data into the training and validation sets.

2.2. Information Transmission and ANN Training

The tier structure used to model information transmission in this study is as follows. Denote D as the total number of tiers. In the first tier ($d = 1$), a population of J ANNs use empirical data from the New Zealand economy to predict patterns in the growth rate of final private consumption expenditures per worker (c_t). Members of this tier can be thought of as 'experts' who base their understanding of consumption spending on empirical analysis of the economy directly. The data they use as inputs include the growth rate of final private consumption expenditures per worker in the previous period (c_{t-1}), the growth rate of GDP per worker (y_t), the point change in the unemployment rate (u_t), the point change in the money market interest rate (r_t), the CPI inflation rate (p_t) and the point change in the nominal effective exchange rate (q_t). These variables are commonly seen in both theoretical and empirical business cycle studies in relation to the household sector. All data pertains to the New Zealand economy, 1992Q2 – 2011Q3, and is re-scaled before use.[8] Data for interest rates and exchange rates are sourced from the International Monetary Fund (2011) while all other data are sourced from the OECD (2011). The out-of-sample data is fixed to be the last 24 periods (2005Q4 – 2011Q3) of this data set (for this and all other tiers). After each of the J ANNs in the first tier is trained using the in-sample data, the inputs from both the in-sample and out-of-sample data are re-fed through the network to produce a sequence of estimates for c_t which cover the entire time span.

In subsequent tiers ($d = 2…D$), populations of J ANNs are once again trained to form predictions for c_t. In these tiers, however, the ANNs do not use the empirical data directly as inputs. Instead, they collect predictions for c_t generated by M randomly-selected ANNs trained in the tier that came directly before them. All in-sample data and is randomly divided between the training and validation set as described above. Once the ANN is trained, the inputs from both the in-sample and out-of-sample data sets are again fed through the ANN to produce a sequence of estimates for c_t covering the entire time span. ANNs in the next tier may then call upon this information when producing their own patterns. Members of these 'lower' tiers can be thought of as users of secondary information – they base their understanding of c_t on predictions made by others. As a result of this structure, ANNs trained in more distant tiers are reliant on highly-processed information compared to ANNs trained in higher tiers.

3. RESULTS

In the simulation results that follow, the following calibrations are used:

H = number of activation functions = 3. A_1 makes no non-linear transformations; A_2 and A_3 are tanh sigmoid functions. Farhat (2012) suggests that a low number of non-linear activation functions is sufficient for generating significant improvements in pattern production for c_t over a linear benchmark model (albeit over a shorter forecast set).

D = number of tiers = 12. While a larger number of tiers can be explored, this value provides a sufficient starting point for observing the impact of transmitting estimates 'down-stream'.

M = number of randomly-selected input patterns from the previous tier for tiers $d = 2...D$ = 2, 5 and 10 (experimented with separately).

J = number of ANNs trained in each tier = 500. This number is much lower (per tier) than Farhat (2012), it provides a useful starting point for looking at the impact of heterogeneously processed information on pattern production. If J is too large relative to M, the patterns produced by any single tier 1 ANN would have little impact on the subsequent tiers (as the probability it's patterns are used by a tier 2 ANN are low).

How well an ANN replicates out-of-sample patterns is measured by the error between its estimate for the output variable and the empirical value of the output variable in each period of the forecast set. Since the data for c_t is re-scaled when used in the ANN training process, these errors are also 'adjusted' measures (i.e. measured in terms of number of empirical standard deviations from the empirical mean of c_t).

3.1. Experiment 1: Tier 1 ANN Performance

Fig. 2 shows how a batch of tier 1 ANNs performs at out-of-sample pattern production for each period in the forecast set. The average scaled errors, shown by bars in the figure, represent overall accuracy. In several periods, patterns are produced quite accurately on average (see 2006Q2, 2007Q1 – 2007Q3 and 2009Q4). In other periods the model has difficulty reproducing c_t (see 2008Q4 as the most obvious example). The spread of predictions across tier 1 ANNs is shown by adding ± 2 standard deviation points to the figure (shown as a dotted line). This serves as an illustration of cohesion (or dissention) amongst ANNs. Several periods indicate a modest degree of cohesion (for example, 2006Q3 and 2010Q1 – 2010Q3) while other periods showed a high degree of variation (for example, 2006Q2 and 2010Q4). The longest period with intense variability (2008Q2 – 2009Q1) coincides with the Great Recession reaching New Zealand, perhaps reflecting the ability of the ANN approach described in this study to identify periods of confusion and uncertainty. These results are consistent with those in Farhat (2012).

[Figure 2 here]

3.2. Experiment 2: Information Transmission with M = 5

The analysis is now extended to lower tiers. Figs. 3-4 show how the ANNs trained in each tier perform when input data is collected from 5 randomly-selected sources from the preceding tier (except for tier 1 which uses primary data). The results are mixed. For some time periods, lower tier ANNs perform worse on average than higher tier ANNs (see 2007Q1 for the most obvious example). There are several cases, however, where average accuracy increases as estimates are transmitted downstream. When this occurs, it is common for average errors to worsen before they improve as patterns are passed from one tier to another (2006Q2 is a clear example of this). For most time periods, lower tier ANNs generate a larger variety of estimates than tier 1 ANNs (2010Q4 is an extreme instance). This suggests that distortion occurs and is exacerbated as information is passed.

