IPSS Discussion Paper Series

(No.2011-E03)

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

Masayoshi Hayashi
(The University of Tokyo)

April 2012

国立社会保障・人口問題研究所

National Institute of Population and Social Security Research
Hibiya-Kokusai-Building 6F 2-2-3 Uchisaiwai-Cho
Chiyoda-ku Tokyo, Japan 100-0011

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

Masayoshi Hayashi

Faculty of Economics, The 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 (DI)

Fax: +81-3-5841-5521

Abstract

Forecasting welfare caseloads has become more important than ever in Japan. One reason is the magnitude of the recent increase in its welfare caseloads. Given that most previous studies only concern US cases and have not exploited recent developments in the literature, this study employs several methods (exponential smoothing, ARIMA, LSTAR, VAR, and a set of forecast combinations) to forecast Japanese welfare caseloads and compare their performances. While a VAR model and a forecast-combination model tend to outperform the other methods in pseudo real-time forecasting, a simple average forecast-combination method appears to outperform the other methods in real-time forecasting. In particular, the method predicts that PA caseload in Japan would surpass 1.7 million by the beginning of 2016, an approximately 20% increase from that at the beginning of 2011.

Keywords*:* Public assistance, welfare caseloads, forecast combinations, Japan

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

[*] This is an output of the project “Empirical Studies for Relations between Socio-economic and Human Resources of Households and Functions of Social Security (the National Institute of Population and Social Security Research, 2009-2011).”

1. Introduction

Forecasting welfare caseloads has become more important than ever in Japan. One reason is the magnitude of the recent increase in Public Assistance (PA) caseload in the country.[1] Figure 1 shows the total monthly PA caseload volume between 2001 M01 and 2011 M02, and Figure 2 exhibits their growth rate, along with the monthly entries and exits. PA caseloads have been growing since the early 1990s. In particular, its growth accelerated after 2008, reaching the current record high of 1.5 million. While the caseload levels have been as low as approximately 750,000 at the beginning of this century, the decade that followed saw this number double. Rapidly increasing caseloads may cause logistical problems for eligibility assessments and service delivery operations by welfare agencies. Equally importantly, it could entail financial problems in the appropriations process at different levels of the government. In Japan, local governments implement PA programs according to nationally uniform rules, described as follows.[2] First, the recent trend in PA caseloads has increased the share of PA expenditure in local expenditures. In some urban municipalities, PA expenditure has now amounted to as high as 20% of their total expenditures. Second, since the central government fiscally supports local PA programs through both categorically matching subsidies as well as general grants, the rapid increase also affects the central budgeting process.[3] Especially given the ongoing pressure to reduce public spending and the

[1] This paper uses “social assistance” as a generic term for needs-based and tax-financed programs that aim to maintain the minimum costs of living of their recipients. On the other hand, “Public Assistance (PA)” here refers to a specific social assistance scheme in Japan. The Japanese PA covers all types of households including the elderly, single mothers, disadvantaged, and injured/sick who are considered unable to earn, through their own best efforts, incomes above the basic costs of living. PA thus aims to guarantee that all citizens maintain their basic costs of living. As PA benefits supplement what individuals cannot earn, caseworkers conduct means tests to assess the eligibility of applicants.

[2] The Japanese local system consists of two levels of government, with municipalities (cities, towns, and villages) as the first tier and prefectures as the second tier. Cities implement PA programs through their welfare offices. While a small number of towns and villages set up their own welfare offices to implement PA, they are not required to set up such offices. Prefectural welfare offices cover residents in towns and villages that do not have their own welfare offices.

[3] The Central Government Subsidy (CGS) matches PA benefits at 75%, and the Local Allocation Tax (a general-purpose grant) accounts for the costs of the PA programs not covered by the CGS (the remaining

current trends of offloading responsibilities from the central government to localities in Japan, the accurate forecasting of PA caseloads should thus help the budgeting process both at the central and local levels.

Figures 1 and 2

The literature on welfare caseload has two strands. The first stand has extensively explored the determinants of welfare caseloads in the US. While most of these studies have examined the effects of economic factors such as unemployment (e.g., Blank 2001; Ziliak et al. 2001), many have also explored the effects of institutional schemes, which includes such determinants as state demonstration programs (Schiller and Brasher 1993; Johnson et al. 1994), waivers from Aid to Families with Dependent Children (AFDC; Schiller 1999), and the introduction of the Temporary Assistance for Needy Families (TANF; Moffit 2003). Meanwhile, other studies have examined other factors such as benefits levels (Shah and Smith 1995), local labor markets (Hill and Murray 2008), and minimum wages (Page et al. 2005). There are also analogous studies for other countries such as Canada (Spindler and Gilbreath 1979), Sweden (Gustafsson 1984), Spain (Ayala and Pérez 2005), and Japan (Suzuki and Zhou 2007).

The second strand of studies has attempted to forecast welfare caseloads to inform policy makers in budgeting and operation (Plotnick and Lidman 1987; Albert and Barth 1996; Opitz and Nelson 1996; Conte et al. 1998; Grogger 2007; Gurmu and Smith 2008; Lazariu et al. 2010), to which the current study belongs. However, this strand has two shortcomings. First, it has only focused on US cases. Since forecasting welfare caseloads is also an integral part of the oversight of social assistance programs in other countries, studies for these other countries should contribute to the literature. Second, it has not exhausted the recent developments in forecasting methods. While there have been extensive studies on the comparative forecasting performance of different methods

25% of PA benefits plus the costs of welfare personnel and welfare office maintenance).

in other variables such as unemployment and inflation, there have not been analogous studies on welfare caseloads with the exception of Lazariu et al. (2010), who only considered ARIMA and VAR models. Forecasting caseloads with the other methods should thus contribute to the prediction of future levels of welfare caseloads.

This study aims to improve on these previous studies by applying a variety of forecasting methods to the PA caseloads in Japan. The caseloads in the sample have a monthly frequency and include the period 2001 M01 to 2011 M02, which are then split into a regression sample (2001 M01 to 2010 M02) and a validation sample (2010 M03 to 2011 M02). In this study, I apply several methods as if I were actually forecasting using the regression sample, the results of which are then compared to that using the validation sample. I anticipate that while a VAR model and a forecast-combination method will tend to outperform other methods in pseudo real-time forecasting, a simple average forecast combination will perform well in real-time forecasting. The method predicts that PA caseloads would reach at least 1.7 million by the end of 2015, a more than 20% increase from 2011.

The remaining parts of the paper are structured as follows. In Sections 2 and 3, I introduce the forecasting methods I utilize in this study. In particular, Section 2 introduces and applies the Markov forecasting by Grogger (2007), while Section 3 discusses and estimates additional, more popular models, namely an autoregressive integrated moving average (ARIMA) model, two logistic smooth transition autoregressive (LSTAR) models, a vector autoregression (VAR) model, and two variants of forecast combinations. Section 4 utilizes these methods to generate the PA caseload forecasts, and conducts a forecast evaluation in the validation period. In addition, it also conducts a real-time forecasting exercise for the period beyond 2011 M02. Section 5 then presents the conclusions of the paper.

2. Markov Forecasting

Grogger (2007) proposes a method designed for welfare caseload forecasting. Since his method is less popular than the other methods in the next section, I dedicate the current section to introducing his innovative method, called the Markov forecasting, which I then apply to PA caseloads in Japan. The method exploits the fact that current caseload C_t equals to previous caseload C_{t-1} plus entries E_t net of exits X_t, that is, $C_t \equiv C_{t-1} + (E_t - X_t)$, or

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

where $x_t \equiv X_t/C_{t-1}$ is the exit rate. Grogger regards Eq. (1) as a first-order Markov chain that depends on current exits rates and entries, along with the first lag of the caseload. In a steady state where the exit rate and the entries are constant ($x_t = x$, and $E_t = E$), Eq. (1) converges to $\bar{C} = E / x$. This then implies that *if* the exit rates x_t and entries E_t were held constant in the coming periods, the future caseload would converge to

$$\bar{C}_t = E_t / x_t \qquad (2)$$

which Grogger refers to as "implied steady state (ISS)."

