CRIME AND JUSTICE
Bulletin

NSW Bureau of Crime Statistics and Research

Contemporary Issues in Crime and Justice Number 219

This bulletin has been independently peer reviewed.

November 2018

Forecasting male and female inmate numbers: A comparison of ARIMA and ETS modelling results

George Athanasopoulos[1] and Don Weatherburn[2]

[1] Associate Professor, Econometrics & Business Statistics. Monash University, Victoria, Australia: Email: George.Athanasopoulos@monash.edu.

[2] Executive Director, NSW Bureau of Crime Statistics and Research.

Aim: *The aim of this report is to generate point and interval forecasts for the monthly average male and female prison population in NSW.*

Method: *Separate ARIMA and Exponential Smoothing (ETS) models were fitted to data on the daily average number of male and female prisoners held in NSW correctional centres between July 1997 and April 2018. Model residuals were examined to check model adequacy. Comparative model evaluation began with a minimum training window of 36 months. The training window was then expanded one observation at a time until the end of the sample. Models were re-identified and re-estimated with each step and forecasts were generated and evaluated against actual observations using absolute percentage errors (APEs) and forecast interval coverage was calculated.*

Results: *Over the short-term, both classes of models returned fairly accurate forecasts for both male and female prisoners. For the male series the accuracy of the models varied from 0.51% to 2.86% for the mean APE across the 1 to 12-steps ahead forecast horizons. The forecasts for the females were slightly less accurate, varying from 1.67% to 6.31%. Over the longer term the forecasts from the two models began to diverge.*

Conclusion: *For both the male and female series it seems that ARIMA forecasts are slightly more accurate and are probably preferable to those generated by ETS models.*

Keywords: *prison population, forecasting accuracy, ARIMA, exponential smoothing, absolute percentage error*

INTRODUCTION

Correctional administrators have no control over the number of persons remanded in custody or sentenced to prison. Accurate forecasts of inmate numbers are therefore critical for effective prison management. Prison forecasting, however, can be a hazardous enterprise. The size of the prison population at any point in time is a complex function of the number of people offending, the risk of arrest for offending, the risk of bail refusal given arrest, the risk of conviction given arrest, the risk of imprisonment given conviction, the amount of time spent in custody and the likelihood of return to custody. These factors can change very rapidly. Changes in the likelihood of bail refusal, in particular, can have a very marked effect on the size of the prison population (Halloran, Watson & Weatherburn 2017).

Different approaches to prison forecasting are generally employed, depending on whether the requirement is for a long-term forecast (e.g. more than three years) or a short-term forecast (one to two years). In the former case the usual practice is to base the forecasts on changes to age-specific rates of imprisonment and projected changes to the age structure of the population (e.g. Donnelly et al. 2015). This approach can also be informed by simulation modelling designed to project the likely impact of changes in policing, bail or sentencing policy. In the latter case the usual approach is to employ pure time series models. These models are typically univariate models. The forecasts they produce are based solely on information contained (but not necessarily apparent) in past values of the prison population (e.g. seasonal changes or changes in trends in inmate numbers). In this report with employ both ARIMA and Exponential Smoothing

Suggested citation: Athanasopoulos, G. & Weatherburn, D. (2018). *Forecasting male and female inmate numbers: A comparison of ARIMA and ETS modelling results* (Crime and Justice Bulletin No. 219). Sydney: NSW Bureau of Crime Statistics and Research.

(ETS) models to generate forecasts (for an introduction to these see Hyndman & Athanasopoulos 2018).

Univariate forecasting models are valuable where forecasts need to be done quickly or need to be automated and/or where the factors influencing a process are unknown or interact in too complex a fashion to permit construction of a reliable multivariate model. The NSW Bureau of Crime Statistics and Research (BOCSAR) currently relies on an ARIMA model developed by Donnelly and Wan (2016) to generate short-term forecasts of changes in inmate numbers in NSW 12-months ahead. At present, however, the model does not provide separate forecasts for male and female prisoners. This is unfortunate, as in Australia and most other parts of the world, the correctional systems for male and female prisoners are distinct. The present report therefore updates previous studies in presenting the results of an effort to produce reliable and separate forecasts for male and female prisoners.

TRENDS IN MALE AND FEMALE INMATE NUMBERS

Figure 1 shows the daily average number of male and female prisoners in NSW between July 1997 and April 2018. Both series exhibit increase in prison numbers, especially since 2015, as indicated by the strong positive trend in the data. There is, however, a large dip in both male and female prisoner numbers between 2010 and 2012 (marked by the red line). We have been advised that the dip is largely due to a re-organisation in the NSW Police Service in which the functions of the Highway Patrol were first subsumed within each Local Area Command before being re-established as a separate command in 2012 (Blanchette 2018). The change caused the number of persons proceeded against to court for road traffic and motor vehicle regulatory offences to fall from 55,621 in 2009 to 41,315 in 2012. Thereafter, the number proceeded against for this offence began rising again.[1]

METHOD

DATA SOURCE

Data for the study were extracted from OIMS, the offender information management system maintained by Corrections NSW. The data consists of the daily average number of male and female prisoners held in full-time custody between July 1997 and April 2018. The total includes all those held in correctional centres, transitional centres and police/court cell complexes managed by CSNSW.

Figure 1. Total monthly averages based on daily totals for male and female persons in prison for NSW

EXPLORATORY DATA ANALYSIS

Figure 2 shows seasonal sub-series plots for both male and female prisoners. Seasonal sub-series plots collect and plot together the seasons within the data, in this case the months. Hence the first line in the panels are all the Januaries plotted together, then all the Februaries, and so on. The horizontal blue line is the average of the corresponding month across the entire sample. These plots reveal a weak seasonal component in the data for both males and females with an increasing average in the first four months reaching a peak in April but not much variation after that. The seasonal component is likely due to the fact that the courts close down over the Christmas/New Year holiday period and this temporarily slows down the rate at which people remanded in custody are released.

