

CIRJE-F-846

**Forecasting Welfare Caseloads: The Case of the  
Japanese Public Assistance Program**

Masayoshi Hayashi  
The University of Tokyo

April 2012; Revised in October 2013

CIRJE Discussion Papers can be downloaded without charge from:

<http://www.cirje.e.u-tokyo.ac.jp/research/03research02dp.html>

Discussion Papers are a series of manuscripts in their draft form. They are not intended for circulation or distribution except as indicated by the author. For that reason Discussion Papers may not be reproduced or distributed without the written consent of the author.

# Forecasting Welfare Caseloads: The Case of the Japanese Public Assistance Program

**Masayoshi Hayashi**

Faculty & Graduate School of Economics, University of Tokyo

Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan

**E-mail:** hayashim<<at>>e.u-tokyo.ac.jp

**Phone:** +81-3-5841-5513

**Fax:** +81-3-5841-5521

## **Abstract**

Forecasting welfare caseloads has grown in importance in Japan because of their recent rapid increase. Given that the forecasting literature on welfare caseloads only focuses on US cases and utilizes limited classes of forecasting models, this study employs multiple alternative methods in order to forecast Japanese welfare caseloads and compare forecasting performances. In pseudo real-time forecasting, VAR and forecast combinations tend to outperform the other methods investigated. In real-time forecasting, however, a simple version of forecast combinations seems to perform better than the remaining models, predicting that welfare caseloads in Japan will surpass 1.7 million by the beginning of 2016, an approximately 20% increase in five years from the beginning of 2011.

**Keywords:** Public assistance, welfare caseload, forecast combination, Japan

**JEL classification:** C53, H75, H68, I32

**October 2013.**

# 1. Introduction

The volume of Public Assistance (PA) caseloads has been increasing since the early 1990s in Japan.<sup>1</sup> With its acceleration after 2008, caseload growth has been more rapid than ever, as shown in **Figure 1**. Indeed, the number of PA-receiving households has almost doubled in the 12 years since 2001, increasing from 767,000 to 1,528,000. This rapid growth highlights potential problems that make forecasting PA caseloads more important than before. For example, given the typical sluggish adjustments of personnel in the government sector, caseload growth builds up caseworkers' workloads, exacerbating the logistical difficulties in delivering assistance to the poor. Accurate forecasts of welfare caseloads could thus help the public sector manage its personnel in order for it to meet future PA needs. In addition, growing caseloads may also cause appropriation problems among central and local governments.<sup>2</sup> The proportion of PA expenditure in local budgets reached as high as 20% in some municipalities, which also expands central budgets since the central government shares a fixed proportion (75%) of PA benefits, thereby complicating the issue of inter-governmental cost sharing for PA programs. Hence, accurate forecasts of PA caseloads would also help design better central–local fiscal relations for administering PA programs.

## Figure 1

The empirical literature on welfare caseloads has two strands. The first group of studies has explored the determinants of welfare caseloads. While the majority of them focus on US cases, analogous studies exist for Canada [40, 62], Sweden [30, 63], Spain

---

<sup>1</sup> PA here refers to the comprehensive social assistance scheme in Japan that aims to guarantee that all citizens maintain their basic costs of living by providing benefits to those considered to be unable to earn incomes above the basic costs of living.

<sup>2</sup> The Japanese local government consists of two levels, with municipalities (cities, towns, and villages) as the first tier and prefectures as the second. Cities implement PA programs through their welfare offices. Towns and villages are not required to do so, but some of them do so with their own welfare offices. Prefectural welfare offices cover residents in towns and villages that do not implement PA programs.

[5], and Japan [67]. While most studies examine the effect of unemployment, some explore such economic factors as the industry composition [33, 43] and urbanization [16, 43, 57] of local economies. Furthermore, the impact of personal income is examined by analyzing poverty rates [16, 56, 57, 66] or income quintiles [10, 35, 40] in addition to the traditional measure of per capita income. Other measures are population proportions of demographic characteristics, such as educational background [10, 12, 34, 35], gender [33, 43], single motherhood [5, 10, 48, 52, 57, 66], age [10, 24, 52], and race [10, 12, 33, 35, 36, 46, 47, 52, 57]. Some researchers have also examined the political orientation of public sector [10, 16, 35, 36, 52, 62]. Finally, some works focus on the impacts of policy variables such as benefits levels [10, 11, 15, 34–36, 40, 46–49, 52, 56, 62, 64, 66, 71], minimum wages [15, 36, 43, 52, 62], other assistance programs [5, 10, 15, 24, 30, 34, 35, 36, 46, 48, 52, 63], and institutional changes.<sup>3</sup>

However, these “determinant” studies do not help much forecast caseloads, since the forecasts require the future values of such determinants.<sup>4</sup> In other words, to obtain forecasts from the determinant studies, one needs to either forecast the values of the determinants themselves or make a priori scenarios that predetermine the specific sequences of their future values. Doing so may be sensible for certain variables such as an intercept and a linear trend, but more difficult than forecasting the dependent variable (welfare caseloads) itself [53] or simply not justified [23].

---

<sup>3</sup> US studies have examined the effects of changes in the Aid to Families with Dependent Children (AFDC) made by the Omnibus Budget Reconciliation Act [55], Deficit Reduction Act [55], Job Opportunities and Basic Skills Training Programs [43, 46, 47, 48], mandated AFDC-UP [10, 52], and “waivers” from AFDC programs [8, 10, 15, 24, 37, 46, 52, 56, 57, 71]. More recent US studies have considered the effect of the replacement of AFDC with the Temporary Assistance for Needy Families [8, 34, 35, 43]. Some researchers have further allowed for the effects of the sub-elements (work requirements, time limits, incentives, diversion) of these welfare reforms [8, 12, 15, 36, 49, 71].

<sup>4</sup> Some works have utilized dynamic models that have included the lagged values of the dependent variable in addition to these measures [5, 33, 39, 40, 43, 44, 71]. However, forecasting still requires the future values of unlagged measures [14, 51, 54].

The second strand of the literature therefore aims to overcome these difficulties by estimating self-contained time series models. Some such studies are explicit in expressing the prospects of utilizing their analyses for budgeting and operation in practice [28, 29, 42]. However, two gaps must be bridged. First, previous studies have only focused on US cases. Since forecasting welfare caseloads is integral to administering social policy in all countries, analogous analysis for different countries would contribute not only to policy practice in that country but also to our general understanding of welfare policy. Second, previous authors have not exploited forecasting models that go beyond the autoregressive integrated moving average (ARIMA) [3, 13, 42] or vector autoregression (VAR) models [4, 42, 58, 61]. It would thus be interesting to explore alternative forecasting methods that analogous studies in other areas have utilized extensively [9, 20, 21, 50].

This study contributes to the literature and policy practices in this regard by applying several forecasting models to PA caseloads in Japan. The caseload data investigated herein have a monthly frequency and span from 2001 M01 to 2011 M02. The forecasting models used include ARIMA, exponential smoothing (ES), Markov forecasting (MF), two logistic smooth threshold autoregression (LSTAR) models, VAR, and variations of forecast combinations (FCs). In pseudo real-time forecasting, VAR and FC tend to outperform the other methods. In real-time forecasting, by contrast, a simple version of FC seems to perform the best among the alternative models, predicting that Japanese welfare caseloads will surpass 1.7 million by the beginning of 2016, an approximately 20% increase in five years from the beginning of 2011.

The remainder of the paper is organized as follows. Section 2 introduces the forecasting methods employed in this study. Section 3 evaluates these methods. Section

4 conducts a real-time forecasting exercise for the period beyond 2011 M02. Finally, Section 5 concludes.

## 2. Forecasting Models

The monthly averages of PA caseloads  $C_t$  are forecasted herein.<sup>5</sup> The sample spans from 2001 M01 to 2011 M02, which is then split into an in-sample period (2001 M01 to 2010 M02) and an out-of-sample period (2010 M03 to 2011 M02). What follows describes the alternative forecasting models along with their estimates.

### 2.1. Autoregressive Integrated Moving Average

Autoregressive integrated moving average (ARIMA) is one of the most popular forecasting methods. ARIMA models can mimic the behavior of diverse types of series and do so adequately without usually requiring a number of parameter estimates in the final choice of the model [17]. Furthermore, they typically serve as a benchmark to evaluate other forecasting models. Indeed, many forecasters have applied ARIMA models to a number of time series in a variety of areas. Therefore, it is natural to start our discussion by applying ARIMA and the Box–Jenkins methodology to the analysis of PA caseloads here.<sup>6</sup>

The ARIMA analysis here examines the natural logarithm of monthly series of PA caseloads:  $c_t \equiv \ln C_t$ . Since the series is potentially seasonal, it may be characterized as  $ARIMA(p,d,q) \times (p_s,d_s,q_s)_{12}$  where  $p$ ,  $d$ , and  $q$  respectively indicate AR lags, the degree of integration, and MA lags, and  $p_s$ ,  $d_s$ , and  $q_s$  respectively refer to *seasonal* AR lags, the

---

<sup>5</sup> While MF directly forecasts  $C_t$ , the other methods forecast the natural logarithm of  $C_t$ ,  $c_t \equiv \ln C_t$ , and then retrieve  $C_t$  as  $C_t = \exp(c_t)$  when evaluating the forecast methods.

<sup>6</sup> In fact, several studies have modeled the sequence of welfare caseloads as an ARIMA process, although only one has used it for forecasting [42]. Others have used it to conduct intervention analyses [3, 13, 38].