Assuming that the ISS constitutes a leading indicator of future caseload levels, Grogger uses the following regression model to forecast future caseloads:

$$C_t = \alpha_0 + \alpha_1 \cdot \bar{C}_{t-L} + u_t \qquad (3)$$

where u_t is an error term and αs are coefficient parameters. In Eq. (3), the ISS in period t predicts the caseload in period $t + L$. However, when calculated with raw monthly data, the ISS exhibits a large volatility. Figure 3 shows the case for the Japanese PA caseload in the first panel. Grogger argues that this volatility contains month-to-month noises. To separate such noises from the underlying information that helps predict caseloads, he conducts a locally weighted scatterplot smoothing (LOWESS) on the monthly series of exit rates and entries, and constructs a smoothed ISS as a ratio of the smoothed entries to the smoothed exit rates. Thus, the actual regression model for the estimation is

$$C_t = \alpha_0 + \alpha_1 \cdot \hat{C}_{t-L}(b_E, b_x) + u_t. \tag{4}$$

where $\hat{C}_t(b_E, b_x) \equiv E_t(b_E)/x_t(b_x)$ is the smoothed ISS, $x_t(b_x)$ and $E_t(b_E)$ are the smoothed x_t and E_t, and b_x and b_E are their respective bandwidths. Grogger suggests performing an ordinary least squares (OLS) estimation on Eq. (4) for the given values of b_E, b_x, and L, which together minimize the mean square error of forecasts (MSE), with the use of a three-dimensional grid search. I conduct the grid search for the three parameters, with L from 10 to 36 by 1, and b_x and b_E from .1 to .9 by .01.[4] The MSE uses one-period-ahead forecasts from rolling regressions with a fixed window of 110 monthly observations (see Section 4). The regression and validation samples are respectively from 2001 M01 to 2010 M02 (110 observations) and from 2010 M03 to 2011 M02 (12 observations).[5] I then obtain 5 for L, .5 for b_x, and .8 for b_E. The three panels in Figure 3 show the monthly series for the smoothed ISS, exit rates, and entries, along with their original series represented by the dots. Since the two bandwidths are rather large, the LOWESS curves do not fit the observed data well.

Figure 3

I then estimate Eq. (4) using OLS and obtain the model:

$$\underset{(31710)}{C_t = 309171} + \underset{(.023)}{.522} \times \hat{C}_{t-5} + u_t, \quad \bar{R}^2 = .826, \quad N = 110 \tag{Markov}$$

where the figures in the parentheses are standard errors. Note that the forecasts are rescaled values of the smoothed ISS pushed five periods ahead. Figure 4 describes the relations among the rescaled smoothed ISS, forecasts (the rescaled smoothed ISS shifted to the right by five periods), and actual caseloads. I anticipate that the

[4] Grogger originally set the grid ranges from 10 to 15 for L by 1 and from .1 to .8 for b_E and b_x. These ranges, however, resulted in corner solutions. I thus expand the ranges as in the text.

[5] The size of my sample is almost half of Grogger's, which consist of observations from 1985 M06 to 2002 M11 (209 observations) for the estimation, and from 2003 M12 to 2005 M03 (28 observations) for the evaluation of the forecasts.

within-sample fits will be poor when compared to the forecasts that I describe in Section 4. In addition, the within-regression sample fits are poorer, since I smoothed the ISS so that the MES is minimized without considering the fits within the regression sample. To articulate this point, I again perform a grid search for b_x, b_E, and L so that they, in this case, minimize the sum of squared residuals from the regression sample, and obtain 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. The resulting model is

$$C_t = \underset{(22390)}{173} + \underset{(.018)}{.887} \times \hat{C}_{t-33} + u_t, \quad \bar{R}^2 = .961, \quad N = 93. \tag{5}$$

Here, the fitted values in period t are rescaled values of the smoothed ISS in period t – 33. I then plot these fitted values against the actual caseloads in Figure 6. This time, while the constant is statistically insignificant, the smoothed ISS appears to be a good leading indicator in the regression period. However, the forecasts do not perform well in the validation period. This is a typical example of the within sample overfitting which results in poor out-of-sample (pseudo real-time) forecasts.

Figures 4 and 5

The strength of Markov forecasting lies in its ability to detect a turning point in the welfare caseload trajectory (Grogger 2007; Gurmu and Smith 2008). However, the PA caseloads here do not exhibit any turning points, which may be the reason why the method does not perform well. Nonetheless, if the ISS is a good leading indicator, it should have detected, for example, changes in the pace of caseload growth toward the end of the validation period. However, this is not the case here.

3. ES, ARIMA, LSTAR, VAR, and Forecast Combinations

3.1. Exponential Smoothing

Exponential smoothing (ES) is a popular method of forecasting (Gardner 1985, 2006). In its very basic formulation, ES has nine types of specifications depending on how we model seasonality and linear trends. Table 1 lists these specifications in error-correction form, where S_t is the smoothed level of the series, T_t is the trend rate, I_t is the seasonal index, u_t is the forecast error, and p is the seasonal span. I obtain these parameters by minimizing the sum of the squared errors. I select a specification among the nine based on the value of the Bayesian information criterion (BIC). Table 2 lists the results. The linear-trend model without seasonality yields the minimum value of the BIC, resulting in

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

where S_t is the log of caseload in period t ($S_t \equiv \ln C_t$).

Tables 1 and 2

3.2. ARIMA Model

Modeling welfare caseload evolution as an ARIMA process is a standard method (Albritton 1979; Klassen 1997; Chang 2007; Gurmu and Smith 2008; Lazariu et al. 2012). I follow the standard Box-Jenkins identification procedure. Figure 6 provides the correlogram for seasonally differenced first-order differences of the natural log of PA caseload ($\Delta\ln C_t - \Delta\ln C_{t-12}$). The autocorrelation function shows little sign of non-stationarity, and the partial autocorrelation function implies autoregression of the third order, which together imply that the process is ARIMA$(3,1,0)\times(0,1,0)_{12}$, where the three elements in the latter parentheses respectively indicate *seasonal* AR lags, the degree of integration, and MA lags. Indeed, a grid search for p and q in ARIMA$(p,1,q)\times(0,1,0)_{12}$ yields both the minimal Akaike information criterion (AIC) and BIC as $p = 3$ and $q = 0$. Table 3 lists the estimated coefficients and their standard errors, while Figure 7 shows the correlogram of the residuals with little indication of autocorrelation of residuals.

Table 3 and Figures 6 and 7

3.3. LSTAR Models

A threshold autoregression (TAR) model combines different autoregression models (branches) with a trigger variable that determines which branch applies. The branches may or may not share the same lag structures. Typically, but not necessarily, a TAR model uses the lag of the time series it explains as a trigger. Meanwhile, a *smooth* transition autoregression (STAR) model specifies the smooth transition from one branch to another, with a function $F(\cdot)\in[0, 1]$. Meanwhile, the logistic STAR (LSTAR) uses the logistic function for $F(\cdot)$ (e.g., Teräsvirta 1994). While the STAR framework has been used to model non-linearity in macro-economic variables such as exchange rate (e.g., Sarantis 1999) and inflation (i.e., Byers and Peel 2000), the model has never been applied to the analysis of welfare caseloads. It may thus be worthwhile to apply the LSTAR model to forecast welfare caseloads here.

I first consider an LSTAR model where two branches have the same lag structure (LSTAR1). With some diagnostics, I formulate and estimate LSTAR1 as

$$\begin{aligned}\ln C_t =\ & \underset{(.029)}{.105} + \underset{(.075)}{.828}\ln C_{t-1} + \underset{(.081)}{.180}\ln C_{t-2} + \underset{(.078)}{.627}\ln C_{t-12} - \underset{(.070)}{.643}\ln C_{t-13} \\ & +1/\{1+\exp[\underset{(10.401)}{27.035}\times(\ln C_{t-10}-\underset{(.008)}{13.933})]\}\times(\underset{(2.398)}{.656} + \underset{(.396)}{.227}\ln C_{t-1} \\ & - \underset{(.356)}{.322}\ln C_{t-2} - \underset{(.363)}{.166}\ln C_{t-12} + \underset{(.514)}{.215}\ln C_{t-13}) + u_t \\ & \bar{R}^2 = .999, \quad N = 110.\end{aligned} \qquad \text{(LSTAR1)}$$

where the figures in parentheses are standard errors. However, the last five parameter estimates in the second branch are all statistically insignificant. I then estimate a restricted version of LSTAR1 (LSTAR2) as

$$\begin{aligned}\ln C_t =\ & \underset{(.029)}{.098} + \underset{(.072)}{.851}\ln C_{t-1} + \underset{(.078)}{.157}\ln C_{t-2} + \underset{(.075)}{.603}\ln C_{t-12} - \underset{(.068)}{.618}\ln C_{t-13} \\ & +1/\{1+\exp[\underset{(8.603)}{38.297}\times(\ln C_{t-10}-\underset{(.002)}{13.930})]\}\times(-\underset{(.068)}{.618} + \underset{(.396)}{1.103}\ln C_{t-1}) + u_t. \qquad \text{(LSTAR2)} \\ & \bar{R}^2 = .999, \quad N = 110\end{aligned}$$

With this restricted version, all of the coefficients are statistically significant.

I use both LSTAR1 and LSTAR2 to generate the forecasts for PA caseloads. However, since the path of the forecasts of an LSTAR model depends on the initial situation (e.g., the distance of the shock from a transition point) due to its non-linearity, the non-stochastic calculation of a series of the forecasts simply maps out just one of a number of possible forecast paths. I thus conduct random simulations to replicate 10,000 forecast paths and average them to produce the point forecasts of the two LSTAR models.