IDENTIFYING TIMES SERIES MODELS

Two general and broadly used classes of models are considered. In particular, ARIMA models and exponential smoothing models. These provide alternative approaches to time series modelling. ARIMA models aim at describing and capturing auto-correlations within the data; while exponential smoothing models are built on the interactions of time series components, such as trend and seasonality, and generate forecasts that are weighted averages of past observations, with the weights decaying exponentially as the observations get older. We implement the modelling frameworks for both classes of models with the forecast package in R (version 8.5) using the auto.arima() and ets() functions. In what follows we present enough details for someone to follow the modelling implemented in the report. For further details and a good introduction to these modelling frameworks and the strategies used please refer to Chapters 7 and 8 of Hyndman and Athanasopoulos (2018).

SELECTING ARIMA MODELS

ARIMA models are defined by the orders of their autoregressive and moving average components as well as the order of differencing required to achieve stationarity in the data. These are represented by ARIMA (p,d,q)(P,D,Q)[m] where (p,d,q) respectively represent the orders of the autoregressive component, differencing and moving average component, (P,D,Q) represent their seasonal counterparts, and m corresponds to the number of observations per year.

The first step in ARIMA modelling is to ensure stationarity in the data by selecting the appropriate orders of differencing. As both the male and female series display a very weak seasonal component no seasonal differencing is required and therefore we set D=0. d is then selected by the KPSS test (Kwiatkowski, Phillips, Schmidt, & Shin, 1992). For both the male and female series first order differencing d=1 is adequate to ensure stationarity. The remaining seasonal dynamics are captured

Figure 2. Seasonal sub-series plots for male and female persons in prison

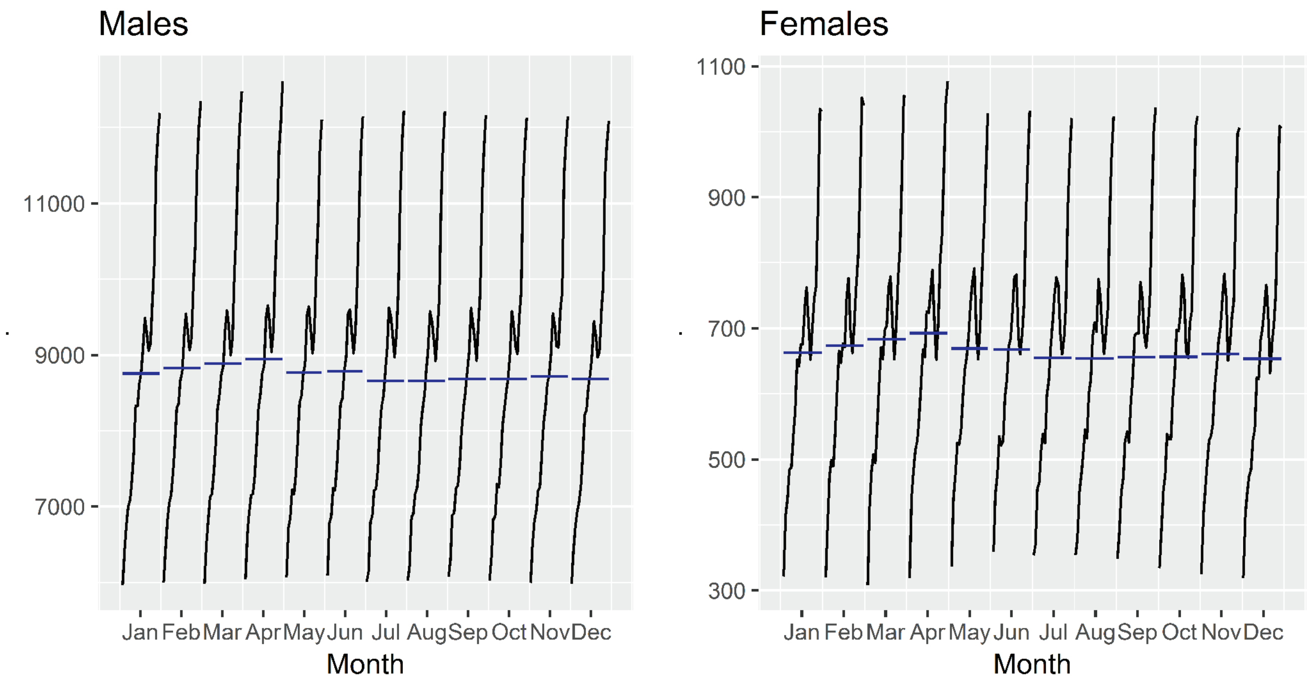
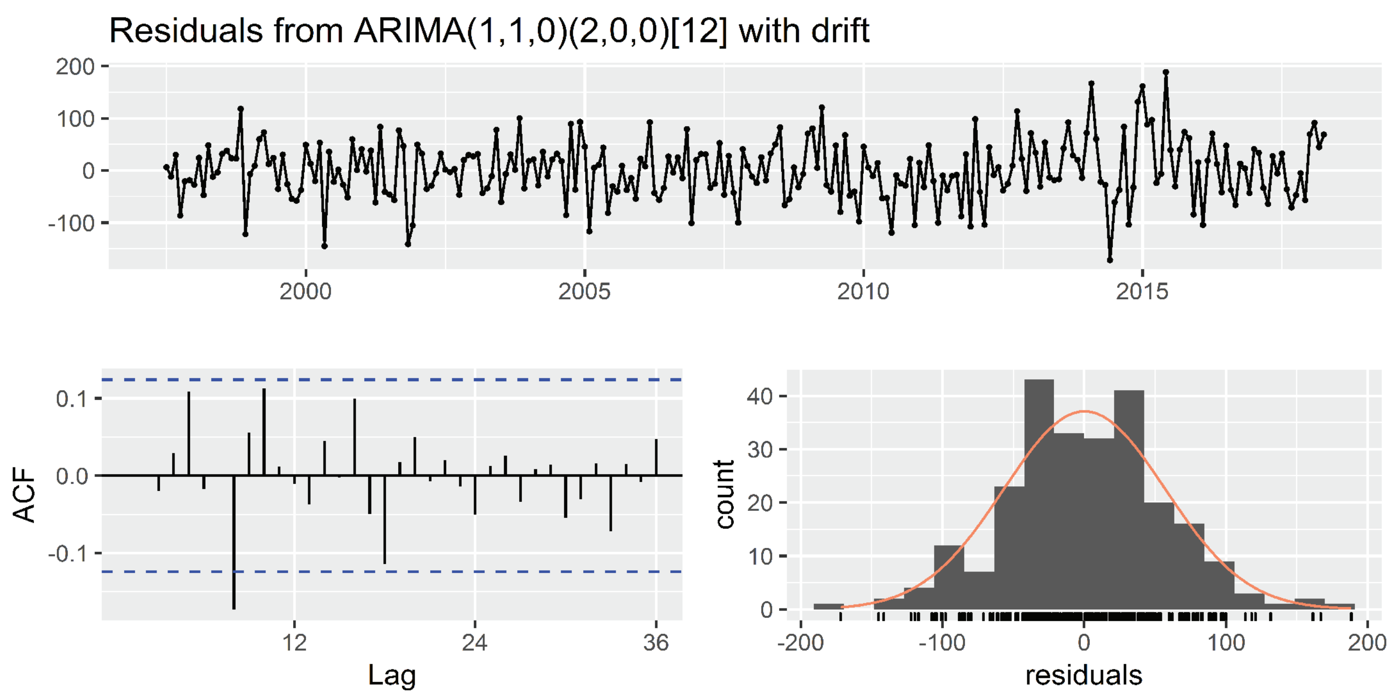