*seasonal* degree of integration, and *seasonal* MA lags. The identification of the ARIMA model here follows the standard Box–Jenkins methodology, which does not need further elaboration here. **Figure 2** shows the correlogram for the *seasonally differenced* first-order differences of the natural logarithm of PA caseloads ( $\Delta c_t - \Delta c_{t-12}$ ), which implies that  $d = d_s = 1$ . Given that  $d = d_s = 1$ , a grid search over  $p$  and  $q$  shows that  $p = 3$  and  $q = 0$  minimize the values of both the Akaike and Bayesian information criterions (AIC and BIC), yielding ARIMA(3,1,0) $\times$ (0,1,0)<sub>12</sub>. Indeed, the autocorrelation and partial autocorrelation in **Figure 2** respectively verify the stationarity and third-order autoregression of the series. The model is estimated by maximum likelihood as

$$\begin{aligned} \Delta c_t - \Delta c_{t-12} = & \underset{(.001)}{.001} + \underset{(.090)}{.224} \cdot (\Delta c_{t-1} - \Delta c_{t-13}) \\ & + \underset{(.088)}{.258} \cdot (\Delta c_{t-2} - \Delta c_{t-14}) + \underset{(.078)}{.391} \cdot (\Delta c_{t-3} - \Delta c_{t-15}) + \varepsilon_t. \end{aligned} \quad (\text{ARIMA})$$

The correlogram of the residuals shows little sign of autocorrelation in **Figure 3**.

### Figures 2 and 3

## 2.2. Exponential Smoothing

Exponential smoothing (ES) focuses on the trend and seasonality of a time series. The literature identifies 15 ES models, which are derived as combinations of the five types of trend (none, additive, damped additive, multiplicative, damped multiplicative) and three types of seasonality (none, additive, multiplicative) [52]. **Table 1** lists the basic specifications, excluding the other six models with damped elements. Among them, the popular specifications are simple ES (no trend, no seasonality), Holt’s linear method (additive trend, no seasonality), Holt–Winters’ additive method (additive trend, additive seasonality), and Holt–Winters’ multiplicative method (additive trend, multiplicative seasonality) [17].

**Table 1**

The three main advantages of utilizing ES techniques are as follows [25, 26]. First, an ES model is easy to interpret, as it generates forecasts as a linear combination of the relevant components (see the models in Table 1). Second, if properly chosen, ES models perform rather well despite their simplicity, since the two main elements (i.e., trend and seasonality) tend to dominate the variance of the series. Third, data requirements and computational effort are not demanding. Despite these benefits, however, other methods are likely to outperform ES if the available data are rich enough to support the use of more elaborate models [17, 25, 26].

Since ES forecasts are linear combinations of the relevant components, the first task is to estimate those weights for a given specification among the nine models in **Table 1**. This exercise estimates the weight parameters by minimizing the sum of the squared errors. The second task is to select a specification among the nine models. The exercise here selects the model that minimizes the value of BIC. The selected model is the following linear-trend model without seasonality:

$$S_t = S_{t-1} + T_{t-1} + .897 \cdot u_t, \text{ and } T_t = T_{t-1} + (.897) \cdot (.205) \cdot u_t \quad (\text{ES})$$

where  $S_t$  is the smoothed level of  $c_t$ ;  $T_t$  is the trend rate; and  $u_t = c_t - c_{t-1}$  is the one-step ahead forecast error.

### 2.3. Logistic Smooth Threshold Autoregression

Threshold autoregression (TAR) allows for the existence of multiple regimes by nesting different linear autoregressive processes in a single time series. For example, a two-regime TAR picks up as the data-generating process (DGP) one such regime after a trigger variable  $s_t$  hits a threshold level  $\mu$  [41]. Typically, but not necessarily, a TAR model uses the lag of the series it explains as the trigger ( $s_t = c_{t-d}$ ), meaning that the model becomes self-contained. In addition, TAR may allow for a smooth transition from



one regime to another by using a transition function  $G(c_{t-d}-\mu) \in [0, 1]$ . Such a process is termed smooth TAR (STAR) [68]. Thus, a two-regime STAR model for  $c_t$  is

$$c_t = \phi_0 + \sum_{p=1}^P \phi_p \cdot c_{t-p} + G(c_{t-d} - \mu) \cdot \left( \psi_0 + \sum_{p=1}^P \psi_p \cdot c_{t-p} \right) + \varepsilon_t.$$

If the transition function is logistic

$$G(c_{t-d} - \mu) = \{1 + \exp[\gamma \cdot (c_{t-d} - \mu)]\}^{-1},$$

the STAR is called logistic STAR (LSTAR), which is one of the standard specifications of a TAR process. Since two-regime LSTAR processes are well suited to characterizing asymmetric cyclical behavior, studies often use LSTAR models to describe cyclical macroeconomic time series [45, 69]. For example, LSTAR models perform well in forecasting unemployment rates [18, 59, 65]. This fact implies that LSTAR models are better suited to characterizing the trend of welfare caseloads, since the determinant studies mentioned in the Introduction have typically shown welfare caseloads to be highly correlated with unemployment rates.

We considered two two-regime LSTAR models and estimated their parameters by using the non-linear least squares method. Assuming that the LSTAR model has an identical lag between the two regimes, LSTAR1 is identified after the selection procedures based on [68]:

$$\begin{aligned} c_t = & .105 + .828 c_{t-1} + .180 c_{t-2} + .627 c_{t-12} - .643 c_{t-13} \\ & (.029) (.075) (.081) (.078) (.070) \\ & + \{1 + \exp[27.035 \cdot (c_{t-10} - 13.933)]\}^{-1} \quad \text{(LSTAR1)} \\ & (10.401) (.008) \\ & \times (.656 + .227 c_{t-1} - .322 c_{t-2} - .166 c_{t-12} + .215 c_{t-13}) + \varepsilon_t. \\ & (2.398) (.396) (.356) (.363) (.514) \end{aligned}$$

However, the five parameter estimates for the second regime are not statistically significant, as seen from the standard errors in parentheses. Further specification tests and diagnostics then yield the restricted version of LSTAR1 (LSTAR2) as

$$\begin{aligned}
c_t = & .098 + .851 c_{t-1} + .157 c_{t-2} + .603 c_{t-12} - .618 c_{t-13} \\
& (.029) (.072) \quad (.078) \quad (.075) \quad (.068) \\
& + \{1 + \exp[38.297 \cdot (c_{t-10} - 13.930)]\}^{-1} \times (-.618 + 1.103 c_{t-1}) + \varepsilon_t \\
& \quad (8.603) \quad (.002) \quad (.068) \quad (.396)
\end{aligned} \tag{LSTAR2}$$

where all coefficients are statistically significant. The forecasts in Section 3 use both LSTAR1 and LSTAR2.

## 2.4. Markov Forecasting

Markov forecasting (MF) is designed for forecasting welfare caseloads [28]. Since the current stock of caseloads equals the previous stock of caseloads plus entries net of exits, caseloads are expressed as the following first-order Markov chain:

$$C_t = C_{t-1} + E_t + X_t = (1 - x_t)C_{t-1} + E_t$$

where  $C_t$ ,  $E_t$ ,  $X_t$ , and  $x_t \equiv X_t/C_{t-1}$  are the caseloads at the end of period, entries, exits, and exit rate, respectively. At the steady state ( $x_t = x$ , and  $E_t = E$ ),  $C_t$  converges to  $C^* = E/x$ . Thus, if the values of  $E_t$  and  $x_t$  were held constant at  $t$ , caseloads would converge to  $C_t^* = E_t/x_t$ . MF regards this value, termed the implied steady state (ISS), to be a leading indicator that predicts caseload value in  $L$  periods ahead by using a linear regression

$$C_{t+L} = \alpha_0 + \alpha_1 \cdot C_t^* + \varepsilon_t.$$

Since the actual ISS tends to be volatile, the use of a smoothed ISS  $\hat{C}_t \equiv E_t(b_E)/x_t(b_x)$  is suggested in place of  $C_t^*$  [28], where  $x_t(b_x)$  and  $E_t(b_E)$  are the values of  $E_t$  and  $x_t$  with LOWESS (locally weighted scatterplot smoothing). Parameters  $\alpha_0$  and  $\alpha_1$  are estimated by using the ordinary least squares method after setting  $b_x = .5$ ,  $b_E = .8$ , and  $L = 5$ . These three values are obtained from a three-dimensional grid search that minimizes the mean square error (MSE) of forecasts.<sup>7</sup> This yields the following forecasting equation:

---

<sup>7</sup> The MSE used one-period ahead forecasts from rolling regressions with a fixed window of 110 observations. The three-dimensional grid search run  $L$  from 10 to 36 by 1, and  $b_x$  and  $b_E$  from .1 to .9

$$C_t = 309171 + .522 \times \hat{C}_{t-5} + \varepsilon_t \quad (\text{MF})$$

(31710) (.023)

where the standard errors are in parentheses.

These forecasts are rescaled values of the smoothed ISS pushed five periods ahead. **Figure 4** shows the rescaled smoothed ISS, forecasts (the rescaled smoothed ISS shifted to the right by five periods), and actual caseloads. The in-sample fits are rather poor, because the method chooses the parameters that minimize the MSE of forecasts rather than the sum of squared residuals (SSR) of the in-sample observations. To illustrate this point, **Figure 4** uses the parameters that together minimize the SSR<sup>8</sup> and plots the fitted values against the actual observations. While the smoothed ISS here is a good leading indicator in the in-sample periods, the forecasts do not perform well in the out-of-sample periods. This is a typical example of where good in-sample fits do not lead to good out-of-sample forecasts.<sup>9</sup>

### Figures 4 and 5

#### 2.5. Vector Autoregression

A vector autoregression (VAR) model refers to a system of interdependent autoregressive processes for a group of series that includes the series in question (here, PA caseloads). Each variable in a VAR model has an equation that explains its trend based on its own lags and the lags of the other variables in the system. VAR models also provide a benchmark as ARIMA models do. Indeed, as they have done with ARIMA models, many forecasters have applied VAR models to a variety of time series in many

---

by .01. The in- and out-samples are from 2001 M01 to 2010 M02 (110 months) and from 2010 M03 to 2011 M02 (12 months), respectively.