3.4. VAR model

I then examine a vector autoregression (VAR) model. The model includes PA caseloads, unemployment rates, and elderly ratio (the ratio of those aged 65 years and above to the total population). While including the unemployment rate is standard (Smith 1991; Lazariu et al. 2012), including the elderly ratio may be specific to the Japanese case. This is because Japanese PA also covers the elderly population, and the recent increase in caseloads partially reflects the increasing number of the elderly who have little or no pension benefits (Hayashi 2010). I take the natural logs of the caseloads (as in the other models except the Markov forecasting) and the elderly ratio as they are both trending variables, and leave the unemployment rates in level (percentage) as they are non-trending. Note that I do not difference the caseload data despite the fact that the ARIMA analysis implied that the series is non-stationary. In ARIMA modeling, appropriate differencing is important since it is impossible to identify the stationary structure of the process using the autocorrelations of an integrated series, and most algorithms used for fitting ARIMA models fail when confronted with integrated data. Neither applies to VAR models. In addition, even if variables are non-stationary and/or co-integrated, the OLS estimators of the VAR coefficients are consistent (Hamilton 1994).

I select the lag p of VAR(p) from 1 to 13 to minimize the MSE in the validation sample. This, in effect, is a performance comparison among a symmetric class of VAR models with the different lag lengths. The model with p = 6 surpasses the other alternatives. Table 4 lists the estimation results for VAR(6). The bottom of the table lists the P values for the Granger causality tests, which employ the lag-augmented VAR (LA-VAR) by Toda and Yamamoto (1995). I use LA-VAR(8), allowing for possible integrated or co-integrated variables for up to the second order. The tests imply that the elderly ratio and PA caseload do not Granger-cause the unemployment rate and elderly ratio, respectively. Nonetheless, I do not exclude them when I forecast the PA caseloads, since doing so weakens the fits of the forecasts.

3.5. Forecast Combinations

Lastly, I perform forecast combinations, which combines the forecasts generated by all or a subset of the six models (i.e., Markov forecasting, ES, ARIMA, LSTAR1, LSTAR2, and VAR). I perform the forecast combinations because even when a model performs worse than the other models, it may not be clear if it is appropriate to ignore the former model completely. We may best view a forecasting model as a simple approximation of a more complicated and/or constantly changing data generating process. If so, such a view implies that a forecasting model is necessarily misspecified. If forecasts from multiple forecast models are available, therefore, combining their forecasts may diversify forecasting errors otherwise not possible from a single forecast model. In fact, the literature shows that combining forecasts from different models outperform forecasts from a single model (Elliott and Timmermann 2008; Aiofli et al. 2011). An issue with this strategy, however, is how to weigh the multiple forecasts.

In this study, I first use a simple average, which the literature shows often outperforms other weights that are deliberately designed (Elliot and Timmermann 2008).

I call the combined forecasts with the equal weights as Combined Forecasts 1 (CF1). In addition, one may improve forecast performance by trimming the group of the models for combination to a subgroup of the better performing models (Aiolfi and Timmermann 2006). For the current case where there are only six models, I exclude only the model that performs worst. I call the simple-averaged forecasts based on the trimmed group as Combined Forecasts 2 (CF2).

4. Performance Comparison

4.1. Pseudo real-time forecasting

I use data for PA caseloads from 2001 M01 to 2011 M02 (T = 122), which I divide into two sub-samples: a regression sample from 2001 M01 to 2010 M02 (N = 110) and a validation sample from 2010 M03 to 2011 M02 (P = 12). I calculate the forecasts for the validation periods to perform *pseudo real-time forecasting* (Elliot and Timmermann 2008), that is, I generate the forecasts f_t for each month in the validation sample as if I were actually forecasting in real time. For each of the forecasting methods (i.e., Markov, ARIMA, LSTAR1, LSTAR2, VAR, CF1, and CF2), I generate three types of forecasts. First, I recursively generate forecasts ($f_{N+1|N}, \ldots, f_{T|N}$) from *a fixed origin t* = N (i.e., 2010 M02) for all the P periods. The *original* regression sample generates forecasts with a *fixed* window. Figure 8 describes their forecasts. Note that this is analogous to the procedure forecasters use in *real time*. Second, I generate one-period-ahead forecasts $f_{t+1|t}$ for t > 2010 M02 using samples with an *expanding window*. I estimate the forecasting model for $f_{t+1|t}$ with *all* the data available from 2001 M01 up to t. I repeat this process P times to generate a series of P one-period-ahead forecasts ($f_{N+1|N}, f_{N+2|N+1}, \ldots, f_{T|N+P-1}$). The regression sample thus *expands* by one observation every time I update the forecast for the next period. Third, I use samples with a *rolling window*. This is analogous to the case of an expanding window *except* that I drop the earliest observation when I update an additional forecast, thereby using the fixed size of observations of 110 months that precede period $t + 1$, that is, the sample for obtaining $f_{N+h|N+h-1}$ consists of $C_{1+h}, \ldots, C_{N+h-1}$.

Figure 8

I then compare the series of forecasts to the observations of PA caseloads in the validation sample. For each of the three modes of forecasting (fixed, expanding, and rolling), I obtain the mean absolute error (MAE = $\sum|C_t - f_t|/P$), root mean squared errors (RMSE = $[\sum(C_t - f_t)^2/P]^{1/2}$), and mean error (ME = $\sum(C_t - f_t)/P$) to evaluate the performance of the forecasting methods. The ME may be less popular for this type of evaluation. However, since budget planning is usually performed annually, large month-to-month errors may not be a serious problem as long as they average to zero within a year (12 months). The ME indeed captures such averaging. Note that I calculate MAE, RMSE, and ME using caseloads in level. Since the five methods except the Markov forecasting transform the caseload into log, I reverse the transformation to obtain the forecasted PA caseload in level.

Table 5 lists the MAE, RMSE, and ME. First, for the fixed sample evaluation. ES performs the worst, followed by Markov forecasting. The best model is either CF2 (a simple average that excludes exponential smoothing) with MAE, or VAR with both RMSE and ME. LSTAR1 fares relatively well, being either the second best or third best model. However, possibly in contrast to a common expectation, CF1 (a simple average of the six methods) shows mediocre performance, which may be due to the fact that I only deal with six forecasting methods. Second, the expanding-window evaluation changes the ranking. In particular, the ARIMA model now performs worst in terms of all the three measures, although the Markov forecasting remains the second worst performer as before. Note that the VAR model and CF2 continue to perform well. Third, the rolling-window evaluation gives another different picture. Here, the ARIMA model is the best model in terms of RMSE and ME, and the second best in terms of MAE. In addition, possibly reflecting the performance of the ARIMA model, CF1 ranks the first, second, and third, respectively, in terms of the MAE, ME, and RMSE. However, the VAR still fares relatively well, being the second best in terms of the RMSE, and the third best in terms of the MAE and ME. Overall, the forecast evaluation for the period 2010 M03 to 2011 M02 indicates that the VAR and one of the combined forecast models perform consistently well in each type of forecast. The good performance of the VAR model is consistent with the previous study on welfare caseload forecasting by Lazariu et al. (2011). The results of the combined forecasts also parallel the findings of Aiolfi et al. (2011), who state that "even if forecast combinations do not always deliver the best forecasts, they do not generally deliver poor performance, thus from 'risk' perspective they represent a safe choice."

Table 5

I next compare more formally a pair of forecasts by conducting the Diebold-Mariano (DM) test (Diebold and Mariano 1995). In this comparison, it is important to note the following points. First, I base the test statistics on the RMSE. Second, I utilize only the forecasts generated by the fixed sample method. This is because this study is interested in *real-time* forecasting, that is, obtaining a series of forecasts and exhausting all the currently available data to determine how the caseload would evolve over the coming months, which I perform for the period beyond 2011 M02, as shown below. Third, I do not compare between LSTAR1 and LSTAR2 since LSTAR1 nests LSTAR2.[6] Table 6 shows the *P* values of the DM tests. 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 CF2 are not rejected against

[6] If we apply the DM test to a pair of models in which one nests the other, the distribution of the test statistics will be non-standard.

any of the alternative models, barring the fact that LSTAR1 is not tested against LSTAR2. Thus, setting aside LSTAR1, VAR and CF2 are among the best models, although their differences are not identified statistically. Note that Ashley (2003) argues that more than 100 forecasts are necessary to establish significant differences in predictive accuracy across models. The results in Table 6 should thus be taken with caution since they rely on only 12 forecast values. However, as all of the *P* values for the well-performing models are close to one except when they are compared among themselves, I do not necessarily have to be too cautious on this point.