Figure 3. Residual analysis of the ARIMA model for the male series



via a seasonal AR component. Once stationarity is achieved, an exhaustive search through all possible combinations of p, q, P and Q is performed, with maximum orders of 5 and 2 for the non-seasonal and seasonal components respectively, while both including and not including a constant. The model with the minimum AICc (bias corrected Akaike Information Criterion) is selected.

The exhaustive search process selects an ARIMA (1,1,0)(2,0,0)[12] with drift. The residual analysis presented in Figure 3 shows that the residuals are well behaved without any significant dynamics left over. The p-value associated with the Ljung-Box test statistic is 0.25. Hence, the null hypothesis of no significant joint autocorrelation for the first 24 lags cannot be rejected at any reasonable level of significance. Furthermore, the residuals seem to be very close to being normal an advantage for generating prediction intervals assuming normally distributed errors. We should note that as a robustness check in what follows we do generate prediction intervals using bootstrap errors but the empirical coverage results are not sensitive to this choice and therefore we do not report those results.

The ARIMA model for the female series is an ARIMA (2,1,0)(1,0,0)[12] with drift. The residual analysis is presented in Figure 4 which shows again well behaved residuals with no left over dynamics. The Ljung-Box statistic has a p-value of 0.39 not rejecting the null hypothesis of no joint autocorrelation for the first 24 lags.

SELECTING EXPONENTIAL SMOOTHING MODELS

The general class of exponential smoothing models we consider are, innovations state space models first introduced in this form by Ord et al. (1997) and subsequently developed in a generalised class of models by various contributors (interested readers can refer to Hyndman et al., 2008 for full details). These are conveniently referred to as ETS(·,·,·) models, reflecting the three components (Error, Trend, Seasonal) that characterise a model and how these components interact. The modelling framework we implement allows for the error process to be either additive or multiplicative; for the trend component to be either, none, additive, or additive-damped; and for the seasonal component to be either, none, additive or multiplicative. Table 1 shows a two-way classification for the trend and seasonal components of exponential smoothing models. For each cell there exist two ETS models, one with additive errors (A) and one with multiplicative errors (M). For each series the model with the lowest AIC among the 18 possible models is selected.