[Figure 3 & 4 here]

3.3. Experiment 3: Information Transmission with M = 2

Figs. 5-6 show how ANNs trained in the 'lower' tiers perform when input data is collected from only 2 randomly-selected ANNs from the preceding tier. Although the ANNs are using fewer sources of information, the general trends observed when $M = 5$ continue to hold: for some periods lower tier ANNs perform worse on average than tier 1 ANNs while in other periods they perform better; for most periods there is the spread of errors is greater in lower tiers.

When we carefully compare figs. 5-6 to figs. 3-4, we see that reducing the amount of secondary information used has had an impact in only a few time periods. Accuracy has been noticeably reduced in 2006Q2, 2007Q2, 2008Q2, 2008Q4 – 2009Q2, whereas accuracy has increased in 2007Q1, 2009Q3 and 2010Q4 – 2011Q1. Reducing the number of input patterns has noticeably increased variation in periods 2006Q2, 2007Q1 – 2007Q2, and 2008Q4-2009Q1. Notably, variation has been reduced in 2006Q1, 2009Q3, 2010Q2 and, most obviously, in 2010Q4.

[Figure 5 & 6 here]

3.4. Experiment 4: Information Transmission with M = 10

Figs. 7-8 illustrate performance when lower tiers collect input data from are larger number of randomly-selected sources from the preceding tier (M =10). In this case, there is a general tendency for lower tier ANNs to be more accurate than tier 1 ANNs. As with $M = 5$, things tend to worsen before they improve as we move down the tiers. Examples of this trend appear in 2006Q3, 2007Q3, 2008Q1 – 2008Q3, and 2011Q1. There are many instances where lower tier ANNs generate less accurate results, but only mildly so (2006Q1 – 2006Q2, 2008Q4 – 2009Q1, and 2010Q1 – 2010Q3 are examples of this).

When we carefully compare figs. 7-8 to figs. 3-4, we see that increasing the amount of secondary information used to construct patterns for c_t has increased accuracy in 2006Q1, 2007Q1, 2007Q4, 2009Q3 and 2010Q4 – 2011Q1 for ANNs in lower tiers. Further, variation has been reduced in 2006Q1 – 2006Q2, 2007Q4 – 2009Q3, 2010Q1 – 2011Q1, and 2011Q3 (particularly amongst ANNs in lower tiers). As one might expect, this result indicates that consensus tends to occur as ANNs rely on a larger amount of similar information to produce patterns.

[Figure 7 & 8 here]

4. DISCUSSION

The questions posed above can now be addressed. First, are the patterns produced by lower tier ANNs less accurate on average than those produced by tier 1 ANNs? The answer varies from time period to time period in the forecast set and how far downstream we go. There are many cases when lower tier ANNs exhibit increased accuracy compared to tier 1 ANNs. However, it should also be noted that not *all* lower tier ANNs outperform higher tier ANNs. There are several cases where average performance first worsens as predicted patterns are transmitted down the tier structure, but then begins to improve for more distant tiers. In these cases, using second-hand information does not lead to wrong-minded predictions for those extremely detached from the empirical data.

Second, do lower tier ANNs vary greatly in their predictions compared to tier 1 ANNs? The answer to this depends on the number of sources a low-tier ANN relies on. When M is modest (here, 2 or 5), it is most common for the diversity of predictions to rise downstream (note this is not uniformly the case, see 2011Q2 for example). When the number of sources is large (M = 10), lower tiers tend to produce relatively uniform predictions. It is likely that this effect is due to the dilution of errant sources and increased information overlap. Interestingly, the tier structure seems to amplify homogeneity of pattern prediction in lower tiers compared to tier 1 ANNs despite elements of randomness in the training process.

Finally, can an ANN in any tier produce more accurate patterns by relying on a greater amount of information from the proceeding tier? While increasing M from 5 to 10 has definitely reduced the *variety* of predictions produced in lower tiers as noted above, it has had relatively little impact on the *accuracy*. While there are some time periods where improvements in accuracy definitely occur for lower tier ANNs (see 2006Q1 and 2990Q3 as examples), these are few. For the most part, increasing the number of sources does not greatly enhance pattern production (nor does it severely diminish it for that matter). In other words, more secondary information is important for reducing the variability of estimates from any tier, and not the average accuracy of those predictions.

5. CONCLUSION

The study above relies on heterogeneous artificial neural networks to reproduce aggregate per-capita consumption growth in New Zealand. The ANNs are arranged in a tier structure which allows information to be transmitted from 'tier 1' ANNs (which are trained using empirical data directly) to 'lower' tier ANNs (which are trained using estimates from randomly-selected ANNs from the preceding tier). The grand model's main purpose is to identify how the use of secondary information affects the accuracy and variability of consumption patterns.