Table 6

4.2. Real-time forecasting

I finally conduct a real-time forecast for the period beyond 2010 M02. In this forecast, it is important to note the following. First, the Markov forecasting can only provide forecasts up to 2010 M07 since it is based on the smoothed ISS with a lag of five (See Eq. 5). In addition, I exclude the Markov forecasting when I construct CF2 for the real-time forecasting since the five forecast values for the methods implies the worst forecasting values, as shown below. Second, since I can use the information from the validation sample, I can now exploit two more types of more elaborate forecast combinations with weights that require information ex-ante to the forecast origin (2010 M02). Note that I exclude the forecasts from the Markov forecasting again when I construct the following two more elaborate weights.

One of the weights comes from inverse mean-squared-errors weighting (Bates and Granger 1969). This scheme, which I call CF3, weighs forecasts made by the *j*th forecasting method with

$$\omega_j = \frac{1/MSE_j}{\sum_i 1/MSE_i}. \quad \text{(CF3)}$$

I use the MSE calculated from the RMSE listed in the third column in Table 5 (excluding the Markov forecasting).

The other weight (CF4) comes from a scheme by Granger and Ramanathan (1984), which takes advantage of the estimates from the following regression model:

$$C_t = \alpha + \sum_j \beta_j \cdot f_{jt} + \text{residuals} \quad \text{for } N < t \leq T. \qquad \text{(CF4)}$$

The five estimated coefficients, in addition to the constant, rescale their corresponding forecasts generated beyond 2011 M02. By construction, these "weights" do not generally add up to unity. As shown in Table 7, the fit of the regression is very good (R^2 = .9999 and $\bar{R}^2$ = .9998). However, these "weights" turn out to be quite different from what we usually think are weights. First, as expected, they do not add up to unity, although, when excluding constant α, the sum of the six coefficients is close to one (1.012). Second, however, 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).

Table 7

I then conduct real-time forecasting to predict the PA caseloads for the period beyond 2011 M02 up to 2016 M02 (i.e., 5-year or 60-month horizon). Using all the available data from 2001 M01 to 2011 M02, I estimate the six forecasting models (i.e., Markov, ES, ARIMA, LSTAR1, LSTAR2, and VAR) as specified in the previous sections, and four forecast combinations (CF1, CF2, CF3, and CF4) as described above. Figure 9 describes their forecasts.

Figure 9

Not surprisingly, each of the individual (i.e., non-combined) forecasts behaves erratically in the longer term. Furthermore, although their behaviors are more or less similar in the validation period (recall Figure 8), they now are quite different. The Markov forecasting, which has only five forecast values, appears to predict the largest level of caseloads for the limited periods of forecasting. The ES also exhibits constantly increasing caseloads that amount to well beyond 2 million toward the end of 2015. Likewise, the ARIMA shows another case of increasing caseloads, which, while not as sharp as that of the ES, is also substantially high at just below 2 million by the end of the period. On the other hand, although the forecasts of the two LSTAR models slowly increased in the early periods, they declined over time. This would be implausible given the current trends in socio-economic factors affecting PA in Japan, such as the rapidly aging population and the stagnant economy. The VAR exhibits another implausible movement. At the earlier stages, its forecasts started decreasing more rapidly than those of the two LSTAR models, then started increasing at the beginning of 2014. This may be striking since the VAR model performed quite well in the pseudo real-time forecasting. These erratic and different behaviors of the individual forecasts are quite different from those in the pseudo real-time forecasting. 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 forecast combinations. Since the Markov method can only generate forecasts up to 2011 M07, recall that I excluded its forecasts from CF3 and CF4. I only calculate CF1 up to 2011 M07, which may not merit discussion here. As for CF2, I drop the forecasts from the Markov model as it appears to be the worst performing model, as shown in Figure 9. I thus construct CF2 as a simple average of the five remaining models. I also exclude the forecasts from the Markov forecasting from CF3 and CF4. It may be surprising that these two more elaborate forecast combinations behave more erratically than the forecasts from some of the individual models. In particular, the forecasts by CF4 starts to decrease as early as in 2011 and continue to do so until mid-2013, and then increases rapidly so that it exceeds the forecasts made by CF2 for 2015. This is because the CF4 has the largest “weight” of

1.696 on the forecasts by the VAR model, and the VAR model behaves quite erratically, both as described above. The behavior of the CF3 forecasts is less erratic when compared to those of CF4. However, it still exhibits a decrease that starts in mid-2012.

The forecasts by CF2, a simple average of the forecasts by all methods except Markov forecasting, appear to be a safe choice in that it demonstrates the most plausible future path of the PA caseloads given the available information that would affect caseloads in the future, including the expected rapid aging of the population, continuing stagnant economy, and changing labor market practices in the country. The caseload forecasts by CF2 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, by the time I completed this forecast calculation (i.e., 2012 M03), the data on PA caseloads have become available to the public for up to 2011 M12. While these values are *not finalized* except those for 2011 M03, the forecasts by CF2 appear to perform well when compared to the other forecasts. While they underestimate the reported values, the monthly differences are only 2,036 (.14%), 1,555 (.11%), 3,858 (.26%), 4,460 (.30%), 4,662 (.31%), 6,391 (.43%), 6,064 (.40%), 5,638 (.38%), 6,772 (.45%), and 7,933 (.52%) from 2011 M03 to 2011 M12. Considering the percentage errors in the parentheses are all 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 may constitute another case of a "forecast combination puzzle," where simple combinations of point forecasts repeatedly outperform other sophisticated forecast combinations (Smith and Wallis 2009).

5. Concluding remarks

Forecasting welfare caseloads has become more important than ever in Japan due to the recent rapid increase in its PA caseloads. Given the fact that most previous studies only concern US cases and have not exploited recent developments in the literature, this study employed several forecasting methods (ES, ARIMA, LSTAR, VAR, and a set of forecast combinations) to predict Japanese PA caseloads and compared their performances. This study showed that while a VAR model and one of the forecast-combination methods tend to outperform the other methods in pseudo real-time forecasting, a simple average forecast combination appears to outperform them in real-time forecasting. In particular, the simple average forecasting predicts that PA caseload in Japan would surpass 1.7 million by the end of the forecast period (2016 M02), approximately a 20% increase from the beginning of 2011.

References

Aiolfi, M., Capistrán, C., Timmermann, A., 2011. Forecast combinations. In: Clements, M.P., Hendry, D.E., (Eds.) The Oxford Handbook of Economic Forecasting. Oxford University Press, Oxford, 355–388.

Aiolfi, M., Timmermann, A., 2006. Persistence in forecasting performance and conditional combination strategies. Journal of Econometrics 135(1/2), 31–53.

Albert, V., Barth, R.P., 1996. Predicting growth in child abuse and neglect reports in urban, suburban, and rural counties. Social Science Review 70(1), 58–82.

Albritton, R.B., 1979. Measuring public policy: Impacts of the Supplemental Security Income Program. American Journal of Political Science 23(3), 559–578.

Ayala, L., Pérez, C. 2005. Macroeconomic conditions, institutional factors and demographic structure: What causes welfare caseloads? Journal of Population Economics, 18(3), 563–581.

Armstrong, J.S., 2001. Combining forecasts. In: Armstrong J.S. (Ed.) Principles of Forecasting: A Handbook for Researchers and Practitioners. Kluwer Academic Publishers, Norwell, MA, 417–440.

Ashley, R., 1998. A new technique for postsample model selection and validation. Journal of Economic Dynamics and Control 22(5), 647–665.

Bates, J. M., Granger, C.W.J., 1969. The combination of forecasts. Operational Research Quarterly 20(4), 451–468.

Blank, R.M. (2001). What causes public assistance caseloads to grow? Journal of Human Resources 36(1), 85–118.

Byers, J.S., Peel, D.A., 2000. Non-linear inflation dynamics: Evidence from the UK. Manchester School 68, 23–37.

Chang, H.J., 2007. Explaining welfare caseload reduction in New York State: The effect of policy or economy? International Review of Public Administration 12(1), 105–117.

Conte, M., Levy, D.T., Shahrokh, F., Staveley, J., Thompson, S., 1998. Economic determinants of income maintenance programs: The Maryland forecasting model. Journal of Policy Modeling 20(4), 461–481.

Danielson, C., Klerman, J.A. (2008). Did welfare reform cause the caseload decline? Social Service Review 82(4), 703–730.

Diebold, F.X., Mariano, R.S., 1995. Comparing predictive accuracy. Journal of Business and Economic Statistics 13(3), 253–263.

Elliot, G., Timmermann, A., 2008. Economic forecasting. Journal of Economic Literature 46(1), 2–56.

Gardner, E.S., Jr., 1985. Exponential smoothing: The state of the art. Journal of Forecasting 4(1), 1–28.