Figure 4. Residual analysis of the ARIMA model for the female series

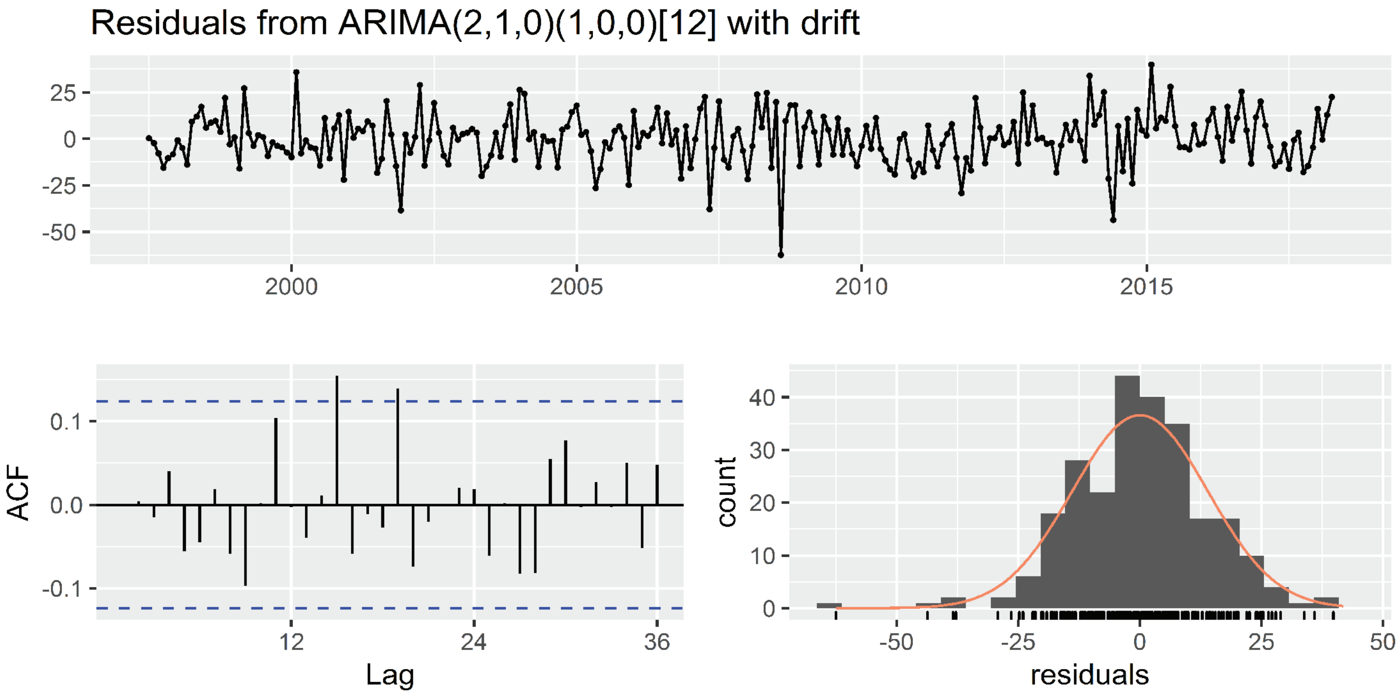


Figure 5. Residual analysis of the ETS model for male series

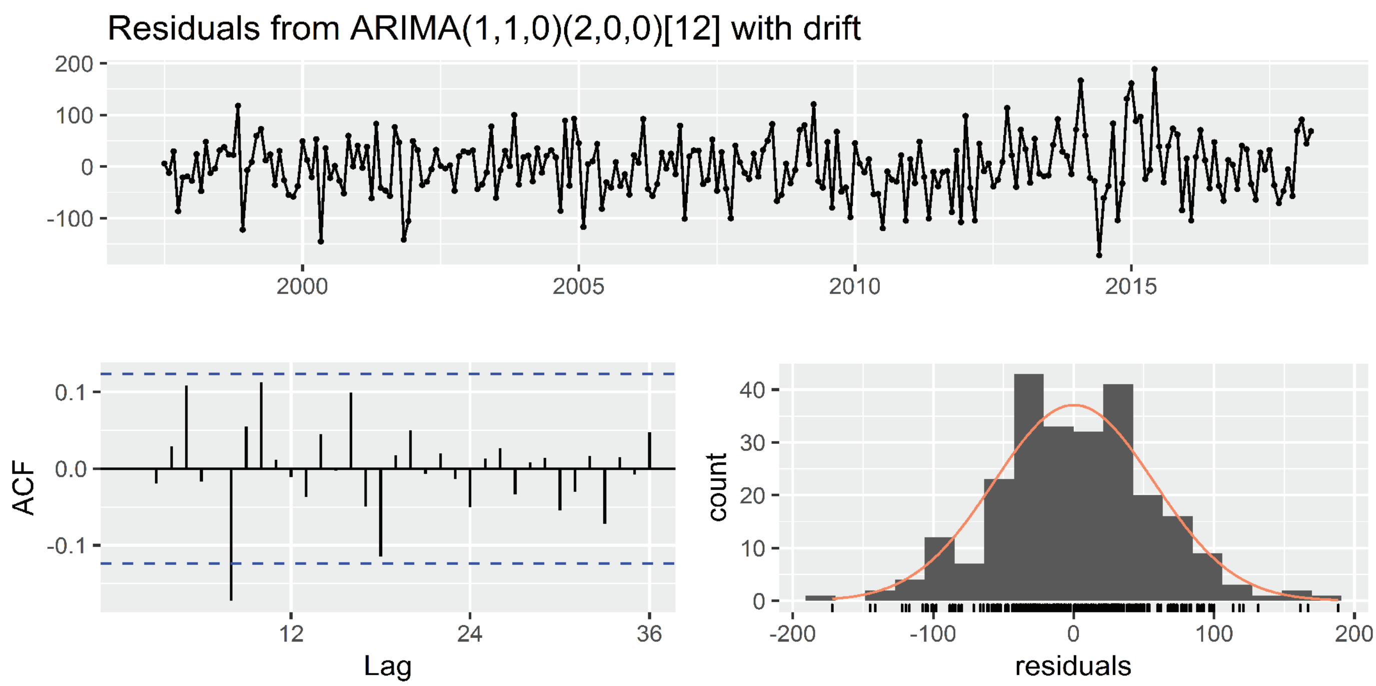


Figure 6. Residual analysis of the ETS model for the female series

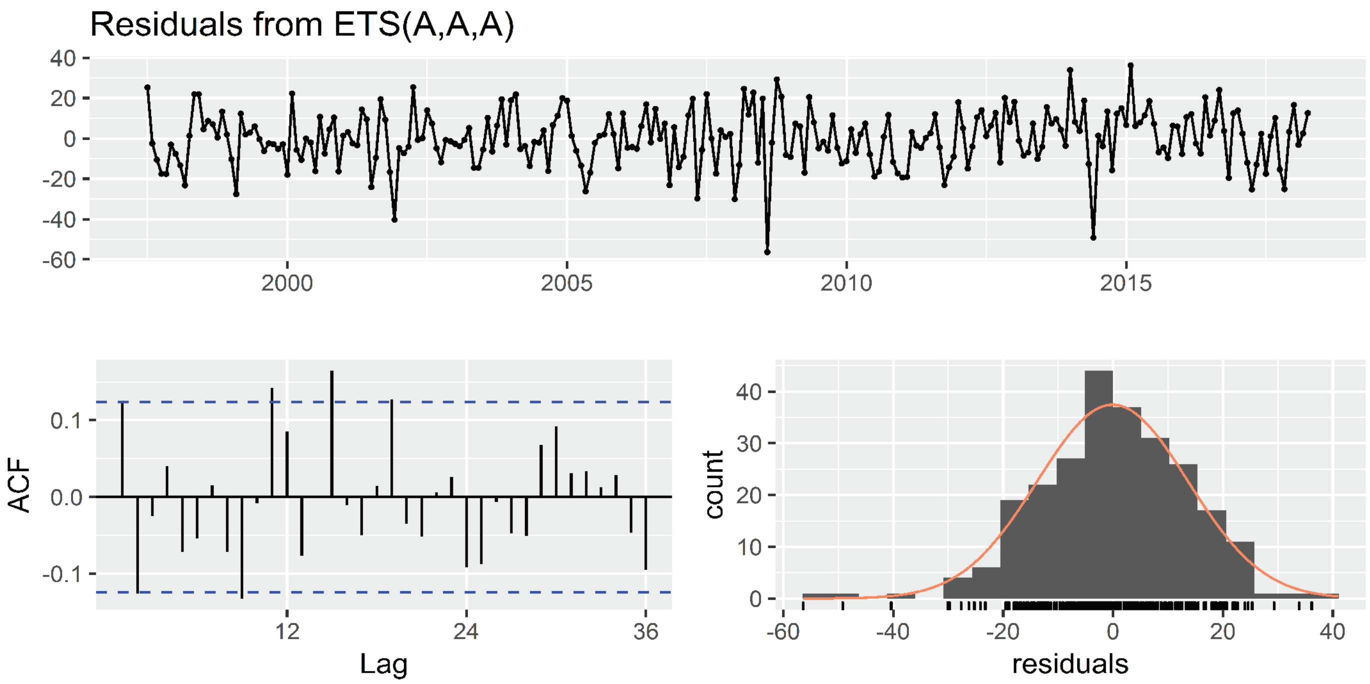


Table 1. A two-way classification for the trend and seasonal components of exponential smoothing models. For each combination of these there exist two ETS models: one with additive errors and one with multiplicative errors