<sup>8</sup> They are estimated as  $\alpha_0 = 173$ ,  $\alpha_1 = .887$ ,  $L = 33$ ,  $b_E = .79$ , and  $b_E = .17$ . Because  $L = 33$ , the OLS estimation now has a smaller sample from 2002 M07 to 2010 M02.

<sup>9</sup> It argued that MF can detect a turning point in caseload trajectory [28]. However, the PA caseloads studied here do not display any turning points or changes in the pace of increase in the in-sample periods. We thus cannot validate the strength of MF in this exercise.

areas. By contrast, few studies have used VAR analysis to examine welfare caseloads [42, 58, 61].

The first task in VAR modeling is to select a group of variables in its system. The VAR here concerns three variables: PA caseloads, unemployment rates, and the elderly ratio (the population proportion of those aged 65 years and over). While the choice of unemployment rates is standard [42, 58, 61], the choice of the elderly ratio may be specific to the Japanese case. Japanese programs also cover the elderly, and caseload growth partly reflects the increasing number of older people who have little or no pension benefits [32]. The model uses the natural logarithms of caseloads and the elderly ratio, and unemployment rates as a percentage.

Although the ARIMA analysis above implied that the series was non-stationary, caseload series are not differenced. In ARIMA modeling, it is necessary to make a series stationary since the Box–Jenkins method only relates to stationary processes. In addition, when estimating ARIMA processes, algorithms are likely to fail if series are integrated. These do not apply to VAR models. Furthermore, even if time series are non-stationary and/or integrated, the OLS estimators of VAR coefficients are consistent [31].

The second task is to select the lag  $p$  of VAR( $p$ ). The estimation starts from  $p = 1$  to 13 in order to find  $p = 6$  that minimizes the MSE in the out-of sample. **Table 2** lists the estimation results for VAR(6). The last three lines of the table consist of the  $P$  values for the Granger tests, based on the lag-augmented VAR (LA-VAR) [70]. The tests use LA-VAR(8), allowing for possible integrated or co-integrated variables up to the second order. The results imply that the elderly ratio and PA caseloads do not Granger-cause unemployment rates and the elderly ratio, respectively. However, the exercise below does not exclude them, as doing so weakens the fits of the forecasts.

**Table 2**

## **2.6. Forecast Combination**

Forecast combination (FC) combines the multiple forecasts made by different models for a single time series. Hence, FC diversifies against model uncertainty [1, 22]. When a diagnostic test fails to reject a hypothesis that two models have an equal predictive accuracy, it is not clear which model to choose when making forecasts. Even if one model seems to be better, it is again not clear if it is optimal to ignore the forecasts from the other models altogether. In addition, assume that a forecasting model can only approximate a more complicated and evolving DGP. Then, the model necessarily misspecifies the DGP. Nonetheless, some models, albeit misspecified, may adapt more quickly to the changing behavior of a time series, while others may not. Thus, when multiple forecasts exist, it may be sensible to combine them in order to diversify forecasting errors.

An issue with FC is how to weight the multiple forecasts. This study first uses a simple average or equal weights (FC1), which often outperform other weights that are deliberately designed [22]. In addition, one may improve forecast performance by trimming a group of models by dropping the least performing ones [2]. The current case, where there are six sets of forecasts, excludes the model that performs worst in order to obtain simple-averaged forecasts based on the trimmed group (FC2).

## **3. Performance Comparison**

### **3.1. Pseudo real-time forecasting**

The exercise presented in this section uses data on PA caseloads from 2001 M01 to 2011 M02 ( $T = 122$ ), which are divided into an in-sample period for initial parameter

estimation and model selection and an out-of-sample period for evaluating forecast performance. The out-of-sample period runs from 2010 M03 to 2011 M02 ( $P = 12$ ). For each of the out-of-sample observations, *pseudo real-time forecasting* is performed to make forecasts  $f_t$  as if they were actually made in real time.

Three schemes are employed for updating forecasts, which are widely used in simulating real-time forecasts [22].<sup>10</sup> The first scheme is iterated multi-period forecasting. It recursively generates forecasts ( $f_{N+1|N}, \dots, f_{T|N}$ ) for all the  $P$  periods from a fixed origin (2010 M02). This uses only the *fixed* in-sample ( $N = 110$  observations from 2001 M01 to 2010 M02). The second scheme makes one-period ahead forecasts  $f_{t+1|t}$  using the in-samples with an expanding data window. It makes forecasts for  $t + 1 \geq N$  exploiting *all* the data available up to  $t$  and repeats the process  $P$  times to generate a series of  $P$  one-period ahead forecasts ( $f_{N+1|N}, f_{N+2|N+1}, \dots, f_{T|N+P-1}$ ). Its in-sample thus expands by one observation when an additional forecast is made. The third scheme again makes one-period ahead forecasts ( $f_{N+1|N}, f_{N+2|N+1}, \dots, f_{T|N+P-1}$ ), but does so using the in-samples with a rolling data window of fixed size  $N$ . This is similar to the second scheme of expanding windows except that it drops the earliest observation for an additional forecast.

There are eight estimated models to forecast PA caseloads. Each model generates three types of forecasts for out-of-sample periods and compares them to the actual observations of PA caseloads in those periods. To evaluate the performance of each method, three measures are used, namely the mean absolute error ( $\text{MAE} = \sum |C_t - f_t|/P$ ), root mean square error ( $\text{RMSE} = [\sum (C_t - f_t)^2/P]^{1/2}$ ), and mean error ( $\text{ME} = \sum (C_t -$

---

<sup>10</sup> Note that when forecasting with the LSTAR model, as is standard, the exercise simulates multiple forecast paths by using a pseudo random number generator that follows the normal distribution and averages the multiple paths to obtain forecast values. The number of replications for the simulation here is 10,000.

$f_i/P$ ),<sup>11</sup> which are all calculated using caseloads in level. In addition, the Diebold–Mariano (DM) test [19] is performed to compare more formally a pair of forecasts from different models. The DM test here only utilizes the test statistics based on the RMSE from the forecasts generated by the first scheme (iterated multi-period forecasting). This scheme is used, since we are interested in conducting real-time forecasting for PA caseloads in multiple periods beyond 2011 M02, where we can only conduct iterated multi-period forecasting based on the currently available (fixed) data.

### 3.2. Results

**Table 3** lists the MAE, RMSE, and ME. First, for iterated multi-period forecasts, ES performs the worst, followed by MF. The best model is either FC2 with MAE or VAR with both RMSE and ME. LSTAR1 fares relatively well, being either the second or third best. However, FC1 shows mediocre performance, possibly because there are only six forecasts. Second, the expanding window evaluation changes the ranking. ARIMA now performs the worst in terms of all three measures, while MF remains the second worst as before. VAR and FC2 continue to perform well. Third, the rolling fixed window evaluation provides another different picture. ARIMA is the best in terms of RMSE and ME and the second best in terms of MAE. In addition, possibly reflecting the performance of ARIMA, FC1 ranks first, second, and third in terms of MAE, ME, and RMSE, respectively. However, VAR still fares relatively well, being the second best according to RMSE and third best with MAE and ME. Overall, these results indicate that VAR and one of the FCs perform consistently well. The good performance of VAR is consistent with the previous result [42]. The results for FC also parallel the previous

---

<sup>11</sup> The ME may be less popular for this type of evaluation. However, since public sectors usually plan budgets on an annual basis, month-to-month errors may not be a serious problem as long as they average to zero over a year (12 months). The ME indeed captures such averaging.

finding that while FCs not always deliver the best forecasts, they do not generally deliver poor performance [1].

### Table 3

**Table 4** shows the  $P$  values of the DM tests. LSTAR1 and LSTAR2 are not compared since LSTAR1 nests LSTAR2.<sup>12</sup> The far-left column lists the null hypothesis (model) to be tested and the top row lists the alternative hypothesis (model) to be tested when the associated null hypothesis is rejected. Among the eight forecasts, LSTAR1, VAR, and FC2 are not rejected for any of the alternative models, barring the fact that LSTAR1 is not tested against LSTAR2. Thus, setting aside LSTAR1, VAR and FC2 are among the best models, although their differences are not statistically significant.<sup>13</sup>

### Table 4

## 4. Forecasting Welfare Caseloads

### 4.1. Re-estimated Models

This section now makes real-time forecasts for the periods beyond 2010 M02, using all the available data from 2001 M01 to 2011 M02. Using the terminology in the previous section, this section performs iterated multi-period forecasting that recursively generates forecasts for the periods beyond a fixed origin 2010 M02, using a fixed in-sample of observations from 2001 M01 to 2011 M02. Since the current estimation uses longer data than those used in Section 2 spanning from 2001 M01 to 2011 M02, the specifications are different from those in Section 2 for some of the forecasting models, and the coefficient estimates are different from those in Section 2 for all models, too.

---

<sup>12</sup> If the DM test relates to a pair of models in which one nests the other, the distribution of the test statistics will be non-standard.