Gardner, E.S., Jr., 2006. Exponential smoothing: The state of the art—Part II. International Journal of Forecasting 22(4), 637–666.

Granger, C.W.J., Ramanathan, R., 1984. Improved method of combining forecasts. Journal of Forecasting 3(2), 197–204.

Grogger, J., 2007. Markov forecasting methods for welfare caseloads. Children and Youth Services Review 29(7), 900–911.

Gurmu, S., Smith, W.J., 2008. Estimating and forecasting welfare caseloads. In: Sun, J., Lynch, T.D. (Eds.) Government Budget Forecasting: Theory and Practice. Auerbach Publications, Boca Raton, FL, 188–222.

Gustafsson, B., (1984). Macroeconomic performance, old age security and the rate of social assistance recipients in Sweden. European Economic Review 26(3), 319–338.

Hamilton, J., 1994. Time Series Analysis. Princeton University Press, Princeton.

Hayashi, M., 2010. Social Protection in Japan: Current State and Challenges. In: Asher, M.G. Oum, S, Parulian, F. (Eds.) Social Protection in East Asia: Current State and Challenges. Economic Research Institute for ASEAN and East Asia (ERIA), Jakarta, 19–54.

Hill, B.C., Murray, M.N. 2008. Interactions between welfare caseloads and local labor markets. Contemporary Economic Policy 26(4), 539–554.

Johnson, T.R., Klepinger, D.H., Dong, F.B. (1994). Caseload impacts of welfare reform. Contemporary Economic Policy 12(1), 89–101.

Klassen, T.R., 1997. Getting it backward? Economy and welfare in Ontario 1985–1995. Canadian Public Policy 23(3), 333–338.

Lazariu, V., Chengxuan, Y., Gundersen, C., 2011. Forecasting women, infants, and children caseloads: A comparison of vector autoregression and autoregressive integrated moving average approaches. Contemporary Economic Policy 29(1), 46–55.

Moffit, R. (2003). The role of nonfinancial factors in exit and entry in the TANF program. Journal of Human Resources 38(supplement), 1221–1254.

Opitz, W., Nelson, H., 1996. Short-term, population-based forecasting in the public sector: A dynamic caseload simulation model. Population Research and Policy Review 15(5/6), 549–563.

Page, M.E., Spetz, J., Millar, J. (2005). Does the minimum wage affect welfare caseloads? Journal of Policy Analysis and Management 24(2), 273–295.

Plotnick, R.D., Lidman, R.M., 1987. Forecasting welfare caseloads: A tool to improve budgeting. Public Budgeting & Finance 7(3), 70–81.

Sarantis, N., 1999. Modeling non-linearity in real effective exchange rates. Journal of International Money and Finance 18(1), 27–45.

Schiller, B.R. (1999). State welfare-reform impacts: Content and enforcement effects. Contemporary Economic Policy 17(2), 210–222.

Schiller, B.R., Brasher, C.N. (1993). Effects of workfare saturation on AFDC caseloads. Contemporary Policy Issues 11(2), 39–49.

Shah, P., Smith, P.K. (1995). Do welfare benefits cause the welfare caseload? Public Choice 85(1/2), 91–105.

Smith, J., Wallis, K.F., 2009. A simple explanation of the forecast combination puzzle. Oxford

Bulletin of Economics and Statistics 71(3), 331–355.

Smith, P.K., 1991. Welfare as a cause of poverty: A time series analysis. Public Choice 75(2), 157–170.

Spindler, Z.A., Gilbreath, W.S. (1979). Determinants of Canadian social assistance participation rates. International Journal of Social Economics 6(3), 164–1974.

Suzuki, W., Zhou, Y. (2007). Welfare use in Japan: Trends and determinants. Journal of Income Distribution 16(3/4), 88–109.

Teräsvirta, T., 1994. Specification, estimation, and evaluation of smooth transition autoregressive models. Journal of the American Statistical Association 89(425), 208–218.

Toda, H.Y., Yamamoto, T., 1995. Statistical inference in vector autoregressions with possibly integrated processes 66(1/2), 225–250.

Ziliak, J.P., Figlio, D., Davis, E., Connolly, L. (2000). Accounting for the decline in AFDC caseloads: Welfare reform or economy? Journal of Human Resources 35(3), 570–586.

Table 1. Basic Models of Exponential Smoothing in Error Correction Form

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

Table 2. Model Selection for Exponential Smoothing

Seasonality	Trend	BIC
None	None	–794.5
Additive	None	–776.1
Multiplicative	None	–776.0
None	Linear	–1,173.2
Additive	Linear	–1,169.2
Multiplicative	Linear	–1,169.1
None	Exponential	–1,168.9
Additive	Exponential	–1,164.8
Multiplicative	Exponential	–1,164.6

Table 3. Estimation Results: Dependent Variables $\Delta\ln C_t - \Delta\ln C_{t-12}$

Variable	Coeff.	S.E.	*P* values
Constant	0.001	(0.001)	0.652
$\Delta\ln C_{t-1} - \Delta\ln C_{t-13}$	0.224	(0.090)	0.014
$\Delta\ln C_{t-2} - \Delta\ln C_{t-14}$	0.258	(0.088)	0.004
$\Delta\ln C_{t-3} - \Delta\ln C_{t-15}$	0.391	(0.091)	0.000
Sample Size	110		
$\bar{R}^2$	0.999		
P-value for Q Stat.	0.11		

Table 4. Estimation Results: VAR(6)

Dependent Variable		Log of Caseloads		Unemployment Rate		Log of Elderly Ratio	
Variable	Lag	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.
Log of Caseloads	1	.897	(.102)***	.131	(.109)	.033	(.069)
	2	.238	(.138)*	.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)**	–.097	(.103)	–0.024	(.065)
Unemployment Rate (%)	1	.104	(.094)	.708	(.101)	0.178	(.064)***
	2	.310	(.115)***	–.077	(.123)	–0.203	(.078)***
	3	–.410	(.121)***	–.134	(.130)	–0.086	(.082)
	4	.317	(.130)**	–.009	(.139)	–0.103	(.088)
	5	–.088	(.133)	.093	(.143)	0.123	(.090)
	6	–.108	(.106)	.011	(.114)	–0.052	(.072)
Log of Elderly Ratio	1	–.206	(.158)	.202	(.169)	1.030	(.107)***
	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)**	–0.046	(.149)
	6	.037	(.149)	–.371	(.160)**	0.020	(.101)
Constant		.291	(.119)**	.069	(.128)	–0.043	(.081)
S.E. of Estimate		.002		.002		.001	
SSR		.000		.000		.000	
DW		1.802		1.998		1.983	
P values for Granger (non-) causality tests with LA-VAR(8)							
Log of Caseloads		.000		.001		.550	
Unemployment Rate		.002		.000		.004	
Log of Elderly Ratio		.066		.236		.000	

Notes: "Elderly ratio" refers to the ratio of those aged 65 years and above to the total population. Asterisks ***, **, and * indicate statistical significance at the .01, .05, and .10 levels, respectively. The Granger non-causality is examined using the lag-augmented VAR.

Table 5. Forecasting Performance: 2010 M03 to 2011 M02

	Fixed			Expanding			Rolling		
	MAE	RMSE	ME	MAE	RMSE	ME	MAE	RMSE	ME
Markov forecasting	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)
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)
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)
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)
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)
CF1 (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)
CF2 (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 6. Diebold-Mariano Tests (*P* values)

Null \ Alternative	Markov Forecasting	ES	ARIMA	LSTAR 1	LSTAR 2	VAR(6)	CF1	CF2
Markov Forecasting	—	0.988	0.017	0.006	0.409	0.005	0.004	0.005
ES	0.012	—	0.012	0.009	0.018	0.008	0.008	0.008
ARIMA	0.983	0.988	—	0.002	0.997	0.001	0.000	0.001
LSTAR 1	0.994	0.991	0.998	—	n.a.	0.486	0.964	0.549
LSTAR 2	0.591	0.982	0.003	n.a.	—	0.002	0.001	0.002
VAR(6)	0.995	0.992	0.999	0.514	0.998	—	0.975	0.588
CF1 (simple average)	0.996	0.992	1.000	0.036	0.999	0.025	—	0.017
CF2 (simple average, excluding ES)	0.995	0.992	0.999	0.451	0.999	0.412	0.983	—

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

Table 7. Regression Weights

	Coeff. "weight"	S.E.	*P* values
Constant	–16,395	32,274	0.630
ES	–0.722	0.336	0.075
ARIMA	0.991	0.234	0.005
MSTAR1	–0.719	0.808	0.408
MSTAR2	–0.234	0.592	0.706
VAR	1.696	0.442	0.009
R^2	0.9999		
$\bar{R}^2$	0.9998		
N	12		