		Seasonal component		
		N	A	M
Trend Component		(None)	(Additive)	(Multiplicative)
N	(None)	(N,N)	(N,A)	(N,M)
A	(Additive)	(A,N)	(A,A)	(A,M)
Ad	(Additive damped)	(Ad,N)	(Ad,A)	(Ad,M)

The model selected for the male series is an ETS(A,Ad,A), hence the model components comprise additive errors, an additive damped trend and additive seasonality. The residual analysis in Figure 5 shows that in general the residuals are well behaved. However, there are some first order dynamics left over. This may affect the prediction intervals from these models, as these may not have the correct empirical coverage. With ETS models there is not much more one can do in capturing these dynamics. At this stage we note this as a disadvantage of the ETS model for the male series and we keep this in mind for when we perform further comparisons between the two classes of models.

The model selected for the female series is an ETS (A,A,A). The residual analysis shown in Figure 6 shows that the model has captured all dynamics with well-behaved residuals resembling white noise and close to normally distributed.

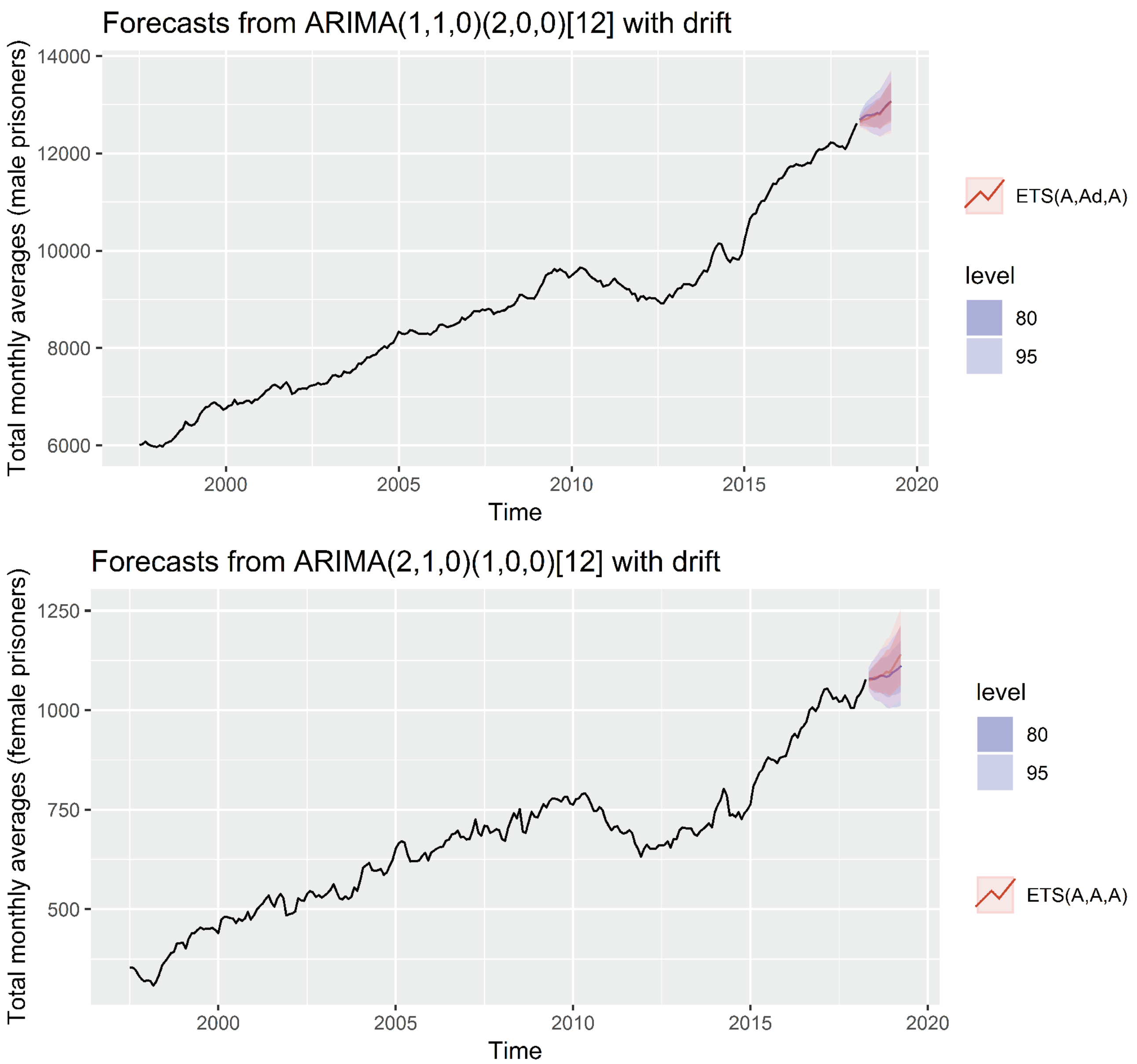
RESULTS

FORECASTING FROM THE ARIMA AND ETS MODELS

Figure 7 shows forecasts generated from the selected ARIMA and ETS models for the male and female series. The forecasts from the two sets of models for up to h = 12-steps ahead do not look very dissimilar and both look sensible with similar prediction intervals. They are both trending upwards with the exponential smoothing models also projecting a weak seasonal component.