The study shows that lower tier ANNs are not always less accurate than tier 1 ANNs. While this result varies depending on the forecast period, it suggests that allowing an ANN model to source input information from an already-trained model will not necessarily affect accurately adversely. Interestingly, ANNs on much lower tiers can, in some cases, perform better than higher-tier ANNs. The model also shows that increasing the number of secondary sources used by ANNs trained in lower tiers leads to enhanced cohesion of produced patterns. However, there is little impact on accuracy when the amount of secondary information used rises. These results may change if we allow more complicated ANN structures or if we incorporate additional features to the tier structure which impact how information is shared downstream. Exploring these is left for future work.

With social and economic simulation becoming a more widely-used form of modeling, the results in the study here lead to relevant insights for models of the household sector. If consumers spend time on information processing and pattern recognition before they form expectations or make decisions within a simulation model, allowing them to share their results with each other may have benefits. Some agents can rely on secondary information to identify patterns in the economy, yet still make reasonable predictions compared to those 'experts' who rely on primary data. Further, how similar the estimates of these 'second-hand data users' become can be influenced by changing the number of sources (i.e. 'peers') they interact with. Incorporating these ideas into a full macroeconomic agent-based simulation framework is a worthwhile endeavor left for future research.

6. FIGURES

Fig. 1. A simple feed-forward artificial neural network

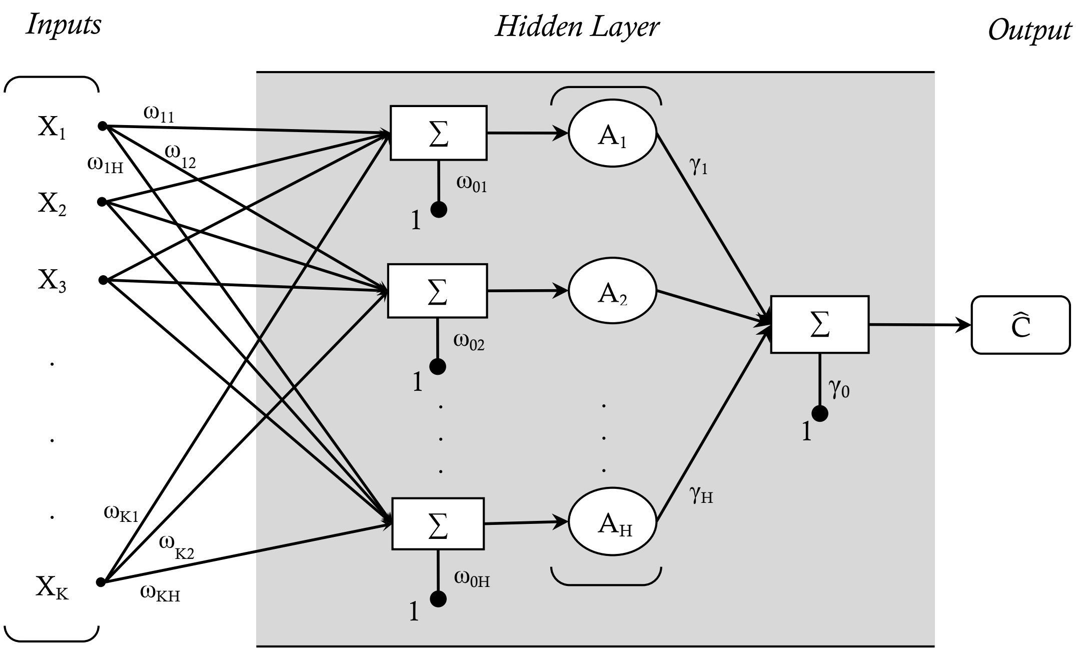


Fig. 2. Forecast set average errors [bars] with ± 2 standard deviations [lines] for tier 1 ANNs