Figure 1. Public Assistance Caseloads in the 2000s

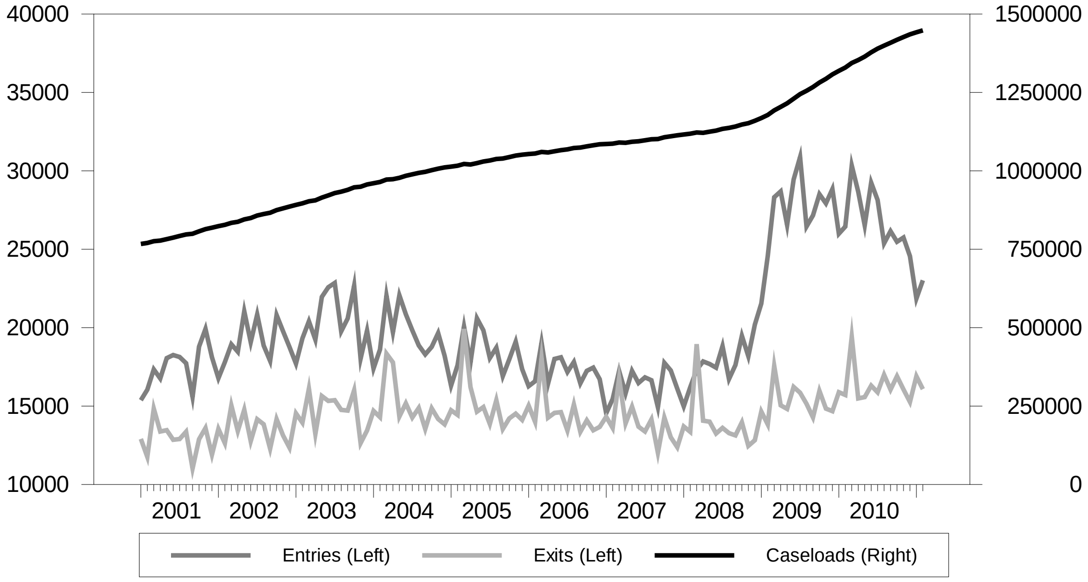


Figure 2. Percent Monthly Changes in Public Assistance Caseloads in the 2000s

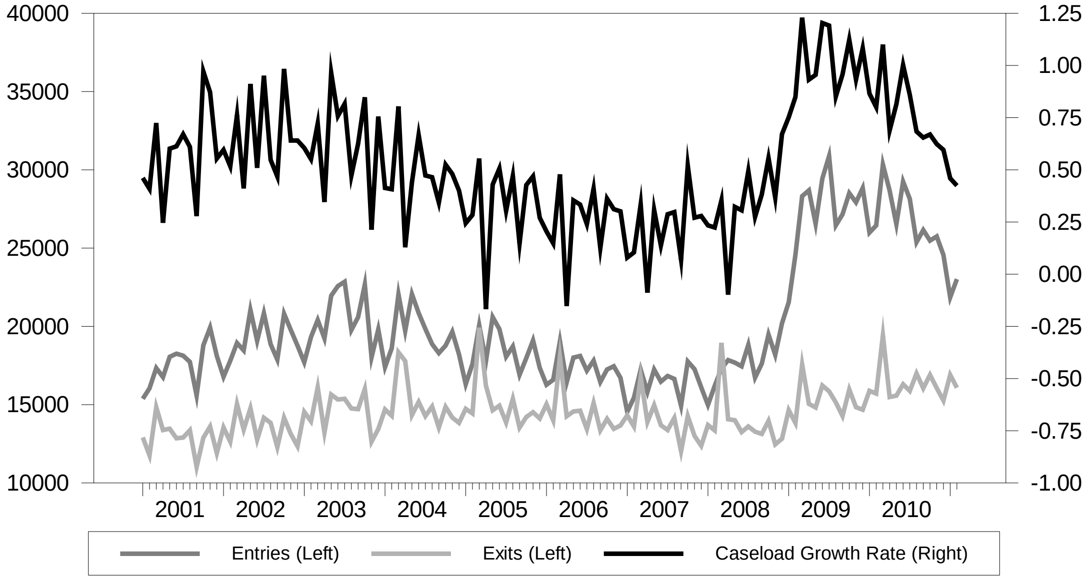


Figure 3. ISS, Entries, and Exit Rates

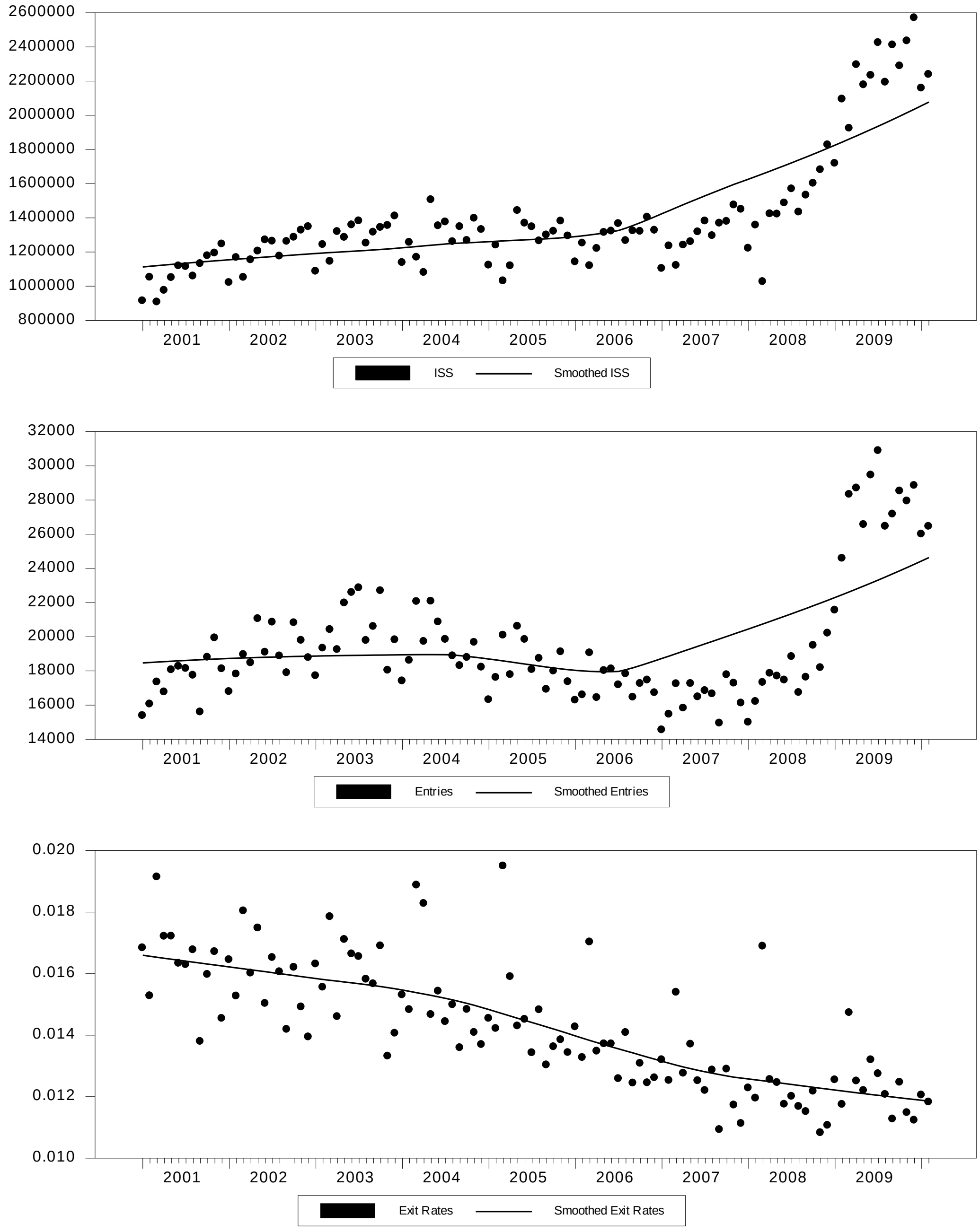


Figure 4. Markov Forecasting: Out-of-sample Fitting

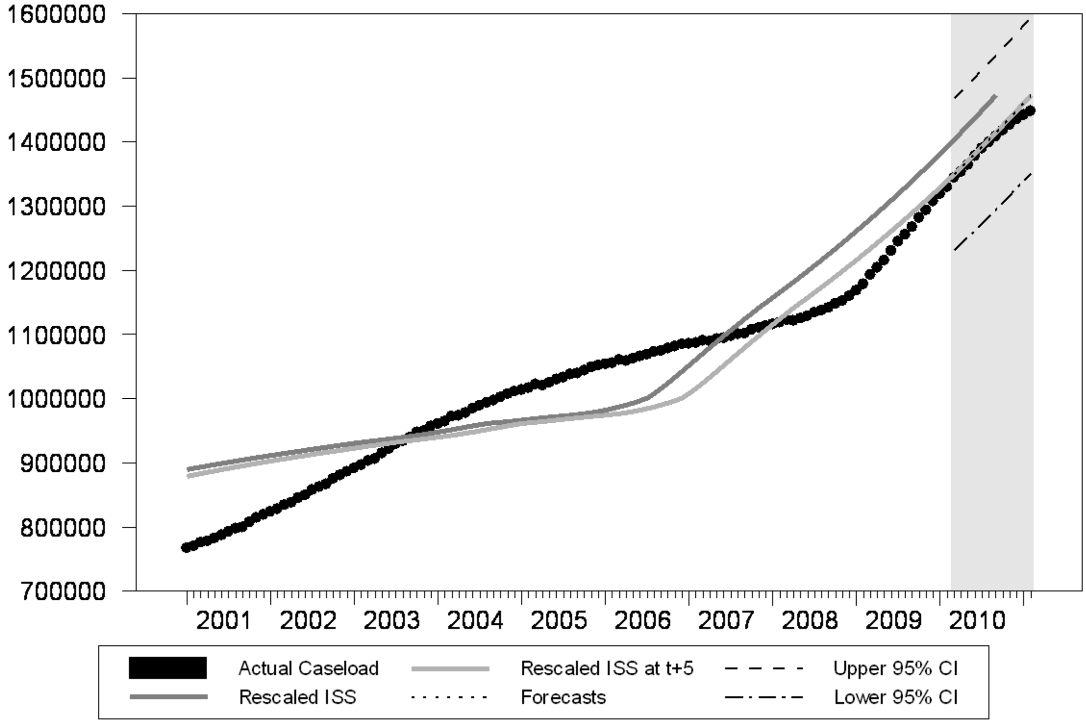


Figure 5. Markov Forecasting: Within-sample Fitting

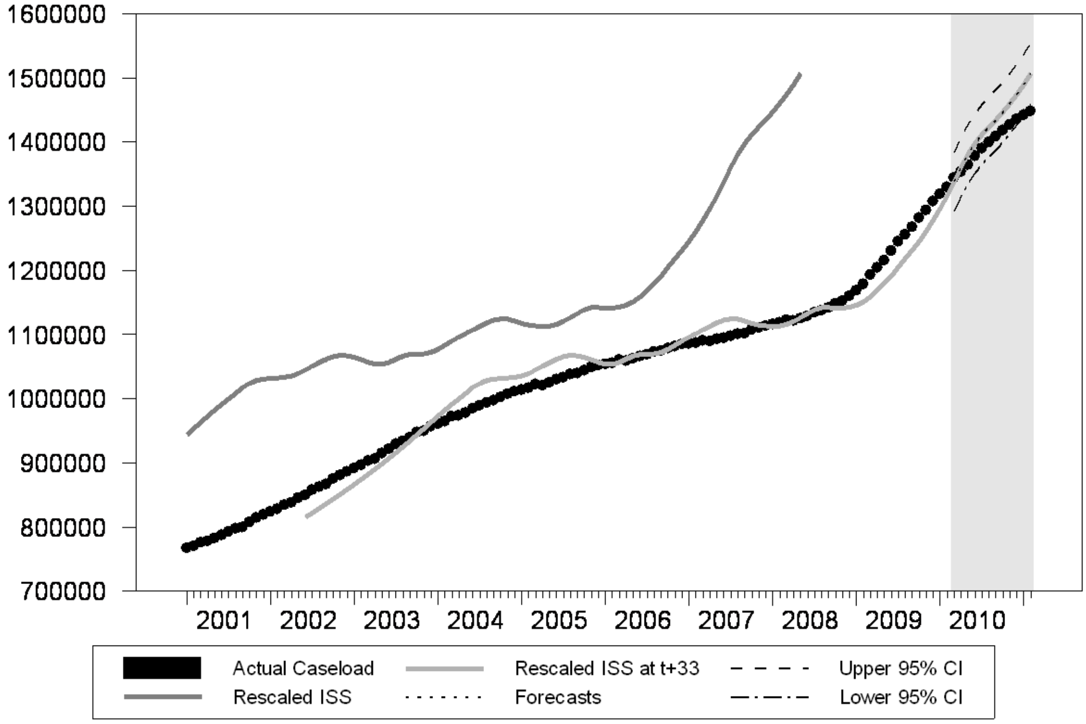


Figure 6. Correlogram for $\Delta \ln C_t - \Delta \ln C_{t-12}$

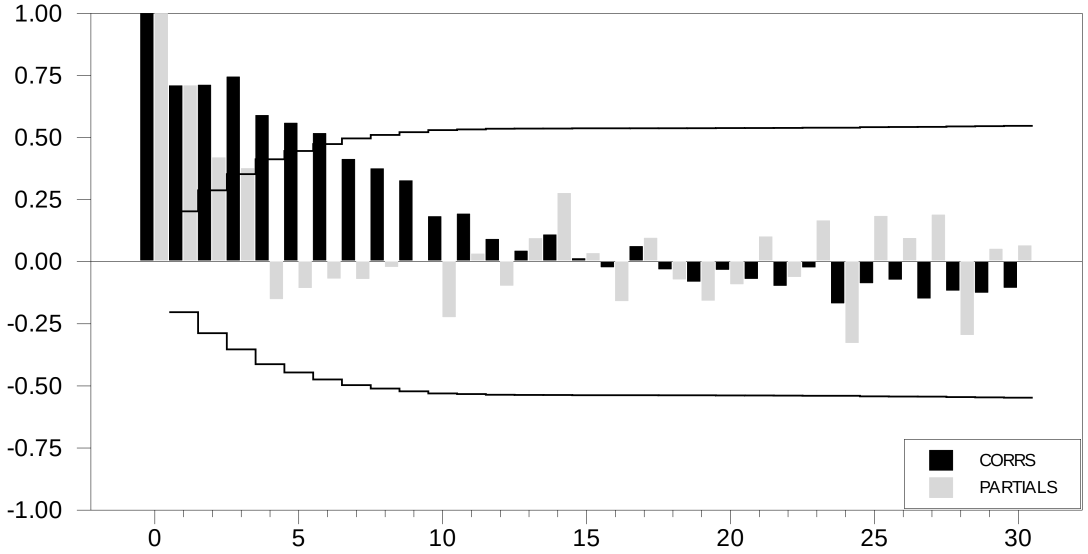


Figure 7. Correlogram for Residuals

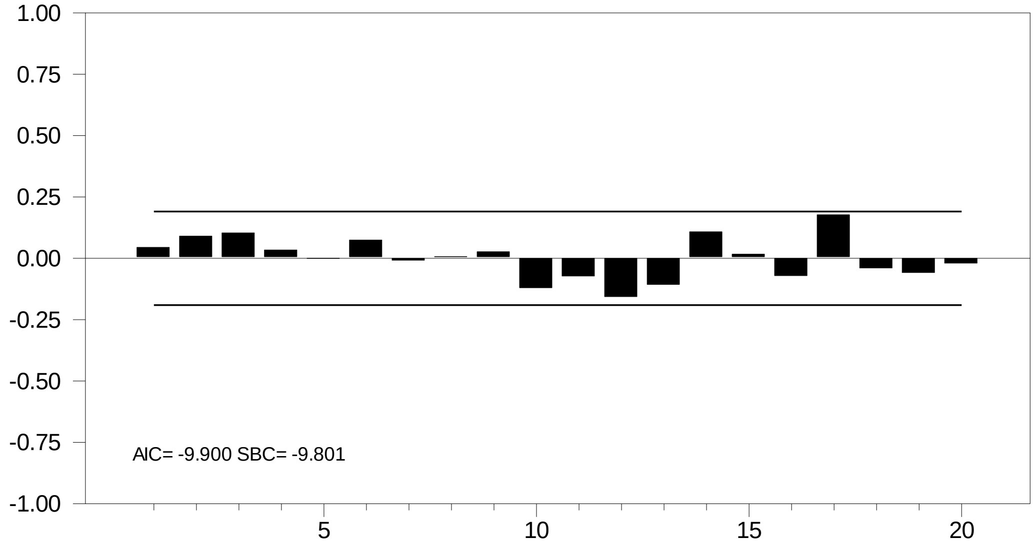


Figure 8. Pseudo Real-time Forecasting