However, there are major differences between the nature of the two sets of forecasts and projections. Although our aim is to project up to 12-steps ahead some longer-term (5-year ahead) forecasts are generated and plotted in Figure 8. These allow us to

Figure 7. Point and interval forecasts for the male and female series from the selected ARIMA and ETS models

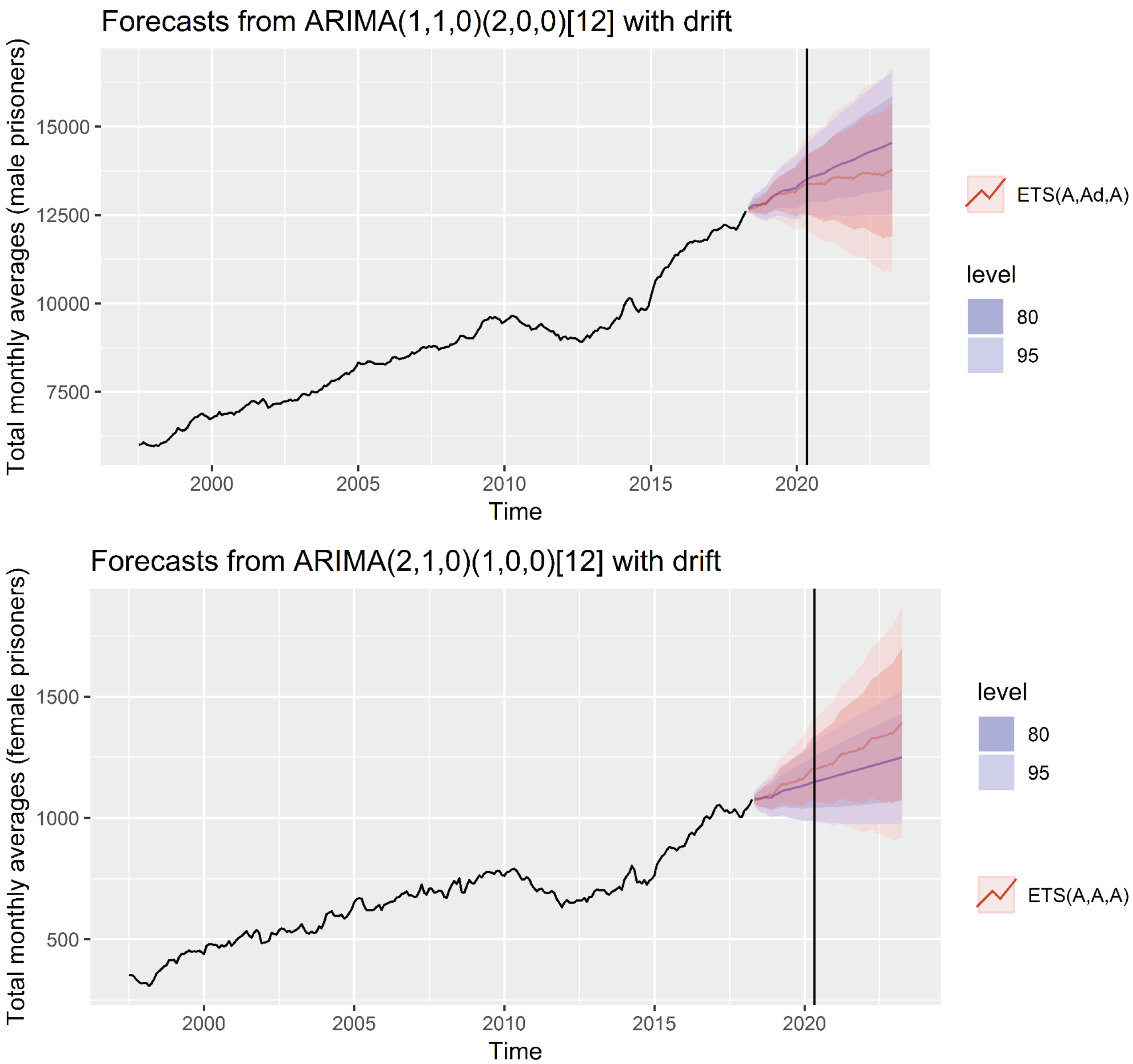


better understand the nature of the forecast trajectories and how these differ between the models. The figure demonstrates that the two sets of forecasts substantially diverge especially beyond 2-years ahead – shown by the vertical black line. In the case of the male series the damping factor in the ETS model really flattens out the trend. In contrast the ETS model for females projects higher growth than the ARIMA model. Also of noticeable difference is the magnitude of the prediction intervals the ETS ones being much wider. It should be noted again that these projections are provided for demonstrative purposes only and for understanding the difference in the model projections.

EXTENSIVE FORECAST EVALUATION OF THE ARIMA AND ETS MODELS

Table 2 provides an evaluation for the forecast accuracy of the two sets of models applied to the male and female series. The evaluation process starts using a minimum training window of 36 months. The training window is then expanded one observation at a time until the end of the sample. Models are re-identified and re-estimated with each step and forecasts are generated and evaluated against actual observations using absolute percentage errors. The evaluation is extensive and comprehensive as it involves 213 1-step ahead, 212 2-steps ahead down to 202

Figure 8. Longer-term (5-years-ahead) point and interval forecasts for the male and female series from the selected ARIMA and ETS models (This is only for illustrative purposes)