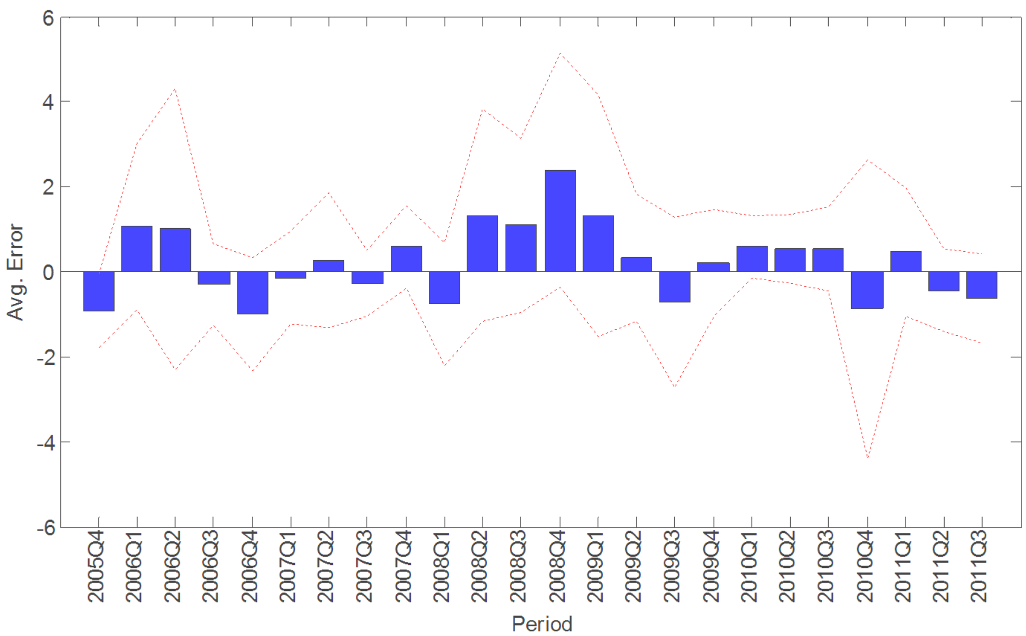


Fig. 3. Forecast set average errors [bars] with ± 2 standard deviations [lines] by tier when M = 5

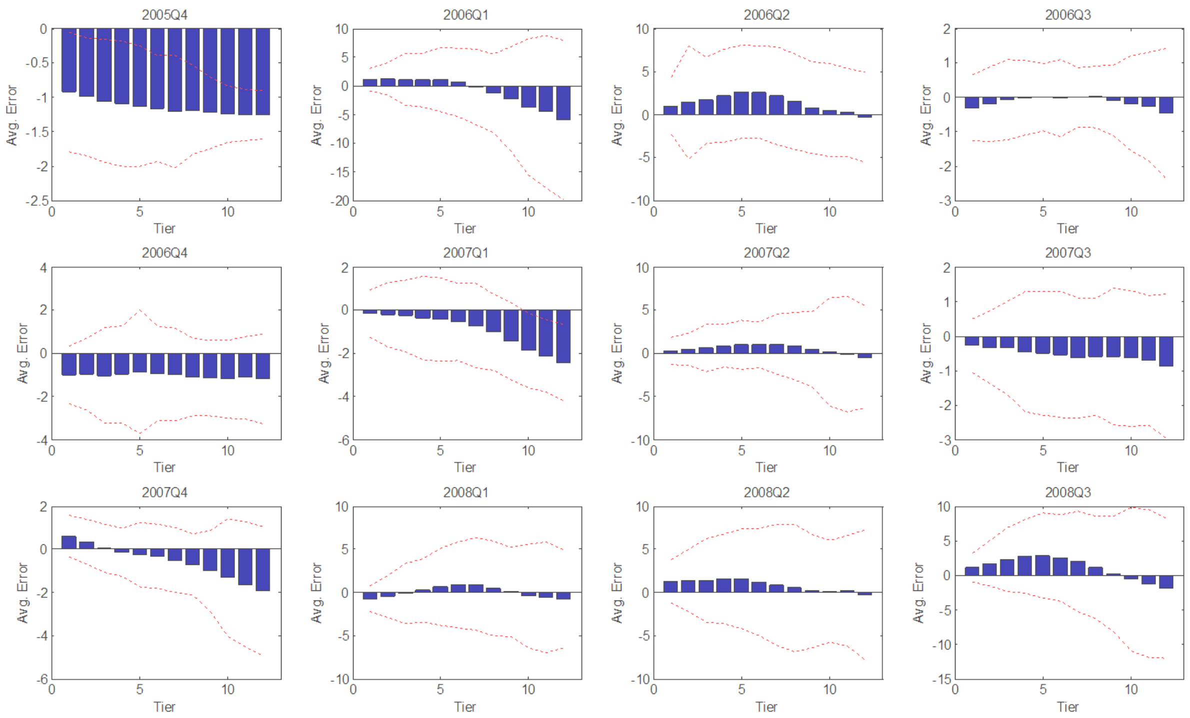


Fig. 4. Forecast set average errors [bars] with ± 2 standard deviations [lines] by tier when M = 5 (cont.)