<sup>13</sup> Since a study argued that more than 100 forecasts may be necessary to establish significant differences in predictive accuracy across models [6], the results in **Table 4** should be taken with caution as they rely on only 12 forecast values. However, as all the  $P$  values for the well-performing models are close to one except when compared among themselves, we may not necessarily have to be too cautious on this point.



First, the Box–Jenkins methodology yielded the following ARIMA specification:

$$\begin{aligned} \Delta c_t - \Delta c_{t-12} = & .061 + 1.350 \cdot (\Delta c_{t-1} - \Delta c_{t-13}) + .157 \cdot (\Delta c_{t-2} - \Delta c_{t-14}) \\ & (.004) \quad (.077) \qquad \qquad \qquad (.143) \\ & - .515 \cdot (\Delta c_{t-3} - \Delta c_{t-15}) + \varepsilon_t - .749 \cdot \varepsilon_{t-12} - .173 \cdot \varepsilon_{t-13} \cdot \\ & (.076) \qquad \qquad \qquad (.112) \qquad \qquad \qquad (.096) \end{aligned} \quad (\text{ARIMA}')$$

Second, the ES model now has a linear trend and multiplicative seasonality

$$\begin{aligned} S_t = S_{t-1} + T_{t-1} + .924 \cdot u_t / I_{t-p}, \quad T_t = T_{t-1} + (.924) \cdot (.263) \cdot u_t / I_{t-p}, \quad \text{and} \\ I_t = I_{t-1} + (.855) \cdot (1 - .924) \cdot u_t / S_t \end{aligned} \quad (\text{ES}')$$

where  $I_t$  is the seasonal index and  $p$  the seasonal span. Third, the two LSTAR models are estimated as

$$\begin{aligned} c_t = & .104 + .830 c_{t-1} + .177 c_{t-2} + .624 c_{t-12} - .639 c_{t-13} \\ & (.028) \quad (.071) \quad (.077) \quad (.074) \quad (.067) \\ & + 1 / \{1 + \exp[29.845 \cdot (c_{t-10} - 13.933)]\} \\ & \qquad \qquad \qquad (6.507) \qquad \qquad \qquad (.003) \\ & \times (.751 + .079 c_{t-1} - .176 c_{t-2} - .081 c_{t-12} + .124 c_{t-13}) + \varepsilon_t, \quad \text{and} \\ & \qquad \qquad \qquad (.166) \quad (.234) \quad (.239) \quad (.204) \quad (.180) \end{aligned} \quad (\text{LSTAR1}')$$

$$\begin{aligned} c_t = & .113 + .847 c_{t-1} + .150 c_{t-2} + .610 c_{t-12} - .615 c_{t-13} \\ & (.026) \quad (.066) \quad (.072) \quad (.068) \quad (.061) \\ & + 1 / \{1 + \exp[41.089 \cdot (c_{t-10} - 13.932)]\} \times (.837 - .059 c_{t-1}) + \varepsilon_t. \end{aligned} \quad (\text{LSTAR2}')$$

Fourth, the MF model is estimated as

$$C_t = 342,734 + .497 \times \hat{C}_{t-5} + \varepsilon_t \quad (\text{MF}')$$

(22,829)    (.015)

where  $b_x = .17$ , and  $b_E = .76$ . Note that MF cannot forecast caseloads in periods more than  $L$ -periods ahead ( $L = 5$ ) in this real-time forecasting, since unlike, the previous pseudo real-time forecasting, the data are unavailable to obtain the regressor (smoothed ISS) beyond 2011 M02. Fifth, the VAR model is still specified as VAR(6) with the different coefficient estimates listed in **Table 5**.

**Table 5**

Lastly, since the information from the validation period is now available, the FCs here consider two types in addition to FC1 and FC2. These two additional FCs use

different types of weights that require information ex-ante to the forecast origin (2010 M02). First, FC3 utilizes weights based on inverse mean squared errors [7], which weigh forecasts made by the  $j$ -th forecasting model with

$$\omega_j = \frac{1/MSE_j}{\sum_i 1/MSE_i} \quad (\text{FC3})$$

where MSE is calculated from the RMSE values listed in the third column in **Table 3**.

Second, FC4 uses the weights proposed by [27]. It thus takes advantage of the estimates from the following regression model:

$$C_t = \alpha + \sum_j \beta_j \cdot f_{jt} + \text{residuals} \quad \text{for } N < t \leq T. \quad (\text{FC4})$$

The estimated coefficients, in addition to the constant, rescale their corresponding forecasts beyond 2011 M02. As shown in **Table 6**, the fit of the regression is very good ( $R^2 = .9999$  and  $\bar{R}^2 = .9998$ ). However, these “weights” turn out to be different from what we usually think weights are. First, they do not add up to unity; however, when excluding the constant  $\alpha$ , the sum of the coefficients is close to one (1.012). Second, there are three negative coefficients ( $-.722$  for ES,  $-.719$  for LSTAR1, and  $-.234$  for LSTAR2). Third, the coefficient (“weight”) on the forecasts from the VAR model is more than unity (1.696). Note that MF only provided forecasts up to 2010 M07 in the validation period since its smoothed ISS had a lag of five periods. For this reason, all the FCs exclude MF.

**Table 6**

## 4.2. Results

**Figure 7** describes the forecasts. Not surprisingly, each of the non-FCs behaves erratically in the long-term. Furthermore, although their behaviors were more or less similar when forecasted for the out-of sample period, they are different here. MF, which

has only five forecast values, seems to predict the largest level of caseloads for the limited periods of forecasting. ES also displays constantly increasing caseloads well beyond two million toward the end of 2015. Likewise, ARIMA also shows increasing caseloads that, while not as sharp as the rise of ES, is also substantially high at just below two million by the end of the period. By contrast, LSTAR1 and LSTAR2 show long-run declines, even though they slowly increased in the early period. This finding is implausible given the current socio-economic trends affecting PA, such as the aging population.

VAR displays another implausible movement. In the earlier stages, its forecasts started decreasing more rapidly than those of the two LSTARs and then increased at the beginning of 2014. This is striking since VAR performed well in the pseudo real-time forecasting. These erratic behaviors of the individual forecasts are different from those of the pseudo real-time forecasts. This result implies that good performance in the validation period, which is typically short for this study, may not constitute a sound foundation for longer-term forecasting.

Finally, let us consider the four FCs. Since MF only generates forecasts up to 2011 M07, FC1 generates its forecasts up to 2011 M07 and FC3 and FC4 exclude its forecasts. FC2 also excludes the forecasts by MF (it seems to be the worst performing, as **Figure 6** indicates). The two more elaborate FCs, FC3 and FC4, behave more erratically than the forecasts from some of the individual models. In particular, the forecasts made by FC4 start to decrease as early as 2011 and continue to decline until mid-2013 before increasing rapidly to exceed the forecasts made by FC2 for 2015. This result occurs because FC4 places the largest “weight” of 1.696 on the forecasts from VAR, which behaved erratically as described above. The behavior of the FC3 forecasts

is less erratic compared with those of FC4; however, it still displays a decrease that starts in mid-2012.

### **Figure 6**

The forecasts made by FC2 seem to be a safe choice in that this model demonstrates the most plausible future path of PA caseloads given the available information that may affect caseloads in the future, including the aging population, continuing stagnant economy, and changing labor market practices in Japan. FC2 caseload forecasts FC consistently increase beyond 2011 M02 with an inflection point around early 2014, reaching more than 1.7 million toward the end of 2015.

In fact, as some time has elapsed since these forecasts were calculated, the additional finalized data on PA caseloads have become available up to 2012 M03. **Table 7** shows both the actual and the forecasted PA caseloads along with their monthly differences (in absolute values). While the forecasts underestimate PA caseloads with an increasing order, the percentage errors all remain below 1%. This performance is outstanding despite the increase in PA caseloads brought about by the Great East Japan Earthquake in March 2011. In addition, this finding may constitute another example of an “FC puzzle,” where simple combinations of point forecasts repeatedly outperform other sophisticated FCs [60].

### **Table 7**

## **5. Concluding Remarks**

Forecasting welfare caseloads has grown in importance in Japan due to their recent rapid increase. Against the background that most previous studies only examine US cases and have not exploited recent developments in the literature, this study employed several forecasting methods (ARIMA, ES, LSTAR, VAR, and a set of FCs) in order to

predict Japanese welfare caseloads and compare their performances. The presented findings showed that while the VAR model and one of the FC methods tend to outperform the others in pseudo real-time forecasting, a simple average FC seems to perform best in real-time forecasting. In addition, simple average forecasting predicts that PA caseloads in Japan will surpass 1.7 million by the end of the forecast period (2016 M02), an approximately 20% increase from the beginning of 2011.

## Acknowledgements

The author is grateful to Vedat Verter and two anonymous reviewers for their helpful comments and suggestions. He also thanks Mototsugu Fukushige, Eiya Hatano, Haruko Noguchi, and the other participants at a workshop held by the National Institute of Population and Social Security Research for their comments. This study was financially supported by the Grant-in-Aid for Scientific Research (B-120330064 & C-23530360) of the Japan Society for the Promotion of Science.