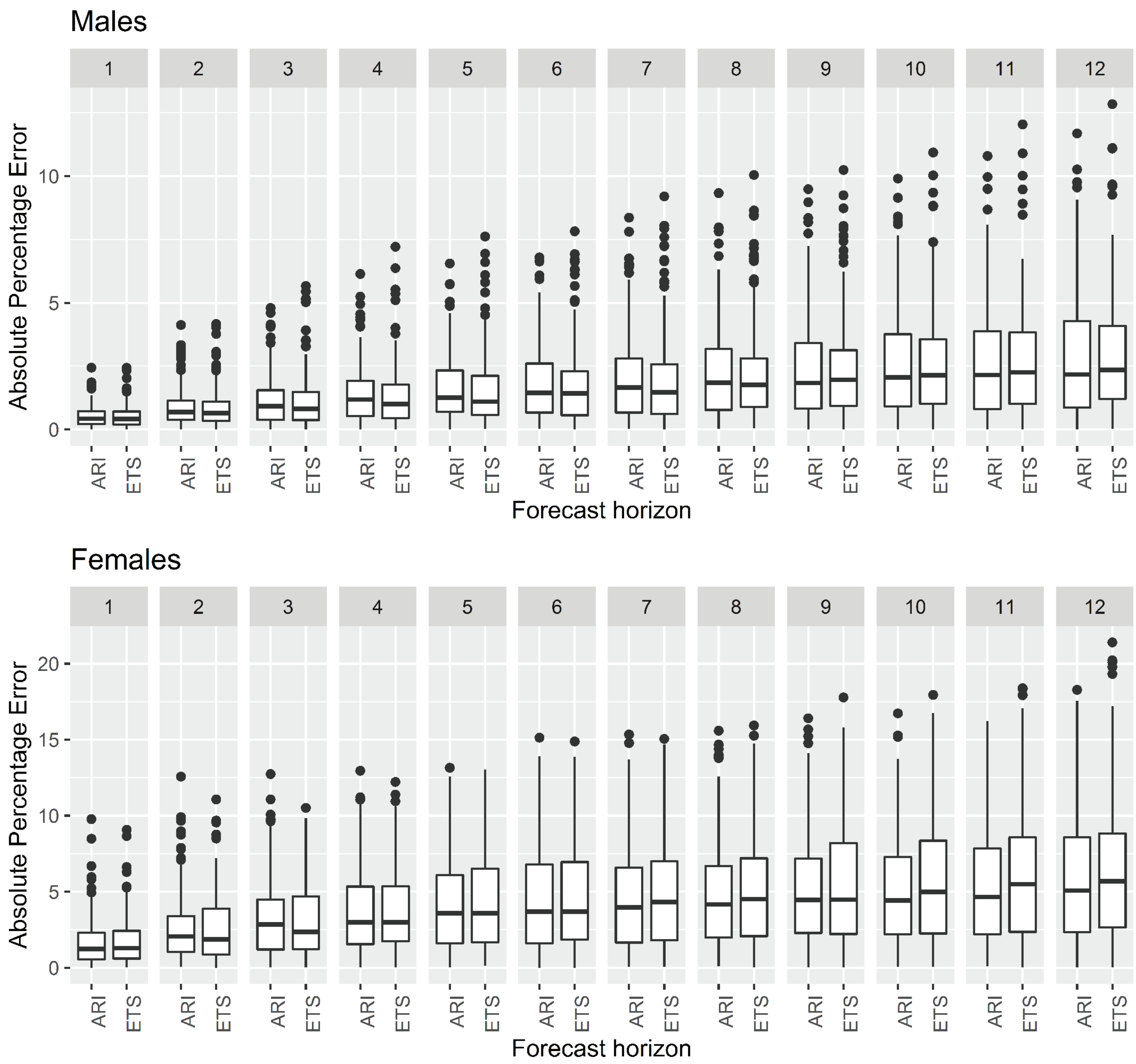
12-steps ahead forecasts. We should note that we also explored forecast evaluation by fixing the models selected for the full sample and only re-estimating the parameters at each step of the expanding window. The results were not substantially different to the ones presented below.

For the male series the accuracy of the models varies from 0.51% to 2.86% for the mean APE across the 1 to 12-steps ahead forecast horizons. These are very accurate forecasts for both ETS and ARIMA models. The forecasts for the females are slightly less accurate varying from 1.67% to 6.31% for the mean APE. Figure 9 shows box plots across each forecast horizon. For both male and female series, it seems that ARIMA forecasts are slightly more accurate and possibly the preferred model.

PREDICTION INTERVAL COVERAGE

Producing estimates of uncertainty is an important aspect of forecasting which is often ignored in practice. Table 3 shows the empirical coverage of the prediction intervals generated from the two sets of models. The coverage of both sets of models is fairly similar. The expectation is that prediction intervals with a 95%

Figure 9. Forecast evaluation for male and female series



(80%) nominal coverage rate should contain the observed values 95% (80%) of the time. It is not surprising that the empirical coverage is lower than the nominal rate. In general, forecasting methods often tend to overestimate the coverage probabilities of the forecast intervals they generate and it is not unusual, especially for pure time series models, to have relatively low empirical coverage. Hence when projecting forward, we should keep these observed coverages in mind especially for the longer forecast horizons.

EXPLORING FORECAST ACCURACY AND COVERAGE FURTHER

Figure 10 shows the point and interval forecasts generated by the ARIMA framework over the rolling test window, starting from a training window of 36 months until the end of the sample, for the male and female series, for forecast horizons 1, 4, 9 and 12. The figure clearly shows that the coverage of the ARIMA models for the female series is closer to the nominal coverage rate due the prediction intervals being wider. This is mainly driven by a larger in-sample error variance for the models estimated for the

Table 2. Mean and median absolute percentage errors (APE) for males and females using an expanding window starting from a minimum training period of 36 months

	Males				Females			
	Mean APE		Median APE		Mean APE		Median APE	
h	ARIMA	ETS	ARIMA	ETS	ARIMA	ETS	ARIMA	ETS
1	0.52	0.51	0.43	0.41	1.67	1.68	1.23	1.28
2	0.85	0.83	0.69	0.65	2.62	2.58	2.07	1.87
3	1.11	1.07	0.88	0.81	3.23	3.22	2.84	2.35
4	1.34	1.31	1.17	1.00	3.70	3.78	2.98	2.98
5	1.56	1.55	1.18	1.10	4.11	4.22	3.59	3.59
6	1.75	1.76	1.44	1.42	4.37	4.61	3.69	3.69
7	1.92	1.95	1.53	1.47	4.59	4.93	3.98	4.31
8	2.10	2.18	1.70	1.76	4.90	5.25	4.16	4.51
9	2.25	2.37	1.79	1.97	5.09	5.51	4.45	4.48
10	2.37	2.54	2.04	2.13	5.19	5.72	4.42	4.99
11	2.48	2.70	2.08	2.26	5.45	6.00	4.65	5.50
12	2.58	2.86	2.07	2.35	5.66	6.31	5.08	5.69
Av.	*1.74*	*1.80*	*1.42*	*1.44*	*4.22*	*4.48*	*3.59*	*3.77*

Table 3. Forecast interval coverage for intervals with nominal coverage 80% and 95% for males and females