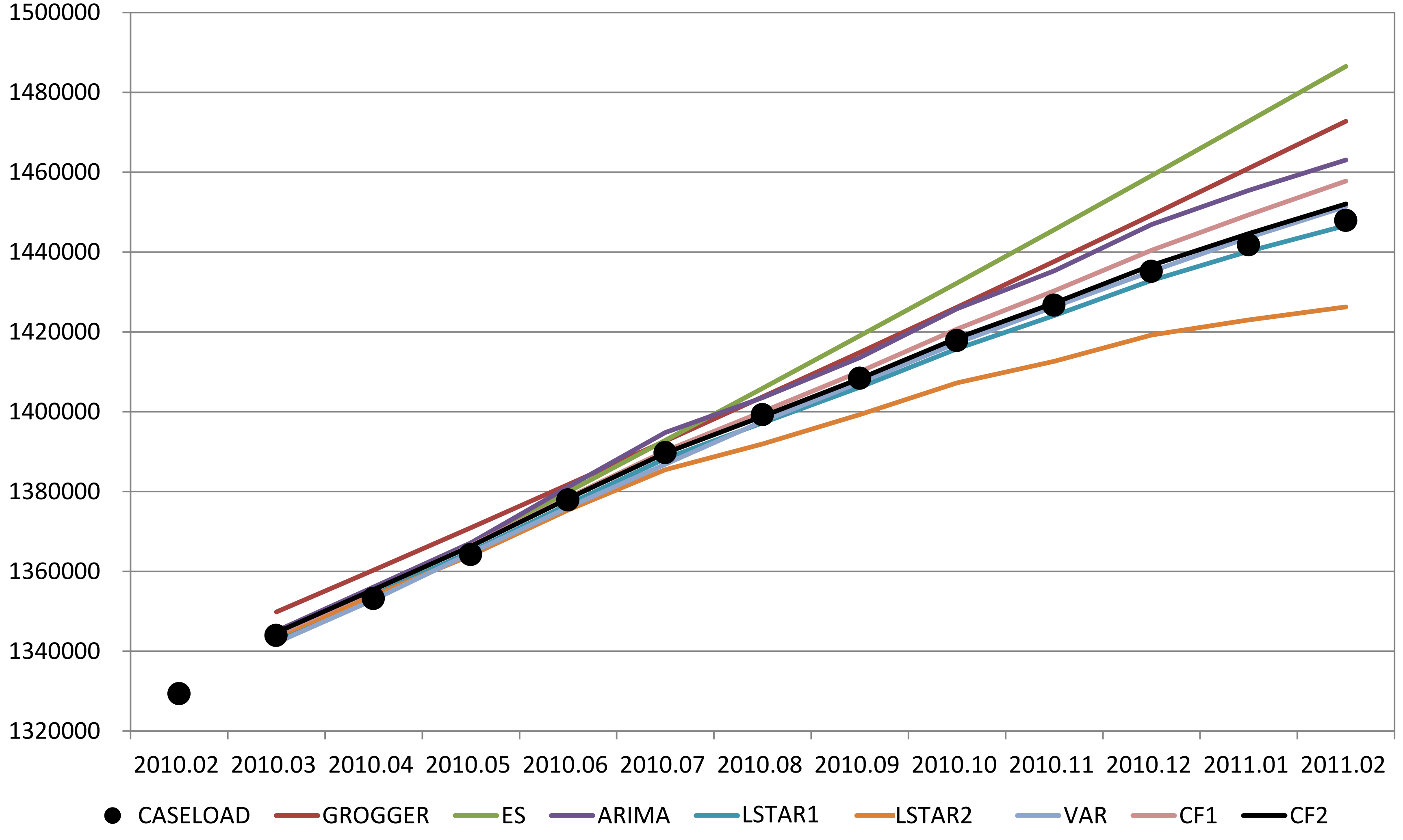
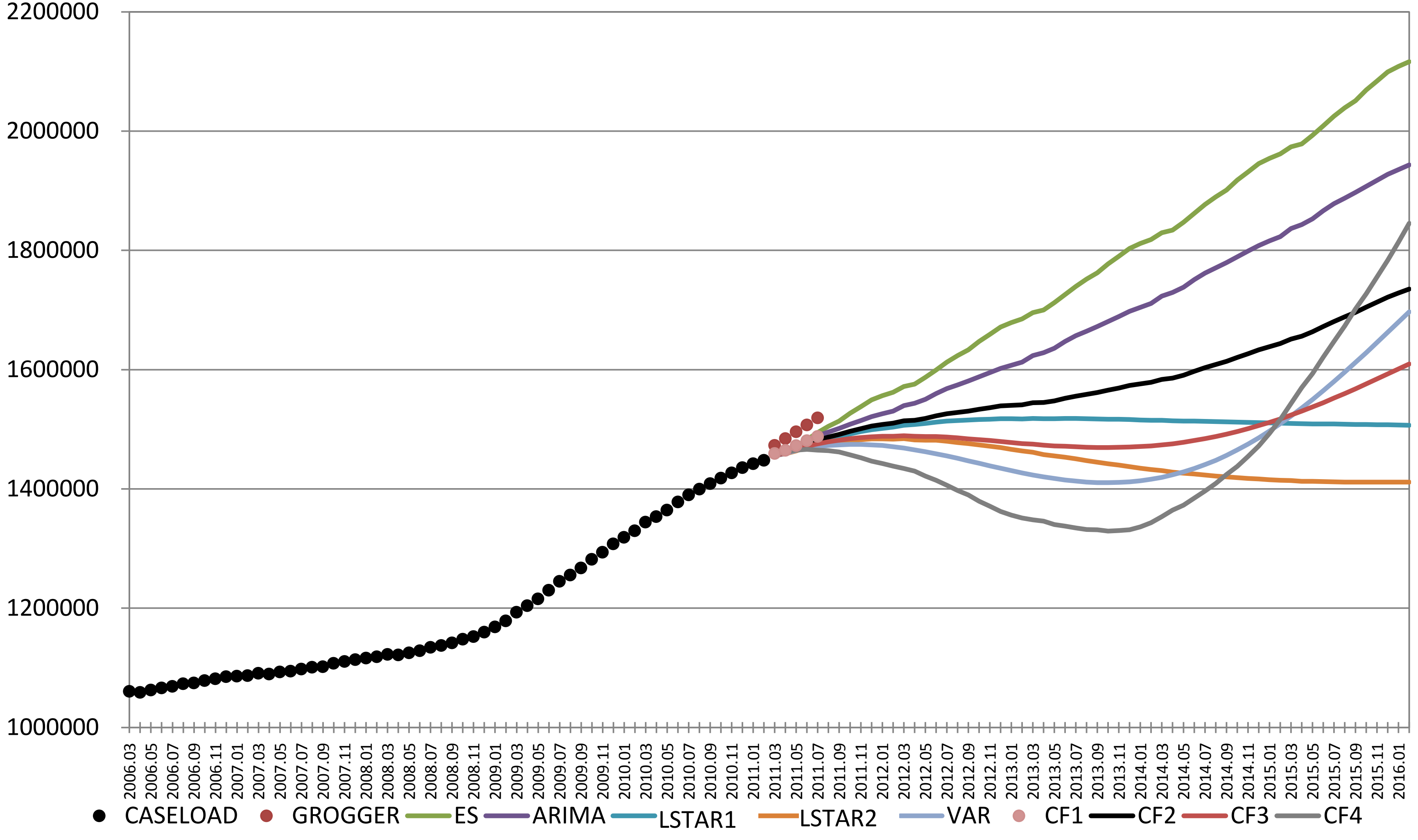


Figure 9. Real-time Forecasting



IPSS Discussion Paper Series（English）

No	Author	Title	Date
2011-E02	Wataru Kureishi and Midori Wakabayashi	Precautionary Wealth and Single Women in Japan	Mar. 2012
2011-E01	Yuka Uzuki	The Effects of Childhood Poverty on Unemployment in Early Working Life: Evidence from British Work History Data	Sep. 2011
2010-E01	Tadashi Sakai and Naomi Miyazato	Who values the family-friendly aspects of a job? Evidence from the Japanese labor market	Jul. 2011
2009-E01	Kazumasa Oguro, Junichiro Takahata and Manabu Shimasawa	Child Benefit and Fiscal Burden: OLG Model with Endogenous Fertility	Jul. 2009
2008-E02	Junya Hamaaki	The effects of the 1999 pension reform on household asset accumulation in Japan: A test of the Life-Cycle Hypothesis	Dec. 2008
2008-E01	Takanobu Kyogoku	Introduction to the theories of social market	Jul. 2008
2007-E02	Tetsuo Fukawa	Household projection 2006/07 in Japan using a micro-simulation model	Oct. 2007
2007-E01	Takanobu Kyogoku	In Search of New Socio-Economic Theory on Social Security	May 2007
2005-07	Aya Abe	Empirical Analysis of Relative Deprivation and Poverty in Japan	Mar. 2006
2005-04	Takashi Oshio and Satoshi Shimizutani	The impact of social security on income, poverty, and health of the elderly in Japan	Oct. 2005
2005-03	Seiichi Inagaki	Projections of the Japanese Socioeconomic Structure Using a Microsimulation Model (INAHSIM)	Oct. 2005