## References

- [1] Aiolfi M, Capistrán C, Timmermann, A. Forecast combinations. In: Clements MP, Hendry DE (Eds.), *The Oxford handbook of economic forecasting*. Oxford: Oxford University Press; 2011.
- [2] Aiolfi M, Timmermann A. Persistence in forecasting performance and conditional combination strategies. *Journal of Econometrics* 2006; 135(1/2): 31–53.
- [3] Albritton RB. Measuring public policy: Impacts of the Supplemental Security Income program. *American Journal of Political Science* 1979; 23(3): 559–578.
- [4] Andini C. Unemployment and welfare participation in a structural VAR: Rethinking the 1990s in the United States. *International Review of Applied Economics* 2006; 20(2): 243–253.
- [5] Ayala L, Pérez C. Macroeconomic conditions, institutional factors and demographic structure: What causes welfare caseloads? *Journal of Population Economics* 2005; 18(3): 563–581.
- [6] Ashley R. A new technique for postsample model selection and validation. *Journal of Economic Dynamics and Control* 1998; 22(5): 647–665.
- [7] Bates JM, Granger CWJ. The combination of forecasts. *Operational Research Quarterly* 1969; 20(4): 451–468.
- [8] Bitler MP, Hoynes HW. The state of the social safety net in the post-welfare reform era. *Brookings Papers on Economic Activity* 2010; Fall: 71–127.
- [9] Bjørnbkabd HC, Gerdrup K, Jore AS, Smith C, Thorsrud LA. Does forecast combination improve Norges Bank inflation forecasts? *Oxford Bulletin of Economics and Statistics*

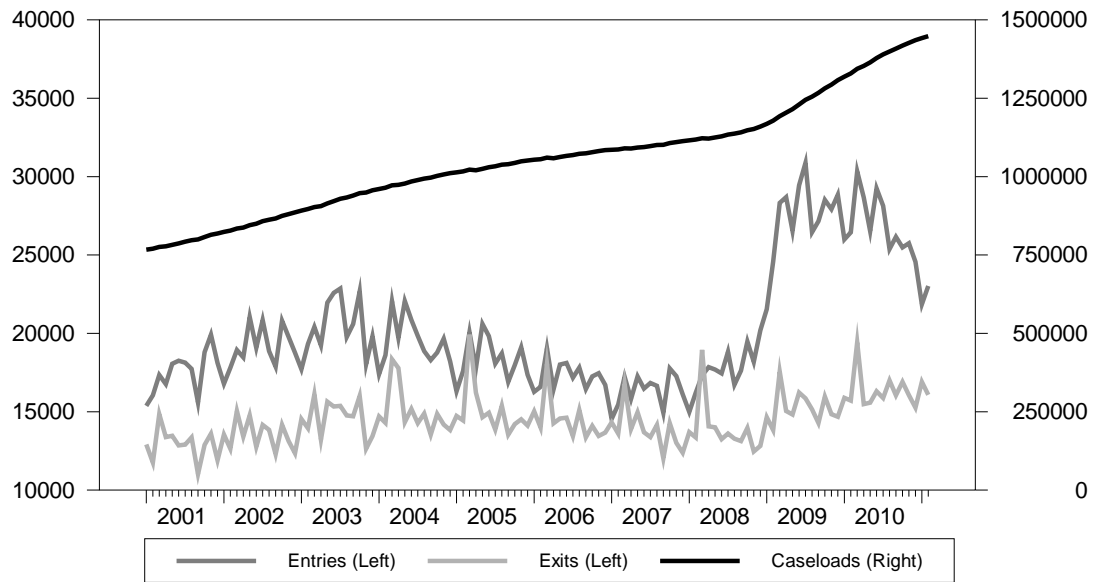
- 2012; 74(2): 163–179.
- [10] Blank RM. What causes public assistance caseloads to grow? *Journal of Human Resources* 2001; 36(1): 85–118.
- [11] Brehm CT, Saving TR. The demand for general assistance payments. *American Economic Review* 1964; 54(6): 1002–1018.
- [12] Cadena B, Danziger S, Seefeldt K. Measuring state welfare policy changes: Why don't they explain caseload and employment outcome? *Social Science Quarterly* 2006; 87(4): 808–817.
- [13] Chang, HJ. Explaining welfare caseload reduction in New York State: The effect of policy or economy? *International Review of Public Administration* 2007; 12(1): 105–117.
- [14] Conte M, Levy DT, Shahrokh F, Staveley J, Thompson S. Economic determinants of income maintenance programs: The Maryland forecasting model. *Journal of Policy Modeling* 1998; 20(4): 461–481.
- [15] Danielson C, Klerman JA. Did welfare reform cause the caseload decline? *Social Service Review* 2008; 82(4): 703–730.
- [16] De Alteriis M. Local governments as implementers of public policy. *Policy Studies Review* 1990; 9(4): 765–773.
- [17] De Gooijer JG, Hyndman RJ. 25 years of time series forecasting. *International Journal of Forecasting* 2006; 22(3): 443–473.
- [18] Deschamps PJ. Comparing smooth transition and Markov switching autoregressive models of US unemployment. *Journal of Applied Econometrics* 2008; 23(4): 435–462.
- [19] Diebold FX, Mariano RS. Comparing predictive accuracy. *Journal of Business and Economic Statistics* 1995; 13(3): 253–263.
- [20] Duasa J, Ahmad N, Ibrahim MH, Zainal MP. Forecasting Inflation in Malaysia. *Journal of Forecasting* 2010; 29(6): 573–594.
- [21] Edlund P-O, Karlsson S. Forecasting the Swedish unemployment rate: VAR vs. transfer function modelling. *International Journal of Forecasting* 1993; 9(1): 61–76.
- [22] Elliot G, Timmermann A. Economic forecasting. *Journal of Economic Literature* 2008; 46(1): 2–56.
- [23] Erricson NR. Predictable uncertainty in economic forecasting. In: Clements MP, Hendry DF (Eds), *A companion to economic forecasting*. Malden, MA: Blackwell Publishing; 2002.
- [24] Garasky S. Analyzing the effect of Massachusetts' ET choice program on the State's AFDC-BASIC caseload. *Evaluation Review* 1990; 14(6): 701–710.
- [25] Gardner ES Jr. Exponential smoothing: The state of the art. *Journal of Forecasting* 1985; 4(1): 1–28.
- [26] Gardner ES Jr. Exponential smoothing: The state of the art—Part II. *International Journal of Forecasting* 2006; 22(4): 637–666.
- [27] Granger CWJ, Ramanathan R. Improved method of combining forecasts. *Journal of Forecasting* 1984; 3(2): 197–204.
- [28] Grogger J. Markov forecasting methods for welfare caseloads. *Children and Youth Services Review* 2007; 29(7): 900–911.
- [29] Gurmu S, Smith WJ. Estimating and forecasting welfare caseloads. In: Sun J, Lynch TD (Eds.), *Government budget forecasting: theory and practice*. Boca Raton, FL: Auerbach Publications; 2008.
- [30] Gustafsson B. Macroeconomic performance, old age security and the rate of social assistance recipients in Sweden. *European Economic Review* 1984; 26(3): 319–338.
- [31] Hamilton J. *Time series analysis*. Princeton: Princeton University Press; 1994.
- [32] Hayashi M. Social Protection in Japan: Current State and Challenges. In: Asher MG.

- Oum S, Parulian F (Eds.), *Social protection in East Asia: current state and challenges*. Jakarta: Economic Research Institute for ASEAN and East Asia (ERIA); 2010.
- [33] Hill BC, Murray MN. Interactions between welfare caseloads and local labor markets. *Contemporary Economic Policy* 2008; 26(4): 539–554.
- [34] Hsing Y. Socioeconomic analysis of the determinants of TANF recipients in the USA and policy implications. *International Journal of Social Economics* 2004; 31(11/12): 1005–1013.
- [35] Huang C-C, Garfinkel I, Waldfogel J. Child support enforcement and welfare caseloads. *Journal of Human Resources* 2004; 39(1): 108–134.
- [36] Jagannathan R. Welfare reform's impact on caseload decline in the United States: An application of latent trajectory model. *Social Science Journal* 2011; 48(4): 703–721.
- [37] Johnson TR, Klepinger DH, Dong FB. Caseload impacts of welfare reform. *Contemporary Economic Policy* 1994; 12(1): 89–101.
- [38] Klassen TR, Buchanan D. Getting it backward? Economy and welfare in Ontario 1985–1995. *Canadian Public Policy* 1997; 23(3): 333–338.
- [39] Klerman JA, Haider SJ. A stock-flow analysis of the welfare caseload. *Journal of Human Resources* 2004; 39(4): 865–886.
- [40] Kneebone R, While KG. Fiscal retrenchment and social assistance in Canada. *Canadian Public Policy* 2009; 35(1): 21–40.
- [41] Kock AB, Teräsvirta T. Forecasting with nonlinear time series models. In: Clements MP, Hendry DE (Eds.), *The Oxford handbook of economic forecasting*. Oxford: Oxford University Press; 2011.
- [42] Lazariu V, Chengxuan Y, Gundersen C. Forecasting women, infants, and children caseloads: A comparison of vector autoregression and autoregressive integrated moving average approaches. *Contemporary Economic Policy* 2011; 29(1): 46–55.
- [43] Lee MA, Harvey M, Neustrom A. Local labor markets and caseload decline in Louisiana in the 1990s. *Rural Sociology* 2002; 67(4): 556–577.
- [44] Lews W, Henry M. Do rural welfare caseloads have a stake in metropolitan area growth? *Southern Business & Economic Journal* 2004; 27(1/2): 22–36.
- [45] Lundbergh S, Teräsvirta T. Forecasting with Smooth Transition Autoregressive Models. In: Clements, MP, Hendry DF. editors. *A companion to economic forecasting*. Malden, MA: Blackwell Publishing; 2002.
- [46] Mead LM. The decline of welfare in Wisconsin. *Journal of Public Administration Research and Theory* 1999; 9(4): 597–622.
- [47] Mead LM. Caseload change: An exploratory study. *Journal of Policy Analysis and Management* 2000; 19(3): 465–472.
- [48] Mead LM. Welfare caseload change: An alternative approach. *Policy Studies Journal* 2003; 31(2): 163–185.
- [49] New MJ. State sanctions and the decline in welfare caseloads. *Cato Journal* 2008; 28(3): 515–533.
- [50] Norman D, Richards A. The forecasting performance of single equation models of inflation. *Economic Records* 2012; 88(280): 64–78.
- [51] Opitz W, Nelson H. Short-term, population-based forecasting in the public sector: A dynamic caseload simulation model. *Population Research and Policy Review* 1996; 15(5/6): 549–563.
- [52] Page ME, Spetz J, Millar J. Does the minimum wage affect welfare caseloads? *Journal of Policy Analysis and Management* 2005; 24(2): 273–295.
- [53] Pindyck RS, Rubinfeld DL. *Econometric models and economic forecasts*. (3rd ed.) New York: McGraw-Hill; 1991.