	Males				Females			
	80%		95%		80%		95%	
h	Arima	ETS	Arima	ETS	Arima	ETS	Arima	ETS
1	75.59	79.34	92.02	92.96	76.53	79.81	92.02	93.9
2	74.06	77.36	90.09	88.21	75.00	75.00	89.62	89.62
3	73.93	72.51	88.63	86.26	72.99	72.51	89.10	91.47
4	72.38	71.43	87.14	85.71	74.29	75.24	89.52	90.48
5	68.42	72.25	86.60	85.65	70.34	74.16	92.34	91.87
6	72.12	71.15	84.62	86.54	71.15	75.48	94.23	90.39
7	71.50	70.05	83.58	84.06	73.91	73.91	93.72	91.79
8	68.93	66.99	84.47	84.95	76.21	75.73	93.69	92.72
9	67.81	66.34	86.34	83.90	77.07	74.15	93.66	92.68
10	66.67	65.69	86.28	81.37	75.49	75.00	94.12	92.65
11	65.03	66.50	86.70	82.76	77.83	76.36	93.60	93.60
12	67.33	68.32	83.66	82.67	78.71	75.74	91.58	95.55
Av.	*70.31*	*70.66*	*86.68*	*85.42*	*74.96*	*75.26*	*92.27*	*92.23*

female series due to the higher 'wiggliness' of the female series compared to the male series. In contrast the relatively smoother male series leads to more accurate point forecasts and prediction intervals that are too tight to achieve the nominal coverage rates.

The plots also show that as the forecast horizon increases it is challenging to forecast the turning points (i.e., to capture the cyclical property of both series). Failing to follow these challenging turns in direction, causes the coverage rates of the forecast intervals to be below their nominal values.

We should note that we also experimented with running the expanding window exercise a sub-period starting from August 2012 for males and January 2012 for females. Hence these did not include in the analysis the structural break period highlighted in Figure 1. The results for both point forecast accuracy and interval forecast coverages improved. However given that this period is considerably short for any robust forecast evaluation we chose not to report these results as they could be misleading.

Figure 10. ARIMA point forecasts (red line) and 95% interval forecasts (grey lines) for the male and female series (black lines) over the test windows for forecast horizons 1, 4, 9 and 12

SUMMARY AND DISCUSSION

The aim of this report is to generate both reliable point and interval forecasts for total male and female daily averages in prison. To this end, ARIMA and exponential smoothing models were constructed using a portion of the data in male and female inmate numbers. Forecasts for each model were then generated and compared with actual male and female inmate numbers. The results indicate that both models give fairly accurate forecasts but the selected ARIMA model slightly outperforms the ETS model, at least over the longer-term (i.e. 8-12 months ahead).

It should not be assumed from this that the recommended ARIMA model is the only tool necessary to understand likely changes in inmate numbers. It will always be challenging picking the turning points that occur in response to exogenous shocks to the criminal justice system (e.g. sudden increases in police arrest rates, sudden changes in the proportion refused bail). Users of prison forecast models should keep a close eye on factors outside the prison system that are likely to impact on inmate growth. This is particularly true of changes in arrest and bail refusal rates as changes in these particular variables can have large and rapid effects on inmate numbers.

There is clearly scope for further improvements in the forecasting process. The prison population (male or female) is made up of different subgroups, the size of which is of intrinsic interest to prison administrators. There are, for example, likely to be different resource requirements for remand and sentenced prisoners, Indigenous and non-Indigenous prisoners, young and elderly prisoners. The ideal forecast models would provide separate forecasts for each of these subpopulations as well as for the prison population as a whole. The authors are currently working on an approach to prison population forecasting that will generate reliable forecasts of subpopulations within the correctional system.

ACKNOWLEDGEMENTS

The authors would like to express their sincere thanks to Simon Corben (Manager Data & Statistics) and Rafal Jurowski (Data Analyst Custodial Operations) of the Corrections Research, Evaluation and Statistics branch of Corrections NSW for providing the data on which this analysis is based.

NOTES

1. Unpublished data from NSW Bureau of Crime Statistics and Research, October 29, 2018. Available on request from the second author.

REFERENCES

Blanchette, J. (2018). Chief Statistician. NSW Police Force. Personal Communication. June 18, 2018.

Donnelly, N., Halstead, I., Corben, S. & Weatherburn, D. (2015). *The 2015 NSW prison population forecast.* (Bureau Brief No. 105). Retrieved November 9, 2018 from NSW Bureau of Crime Statistics and Research website: https://www.bocsar.nsw.gov.au/Documents/BB/BB105.pdf

Donnelly, N. & Wan, W-Y. (2016). *Adult prison population size in New South Wales: Comparative forecasts.* (Crime and Justice Bulletin No. 199) Retrieved November 9, 2018 from NSW Bureau of Crime Statistics and Research website: https://www.bocsar.nsw.gov.au/Documents/CJB/Report-2016-Adult-prison-population-size-in-NSW-Comparative-Forecasts-cjb199.pdf

Halloran, N., Watson, E., & Weatherburn, D. (2017). *The NSW Prison Population Simulation Model: A policy analysis tool.* (Crime and Justice Bulletin No. 203). Retrieved November 9, 2018 from NSW Bureau of Crime Statistics and Research website: https://www.bocsar.nsw.gov.au/Documents/CJB/Report-2017-NSW-Prison-Population-Simulation-Model-CJB203.pdf

Hyndman, R. J., Koehler, A. B., Ord, J. K., & Snyder, R. D. (2008). *Forecasting with Exponential Smoothing: the State Space Approach.* Berlin-Heidelberg: Springer-Verlag.

Hyndman, R.J. & Athanasopoulos, G. (2018). *Forecasting: principles and practice.* (2nd Ed.). Melbourne: OTexts.

Kwiatkowski, D., Phillips, P.C.B., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics, 54,* 159-178.

Ord, J. K., Koehler, A. B., & Snyder, R. D. (1997). Estimation and Prediction for a Class of Dynamic Nonlinear Statistical Models. *Journal of the American Statistical Association, 92,* 1621-1629.

NSW Bureau of Crime Statistics and Research - Level 1, Henry Deane Building, 20 Lee Street, Sydney 2000
bcsr@justice.nsw.gov.au • www.bocsar.nsw.gov.au • Ph: (02) 8346 1100 • Fax: (02) 8346 1298
ISSN 1030-1046 (Print) ISSN 2204-5538 (Online) • ISBN 978-1-925343-68-7