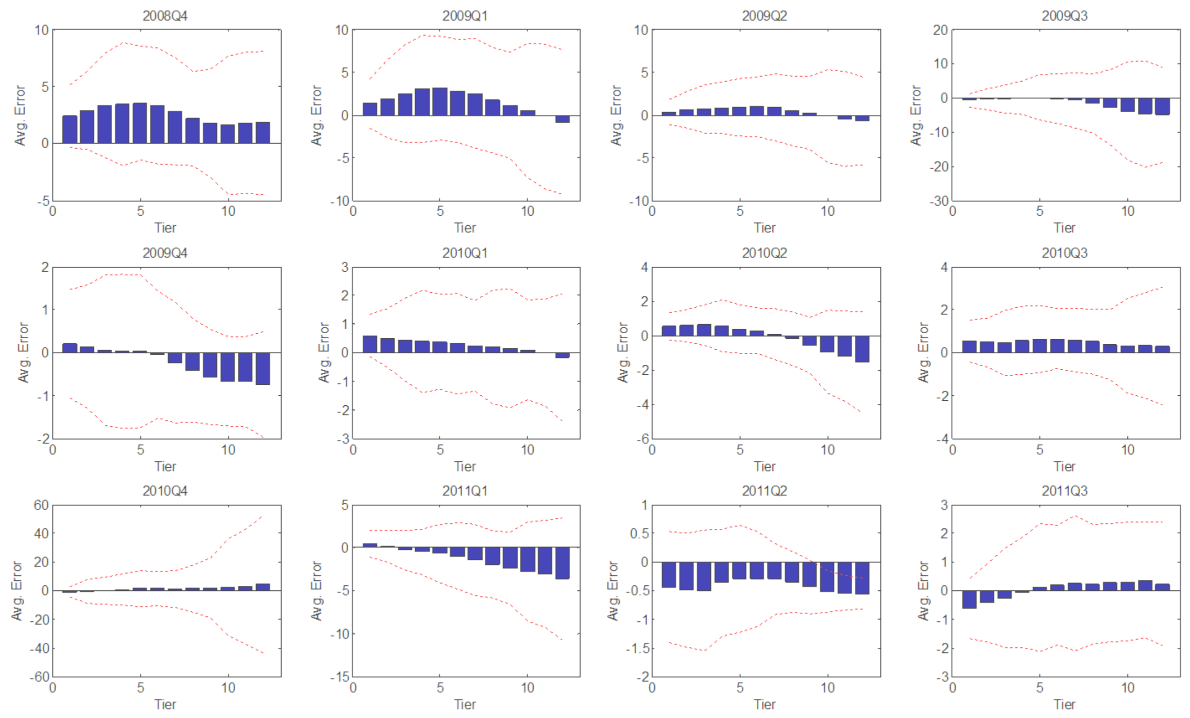


Fig. 5. Forecast set average errors [bars] with ± 2 standard deviations [lines] by tier when M = 2

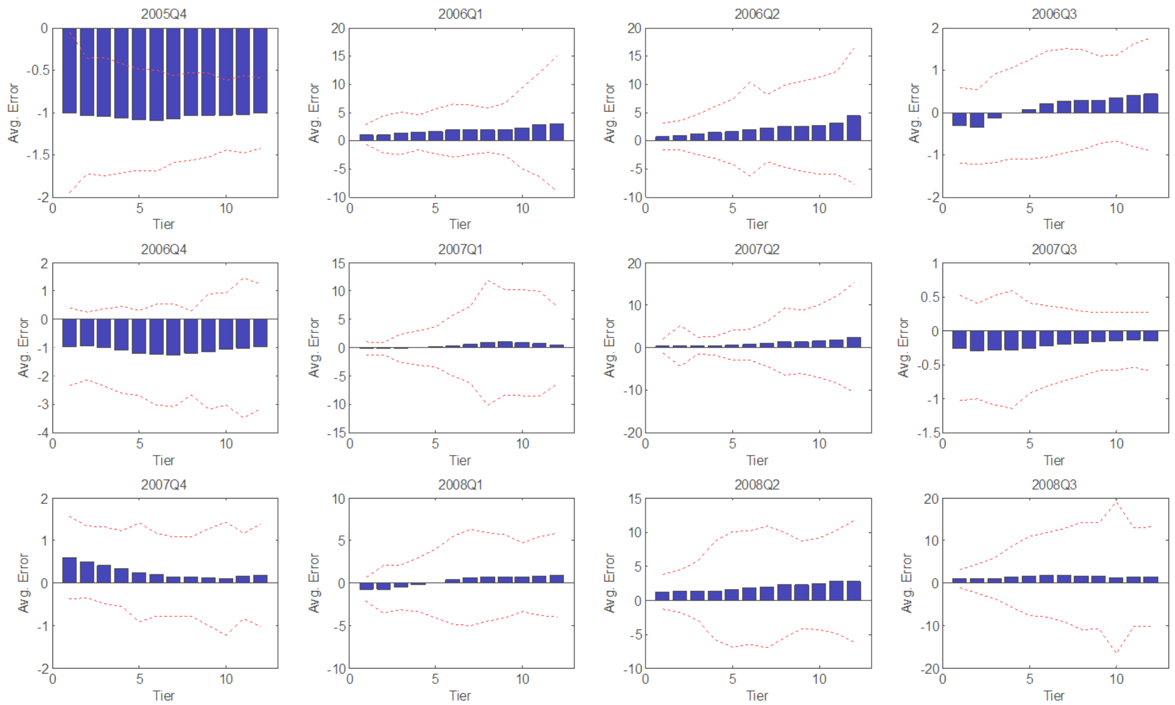


Fig. 6. Forecast set average errors [bars] with ± 2 standard deviations [lines] by tier when M = 2 (cont.)