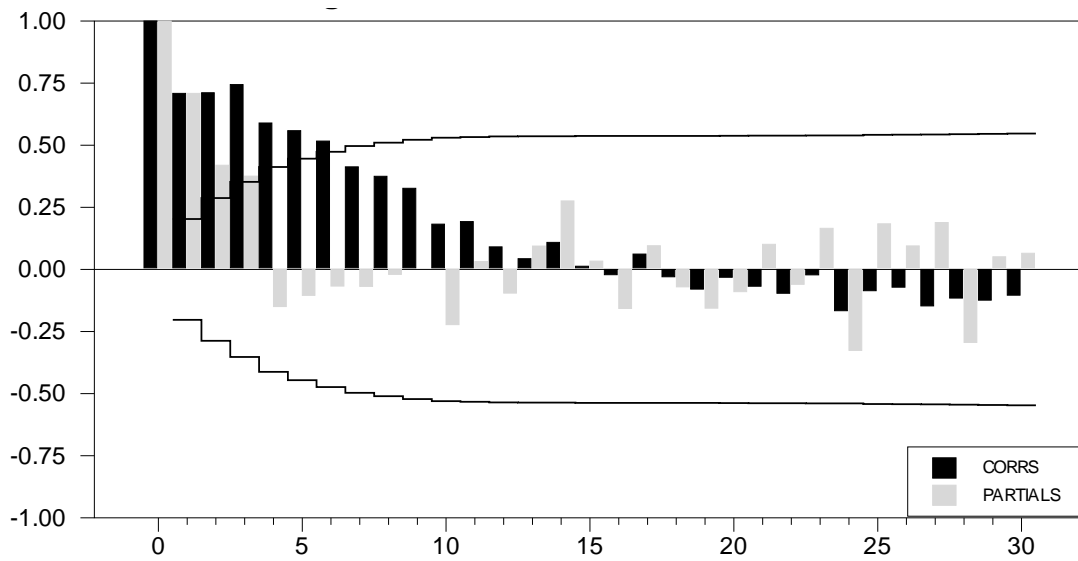
- [54] Plotnick RD, Lidman RM. Forecasting welfare caseloads: A tool to improve budgeting. *Public Budgeting & Finance* 1987; 7(3): 70–81.
- [55] Riggin, LJC, Ward-Zukerman B. Effects on the AFDC-Basic caseload of providing welfare to two-parent families. *Social Science Journal* 1995; 32(3): 265–278.
- [56] Schiller BR. State welfare-reform impacts: Content and enforcement effects. *Contemporary Economic Policy* 1999; 17(2): 210–222.
- [57] Schiller BR, Brasher CN. Effects of workfare saturation on AFDC caseloads. *Contemporary Policy Issues* 1993; 11(2): 39–49.
- [58] Shah P, Smith PK. Do welfare benefits cause the welfare caseload? *Public Choice* 1995; 85(1/2): 91–105.
- [59] Skalin J, Teräsvirta T. Modeling asymmetries and moving equilibria in unemployment rates. *Macroeconomic Dynamics* 2002; 6(2): 202–241.
- [60] Smith J, Wallis KF. A simple explanation of the forecast combination puzzle. *Oxford Bulletin of Economics and Statistics* 2009; 71(3): 331–355.
- [61] Smith PK. Welfare as a cause of poverty: A time series analysis. *Public Choice* 1991; 75(2): 157–170.
- [62] Spindler ZA, Gilbreath WS. Determinants of Canadian social assistance participation rates. *International Journal of Social Economics* 1979; 6(3): 164–1974.
- [63] Stenberg S. Unemployment and economic hardship: A combined macro- and micro-level analysis of the relationship between unemployment and means-tested social assistance in Sweden. *European Sociological Review* 1998; 14(1): 1–3.
- [64] Stephenson EF. Even the underprivileged are rational: The incentive effects of welfare. *Journal of Labor Research* 1997; 18(2): 367–372.
- [65] Stevenson M, Peat M. Forecasting Australian unemployment rates. *Australian Journal of Labour Economics* 2000; 4(1): 41–55.
- [66] Stine WF. The effect of local administrative stringency on the provision of AFDC. *Public Finance Quarterly* 1988; 16(3):284–300.
- [67] Suzuki W, Zhou Y. Welfare use in Japan: Trends and determinants. *Journal of Income Distribution* 2007; 16(3/4): 88–109.
- [68] Teräsvirta T. Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association* 1994; 89(425): 208–218.
- [69] Teräsvirta T, Van Dijk D, Medeiros MC. Smooth transition autoregressions, neural networks, and linear models in forecasting macroeconomic time series: A re-examination. *International Journal of Forecasting* 2005; 21(4): 755–774.
- [70] Toda HY, Yamamoto T. Statistical inference in vector autoregressions with possibly integrated processes. *Journal of Econometrics* 1995; 66(1/2): 225–250.
- [71] Ziliak JP, Figlio D, Davis E, Connolly L. Accounting for the decline in AFDC caseloads: Welfare reform or economy? *Journal of Human Resources* 2000; 35(3): 570–586.



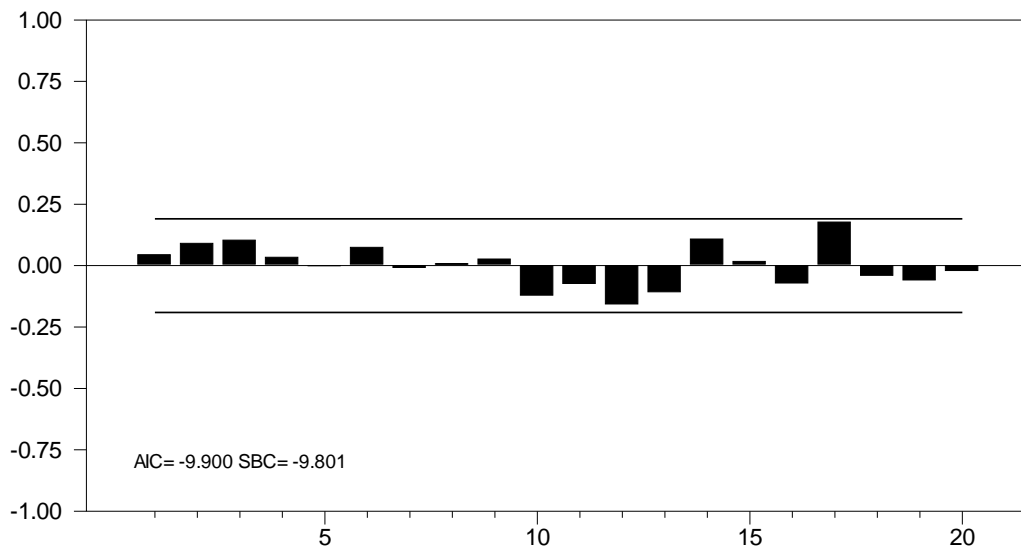
**Figure 1. PA Caseloads in the 2000s**



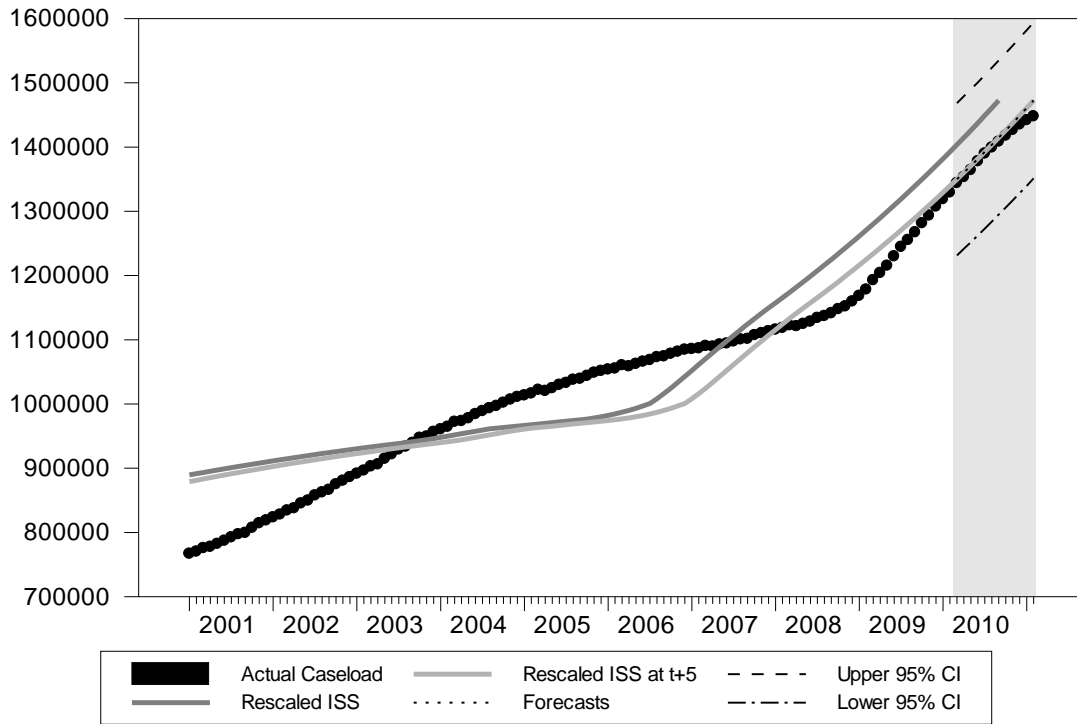
**Figure 2. Correlogram for  $\Delta \ln C_t - \Delta \ln C_{t-12}$**



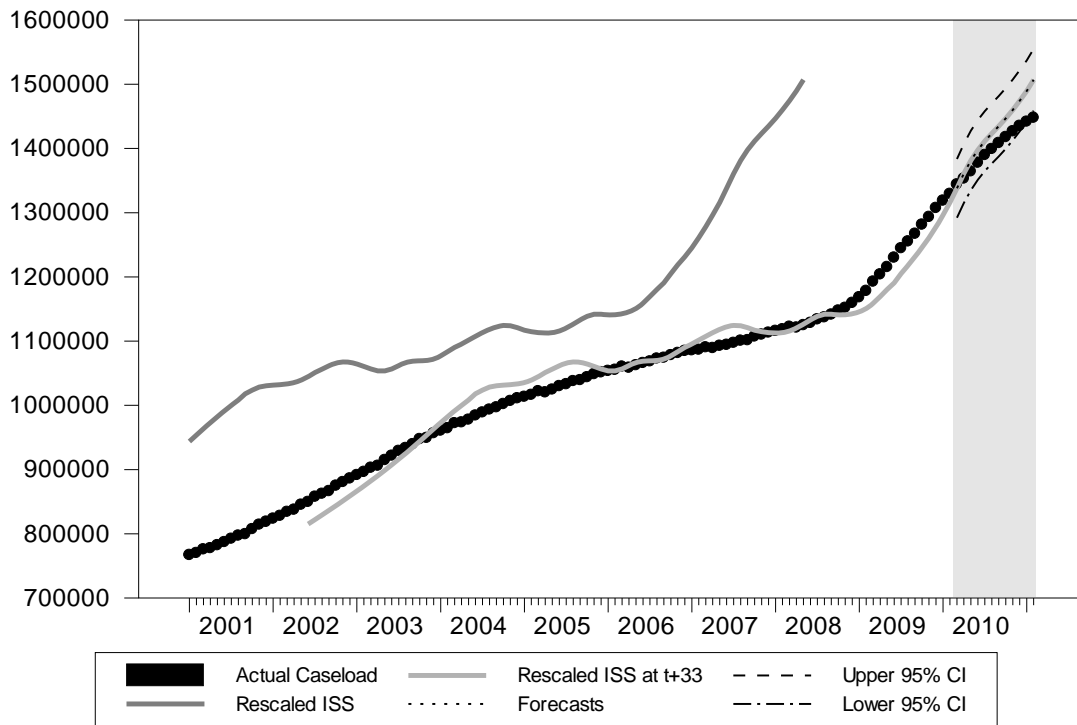
**Figure 3. Correlogram for Residuals**



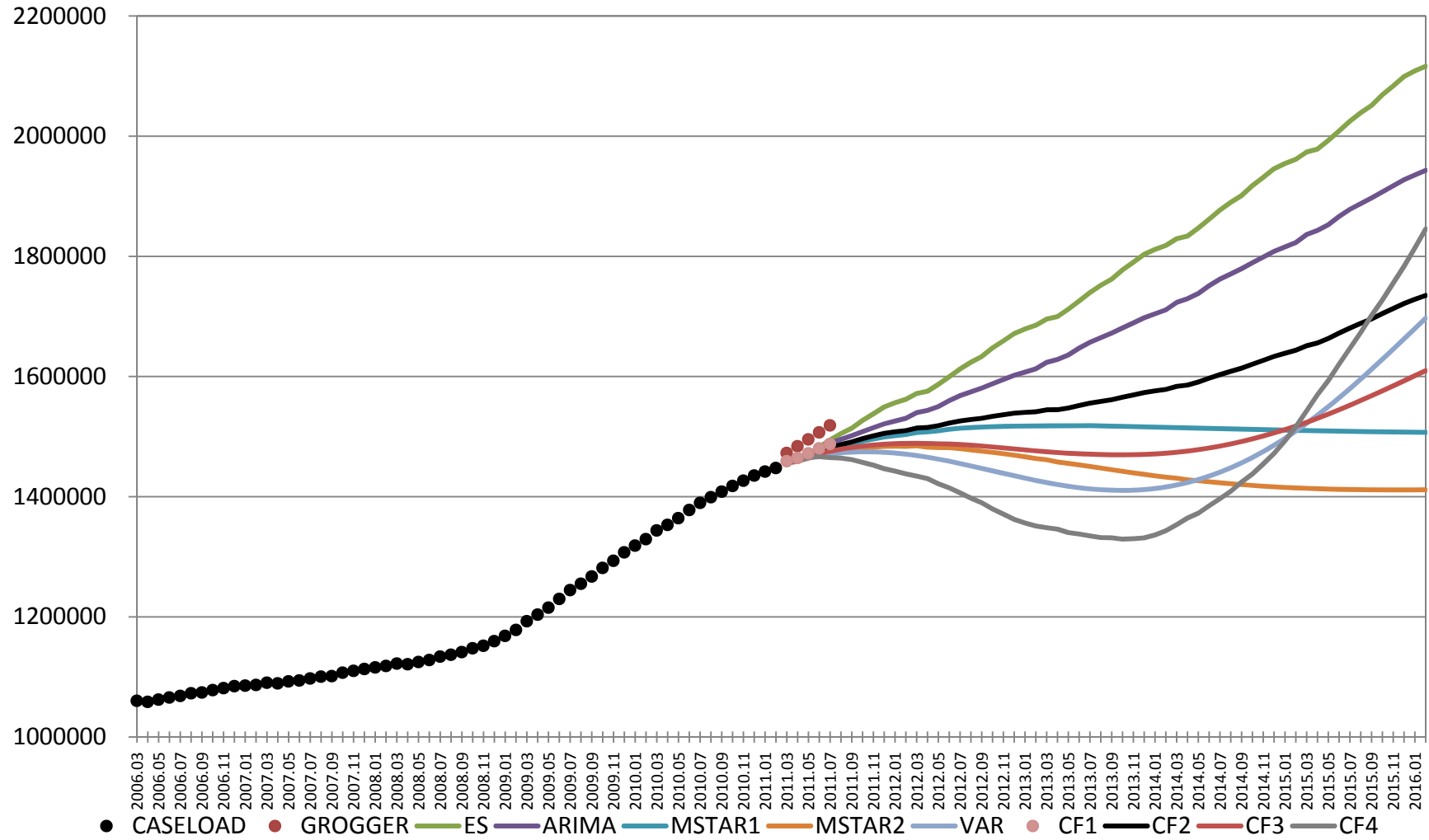
**Figure 4. MF: Out-of-sample Fitting**



**Figure 5. MF: Within-sample Fitting**



**Figure 6. Real-time Forecasting**



**Table 1. Basic Models of ES in Error Correction Form**

Seasonal \ Trend	No Seasonality	Additive	Multiplicative
No Trend	$S_t = S_{t-1} + \alpha u_t$ $c_t(m) = S_t$	$S_t = S_{t-1} + \alpha u_t$ $I_t = I_{t-p} + \delta(1-\alpha)u_t$ $c_t(m) = S_t + I_{t-p+m}$	$S_t = S_{t-1} + \alpha u_t / I_{t-p}$ $I_t = I_{t-p} + \delta(1-\alpha)u_t / S_t$ $c_t(m) = S_t I_{t-p+m}$
Linear	$S_t = S_{t-1} + T_{t-1} + \alpha u_t$ $T_t = T_{t-1} + \alpha \gamma u_t$ $c_t(m) = S_t + mT_t$	$S_t = S_{t-1} + T_{t-1} + \alpha u_t$ $T_t = T_{t-1} + \alpha \gamma u_t$ $I_t = I_{t-p} + \delta(1-\alpha)u_t$ $c_t(m) = S_t + mT_t + I_{t-p+m}$	$S_t = S_{t-1} + T_{t-1} + \alpha u_t / I_{t-p}$ $T_t = T_{t-1} + \alpha \gamma u_t / I_{t-p}$ $I_t = I_{t-p} + \delta(1-\alpha)u_t / S_t$ $c_t(m) = (S_t + mT_t) I_{t-p+m}$
Multiplicative	$S_t = S_{t-1} T_{t-1} + \alpha u_t$ $T_t = T_{t-1} + \alpha \gamma u_t / S_{t-1}$ $c_t(m) = S_t T_t^m$	$S_t = S_{t-1} T_{t-1} + \alpha u_t$ $T_t = T_{t-1} + \alpha \gamma u_t / S_{t-1}$ $I_t = I_{t-p} + \delta(1-\alpha)u_t$ $c_t(m) = S_t T_t^m + I_{t-p+m}$	$S_t = S_{t-1} T_{t-1} + \alpha u_t / I_{t-p}$ $T_t = T_{t-1} + \alpha \gamma u_t / (I_t S_{t-1})$ $I_t = I_{t-p} + \delta(1-\alpha)u_t / S_t$ $c_t(m) = S_t T_t^m I_{t-p+m}$

Notes:  $S_t$  is the smoothed level of the series;  $T_t$  is the trend rate;  $I_t$  is the seasonal index;  $c_t(m)$  is the forecast for  $m$  periods ahead from origin  $t$ ;  $u_t = c_t - c_{t-1}(1)$  is the one-step ahead forecast error;  $p$  is the seasonal span. The forecasts are the current smoothed value for any number of steps ahead  $m$ .