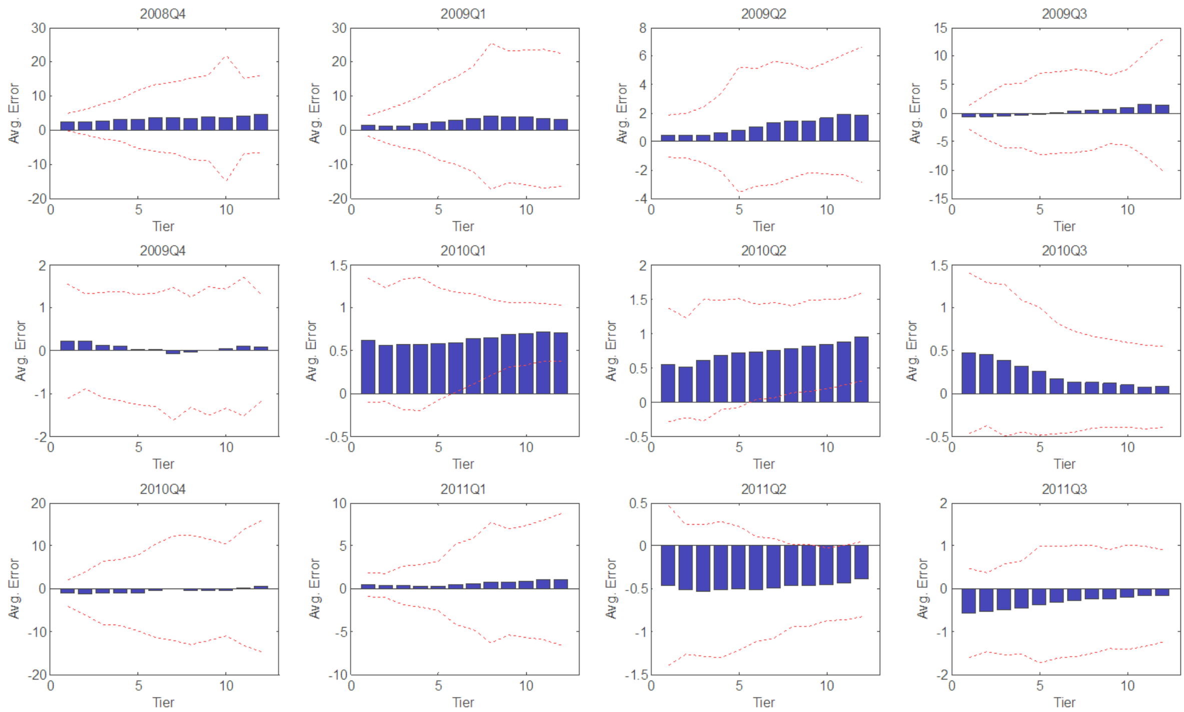


Fig. 7. Forecast set average errors [bars] with ± 2 standard deviations [lines] by tier when M = 10