**Table 2. Estimation Results: VAR(6) for 2001 M01 to 2010 M02**

Dependent Variable		Log of Caseloads		Unemployment Rates		Log of the Elderly Ratio	
Variable	Lag	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.
Log of Caseloads	1	.897	(.102) <sup>***</sup>	.131	(.109)	.033	(.069)
	2	.238	(.138) <sup>*</sup>	.068	(.148)	.091	(.094)
	3	.113	(.138)	.109	(.148)	-.096	(.094)
	4	-.142	(.126)	-.205	(.135)	-.064	(.085)
	5	.094	(.120)	-.010	(.128)	0.063	(.081)
	6	-.218	(.096) <sup>**</sup>	-.097	(.103)	-0.024	(.065)
Unemployment Rates (%)	1	.104	(.094)	.708	(.101)	0.178	(.064) <sup>***</sup>
	2	.310	(.115) <sup>***</sup>	-.077	(.123)	-0.203	(.078) <sup>***</sup>
	3	-.410	(.121) <sup>***</sup>	-.134	(.130)	-0.086	(.082)
	4	.317	(.130) <sup>**</sup>	-.009	(.139)	-0.103	(.088)
	5	-.088	(.133)	.093	(.143)	0.123	(.090)
	6	-.108	(.106)	.011	(.114)	-0.052	(.072)
Log of the Elderly Ratio	1	-.206	(.158)	.202	(.169)	1.030	(.107) <sup>***</sup>
	2	.039	(.222)	-.041	(.238)	-0.076	(.150)
	3	.262	(.219)	-.231	(.235)	0.159	(.148)
	4	.040	(.224)	-.057	(.240)	-0.098	(.152)
	5	-.135	(.220)	.494	(.236) <sup>**</sup>	-0.046	(.149)
	6	.037	(.149)	-.371	(.160) <sup>**</sup>	0.020	(.101)
Constant		.291	(.119) <sup>**</sup>	.069	(.128)	-0.043	(.081)
S.E. of Estimate		.002		.002		.001	
SSR		.000		.000		.000	
DW		1.802		1.998		1.983	
<i>P</i> values for Granger (non-) causality tests with LA-VAR(8)							
Log of Caseloads		.000		.001		.550	
Unemployment Rates		.002		.000		.004	
Log of the Elderly Ratio		.066		.236		.000	

Notes: “Notes: \*\*\*  $p \leq .01$ ; \*\*  $.01 < p \leq .05$ , \*  $.05 < p \leq .10$ . Standard errors are in parentheses. Granger non-causality is examined using the LA-VAR.

**Table 3. Forecasting Performance: 2010 M03 to 2011 M02**

	Iterated multi-period forecasts			One-period ahead forecasting (Expanding windows)			One-period ahead forecasting (Rolling fixed windows)		
	MAE	RMSE	ME	MAE	RMSE	ME	MAE	RMSE	ME
ARIMA	6,823(5)	8,126 (5)	-6,823 (5)	6,823 (8)	8,126 (8)	-6,823 (8)	1,199 (2)	1,366 (1)	-267 (1)
ES	12,962 (8)	17,747 (8)	-12,635(8)	2,182 (4)	2,399 (3)	-1,547 (5)	2,289 (7)	2,491 (4)	-1,727 (6)
LSTAR1	1,636 (3)	1,736 (2)	1,359 (3)	1,708 (3)	2,984 (4)	-757 (3)	1,714 (4)	2,983 (6)	-793 (4)
LSTAR2	8,769 (6)	11,342 (6)	8,664 (6)	2,317 (5)	3,669 (6)	-480 (1)	2,154 (5)	3,619 (7)	-824 (5)
MF	9,572 (7)	11,515 (7)	-9,572 (7)	4,320 (7)	6,283 (7)	-3,750 (7)	4,388 (8)	4,714 (8)	1,895 (7)
VAR(6)	1,363 (2)	1,722 (1)	400 (1)	1,465 (2)	1,722 (1)	-751 (2)	1,347 (3)	1,598 (2)	-524 (3)
FC1 (simple average)	3,101 (4)	4,253 (4)	-3,101 (4)	2,510 (6)	2,984 (4)	-2,351 (6)	1,169 (1)	1,761 (3)	-373 (2)
FC2 (simple average, excluding ES)	1,316 (1)	1,795 (3)	-1,195 (2)	1,177 (1)	1,751 (2)	-827 (4)	2,221 (6)	2,602 (5)	-2,071 (8)

Note: The numbers in parentheses are the rank of the forecasting methods according to their respective loss functions.

**Table 4. DM Tests (*P* values)**

Null \ Alternative	ARIMA	ES	LSTAR1	LSTAR2	MF	VAR(6)	FC1	FC2
ARIMA	—	.988	.002	.997	.983	.001	.000	.001
ES	.012	—	.009	.018	.012	.008	.008	.008
LSTAR1	.998	.991	—	n.a.	.994	.486	.964	.549
LSTAR2	.003	.982	n.a.	—	.591	.002	.001	.002
MF	.017	.988	.006	.409	—	.005	.004	.005
VAR(6)	.999	.992	.514	.998	.995	—	.975	.588
FC1 (simple average)	1.000	.992	.036	.999	.996	.025	—	.017
FC2 (simple average, excluding ES)	.999	.992	.451	.999	.995	.412	.983	—

Note: LSTAR1 and LSTAR2 are not compared since the former nests the latter.



**Table 5. Estimation Results: VAR(6) for 2001 M01 to 2011 M02**

Dependent Variable		Log of Caseloads		Unemployment Rates		Log of the Elderly Ratio	
Variable	Lag	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.
Log of Caseloads	1	.898	(.096)***	.030	(.064)	.129	(.104)
	2	.224	(.130)*	.095	(.086)	.063	(.140)
	3	.154	(.128)	-.085	(.085)	.145	(.139)
	4	-.152	(.118)	-.071	(.078)	-.225	(.128)*
	5	.066	(.113)	.058	(.075)	-.006	(.123)
	6	-.211	(.090)**	-.025	(.060)	-.106	(.098)
Unemployment Rates (%)	1	-.193	(.151)	1.036	(.100)***	.172	(.163)
	2	.037	(.213)	-.073	(.141)	-.045	(.230)
	3	.296	(.210)	.165	(.139)	-.200	(.227)
	4	.030	(.214)	-.099	(.142)	-.044	(.231)
	5	-.182	(.210)	-.051	(.140)	.480	(.228)**
	6	.053	(.142)	.013	(.094)	-.373	(.154)**
Log of the Elderly Ratio	1	.075	(.088)	.166	(.058)***	.710	(.095)***
	2	.304	(.109)***	-.196	(.072)***	-.070	(.118)
	3	-.390	(.115)***	-.093	(.076)	-.130	(.124)
	4	.303	(.122)**	-.110	(.081)	-.017	(.132)
	5	-.110	(.124)	.119	(.083)	.089	(.135)
	6	-.076	(.097)	-.038	(.064)	.020	(.105)
Constant		.345	(.107)***	-.018	(.071)	-.004	(.116)
S.E. of Estimate		.002		.001		.002	
SSR		.000		.000		.000	
DW		1.815		1.989		1.984	

Notes: \*\*\*  $p \leq .01$ ; \*\*  $.01 < p \leq .05$ , \*  $.05 < p \leq .10$ . Standard errors are in parentheses.

**Table 6. Regression Weights**

	Coeff. "weight"	S.E.	<i>P</i> values
Constant	-16,395	32,274	.630
ARIMA	.991	.234	.005
ES	-.722	.336	.075
MSTAR1	-.719	.808	.408
MSTAR2	-.234	.592	.706
VAR	1.696	.442	.009
$R^2$	.9999		
$\bar{R}^2$	.9998		
<i>N</i>	12		

**Table 7. Forecast Errors**

	Actual values	Forecasts (FC2)	Absolute errors (AE)	AE in %
2011 M03	1,458,583	1,456,547	2,036	0.14%
2011 M04	1,462,186	1,460,642	1,544	0.11%
2011 M05	1,471,257	1,467,399	3,858	0.26%
2011 M06	1,479,611	1,475,151	4,460	0.30%
2011 M07	1,486,338	1,481,679	4,659	0.31%
2011 M08	1,493,228	1,486,839	6,389	0.43%
2011 M09	1,497,328	1,491,265	6,063	0.40%
2011 M10	1,502,316	1,496,682	5,634	0.37%
2011 M11	1,507,940	1,501,168	6,772	0.45%
2011 M12	1,513,448	1,505,513	7,935	0.52%
2012 M01	1,516,995	1,508,070	8,925	0.59%
2012 M02	1,521,480	1,510,055	11,425	0.75%
2012 M03	1,528,377	1,514,258	14,119	0.92%