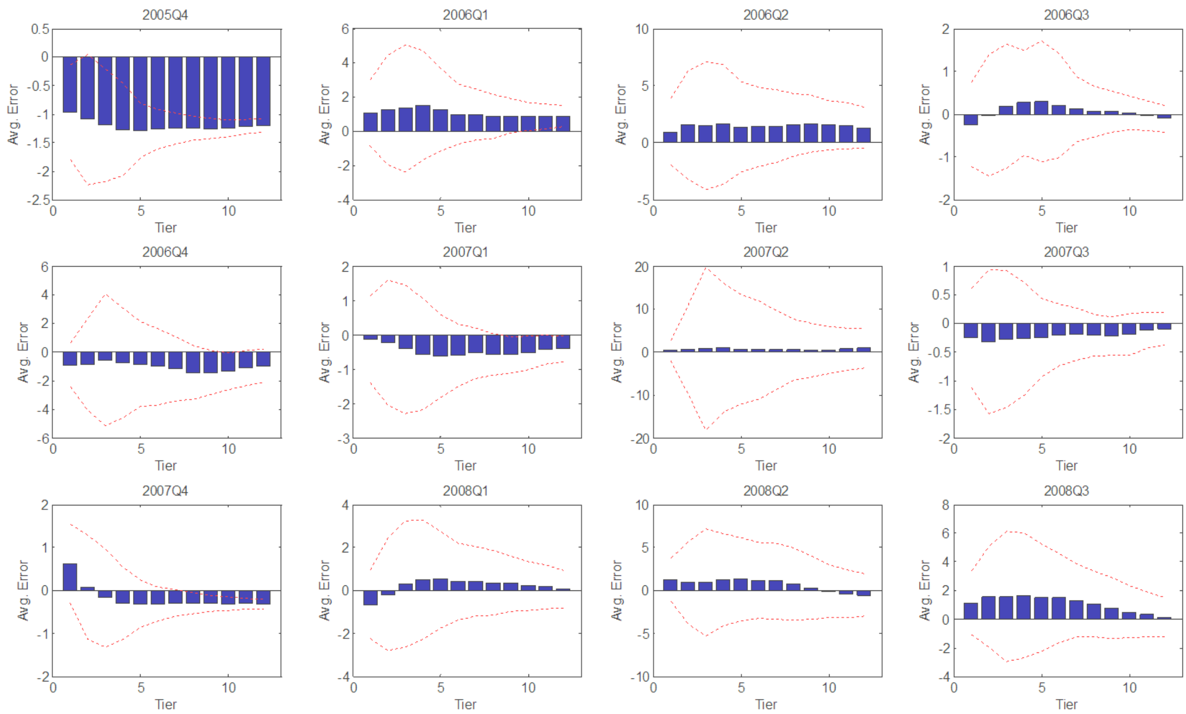
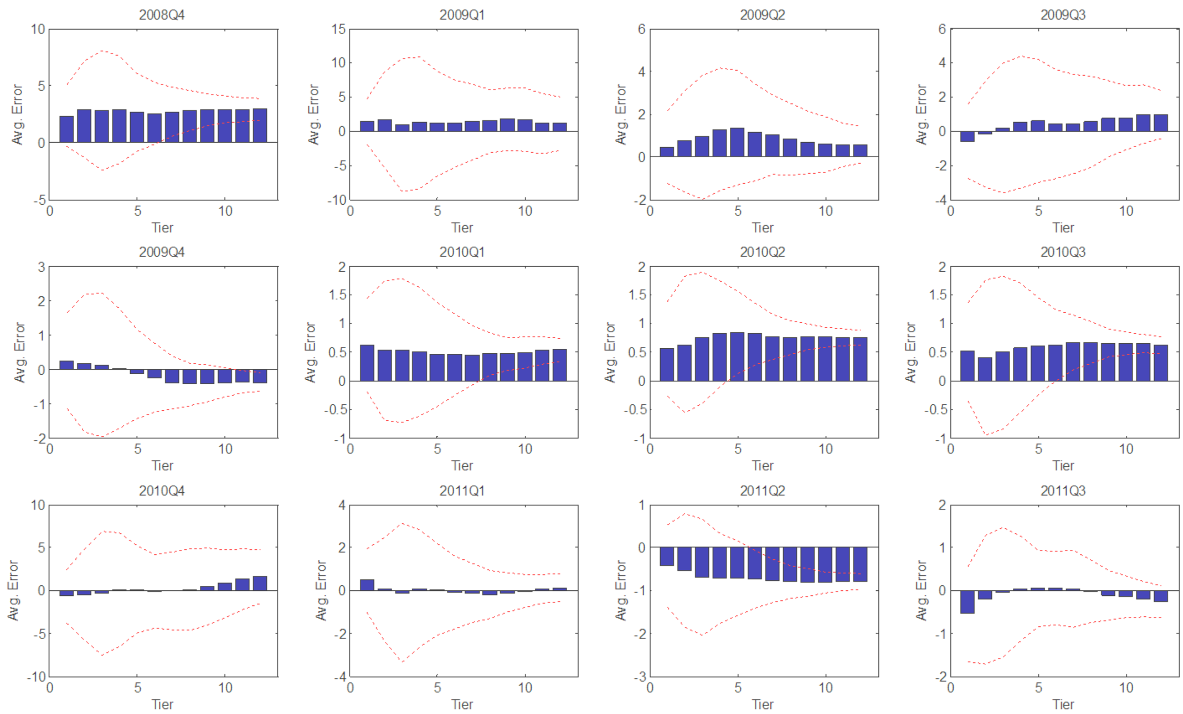


Fig. 8. Forecast set average errors [bars] with ± 2 standard deviations [lines] by tier when M = 10 (cont.)



7. REFERENCES

Aminian, F., Suarez, E. D., Aminian, M. & Walz, D. T. (2006). Forecasting economic data with neural networks. *Computational Economics*, 28(1), 71-88.

Barron, A. (1993). Universal approximation bounds for superpositions of a sigmoidal function. *IEEE Transactions on Information Theory*, 39(3), 930-945.

Beltratti, A., Margarita, S., & Terna, P. (1996). *Neural Networks for Economics and Financial Modelling*. London: International Thomson Computer Press.

Church, K. B. & Curram, S. P. (1996). Forecasting consumers' expenditure: A comparison between econometric and neural network models. *International Journal of Forecasting*, 12(2), 255-267.

Cooper, J. C. B. (1999). Artificial neural networks versus multivariate statistics: An application from economics. *Journal of Applied Statistics*, 26(8), 909-921.

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems*, 2(4), 303-314.

Detienne, K. B., Detienne, D. H. & Joshi, S. A. (2003), Neural networks as statistical tools for business researchers. *Organizational Research Methods*, 6(2), 236-265.

Farhat, D. (2012), Artificial neural networks and aggregate consumption patterns in New Zealand, *Economics Discussion Papers 1205,* September.

Gaffeo, E., Delli Gatti, D., Desiderio, S., & Gallegati, M. (2008). Adaptive microfoundations for emergent macroeconomics. *Universita Degli Studi Di Trento Working Paper No. 2/2008.*

Gatti, D., Di Guilmi, C., Gaffeo, E., Giulioni, G., Gallegati, M., & Palestrini, A. (2005). A new approach to business fluctuations: heterogeneous interacting agents, scaling laws and financial fragility. *Journal of Economic Behavior and Organization*, 56(4), 489-512.

Gonzalez, S. (2000). Neural networks for macroeconomic forecasting: A complementary approach to linear regression models. *Department of Finance Canada Working Paper 2000-07.*

Hornik, K. (1991). Approximation capabilities of multilayer feedforward networks. *Neural Networks*, 4(2), 251-257.

Hornik, K., Stinchcombe, M. & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5), 359-366.

Hornik, K., Stinchcombe, M. & White, H. (1990). Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks. *Neural Networks*, 3(5), 551-560.

International Monetary Fund [IMF] (2011). e-Library Data. Available at http://elibrary-data.imf.org/.

Jain, A. K. & Mao, J. (1996). Artificial neural networks: A tutorial. *Computer*, 29(3), 31-44.

Kourentzes, N. & Crone, S. (2010). Advances in forecasting with artificial neural networks. *Lancaster University Management School Working Paper 2010/023*.

Kuan, C.-M. & White, H. (1994). Artificial neural networks: An econometric perspective. *Econometric Reviews*, 13(1), 1-91.

Mirowski, P. (2007). Markets come to bits: Evolution, computation and markomata in economic science. *Journal of Economic Behavior and Organization*, 63(2), 209-242.

Organisation for Economic Co-operation and Development [OECD] (2011). OECD.StatExtracts. Available at http://stats.oecd.org.

Raberto, M., Teglio, A. & Cincotti, S. (2008). Integrating real and financial markets in an agent-based economic model: An application to monetary policy design. *Computa-*

tional Economics, 32(1-2), 147-162.
Tesfatsion, L. & Judd, K. (eds) (2006). *Handbook of Computational Economics, V.2: Agent-based Computational Economics*. Amsterdam: Elsevier.
Tesfatsion, L. (2002). Agent-based computational economics: Growing economies from the bottom up. *Artificial Life*, 8(1), 55-82.
Tesfatsion, L. (2005). Agent-based computational modeling and macroeconomics. *ISU Economic Report 05023: July 2005.*
Vellido, A., Lisboa, P. J. G. & Vaughan, J. (1999). Neural networks in business: A survey of applications (1992 – 1998). *Expert Systems with Applications*, 17(1), 51-70.
Warner, B. & Misra, M. (1996). Understanding neural networks as statistical tools. *The American Statistician*, 50(4), 284-293.
Zhang, G., Patuwo, B. E. & Hu, M. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting*, 14(1), 35-62.

8. NOTES

1 See Zhang et al. (1998), Vellido et al. (1999), and Kourentzes and Crone (2010) for examples and descriptions of the main advantages and disadvantages of using ANNs.

2 Farhat (2012) explicitly mentions applications to *agent-based models*. Agent-based models (ABMs) are computational simulation models where populations of heterogeneous agents, each with their own characteristics and information, interact with each other in a virtual space. Their interactions and decisions are determined by a prescribed set of rules chosen by the researcher. As they relate to each other at the local level (micro-interaction), aggregate (macroscopic) phenomena emerge. For examples of this work, see Tesfatsion (2002, 2005), Tesfatsion and Judd (2006), Gatti et al. (2005), Mirowski (2007), Gaffeo et al. (2008) and Raberto et al. (2008).

3 A 'bias term' is akin to the 'constant term' in regression analyses.

4 Hornik et al. (1989) and Hornik (1991) show that ANNs can approximate any functional relationship between inputs and outputs with arbitrary precision provided that there are a sufficient number of hidden layers (hence, they are known as 'universal approximators'). Cybenko (1989), Hornik et al. (1990), and Barron (1993) note that ANNs with a single layer (like fig. 1) may also be universal approximators provided that the activation functions used in the model satisfy certain properties (namely, smoothness) and the number of activation functions (*H*) is large enough.

5 A_1 makes no transformation to the input data and γ_1 is fixed to 1. The ANN is, in effect, a linear econometric model of the form $\widehat{C} = \gamma_0 + \sum_{k=1}^{K} \omega_{k1} X_k$ augmented by a non-linear function, $\sum_{h=2}^{H} \gamma_h A(\omega_{0h} + \sum_{k=1}^{K} \omega_{kh} X_k)$.

6 In this study, when an ANN is trained multiple times, the training set and validation set always differ across simulation due to the random allocation of the data. The forecast set, however, will be the same across trained ANNs to make them comparable.

7 See Beltratti et al. (1996) or Warner and Misra (1996) for a more detailed description of the weight updating process for the basic BP algorithm.

8 Data for 1992Q1 is used for the first value of c_{t-1}. The standard score (or z-value) for each data point is used in place of the actual data. As a result, each re-scaled data point is measured as the number of standard deviations the raw data is above its series mean. Using data re-scaled in this way improves the efficiency of the ANN training process.