# Testing a simple averaged model for local and regional population forecasts

Tom Wilson

# Introduction

Subnational population forecasting is challenging

Past forecast errors often high, especially age-specific forecasts

Less detailed, lower quality, and noisier base period data

High user expectations

Often hundreds or thousands of local areas and regions

Staffing and budget limitations

# Introduction

Variety of local and regional population forecasting methods, e.g.

- extrapolation
- various forms of cohort-component model
- dwelling-led models
- regression models
- disaggregation approaches
- land use development and dwelling allocation models
- microsimulation
- large-scale population-housing-employment-transport etc. models

Not much attention given to combining, especially averaging

# Aims

To assess performance of averaged model,
Constant Share of Population–Variable Share of Growth (CSP-VSG)

1. Does the averaged model generate short-term population forecasts for a range of subnational geographies which are of acceptable accuracy and more accurate than those from linear extrapolation?

2. For which areas were CSP-VSG averaged model forecasts successful and where were they less successful?

# Data & methods

**Averaged model**

Constant Share of Population (CSP) model
Variable Share of Growth model (VSG)
} average of models' outputs

CSP: local population is fixed proportion of an independent State forecast
VSG: local population growth is varying share independent State growth

Retrospective forecasts created for:

| Geography | 1991-2001 | 1996-2006 | 2001-2011 |
|---|---|---|---|
| SA2 local areas | | ✓ | ✓ |
| SA3 minor regions | ✓ | ✓ | ✓ |
| SA4 major regions | ✓ | ✓ | ✓ |

# Data & methods

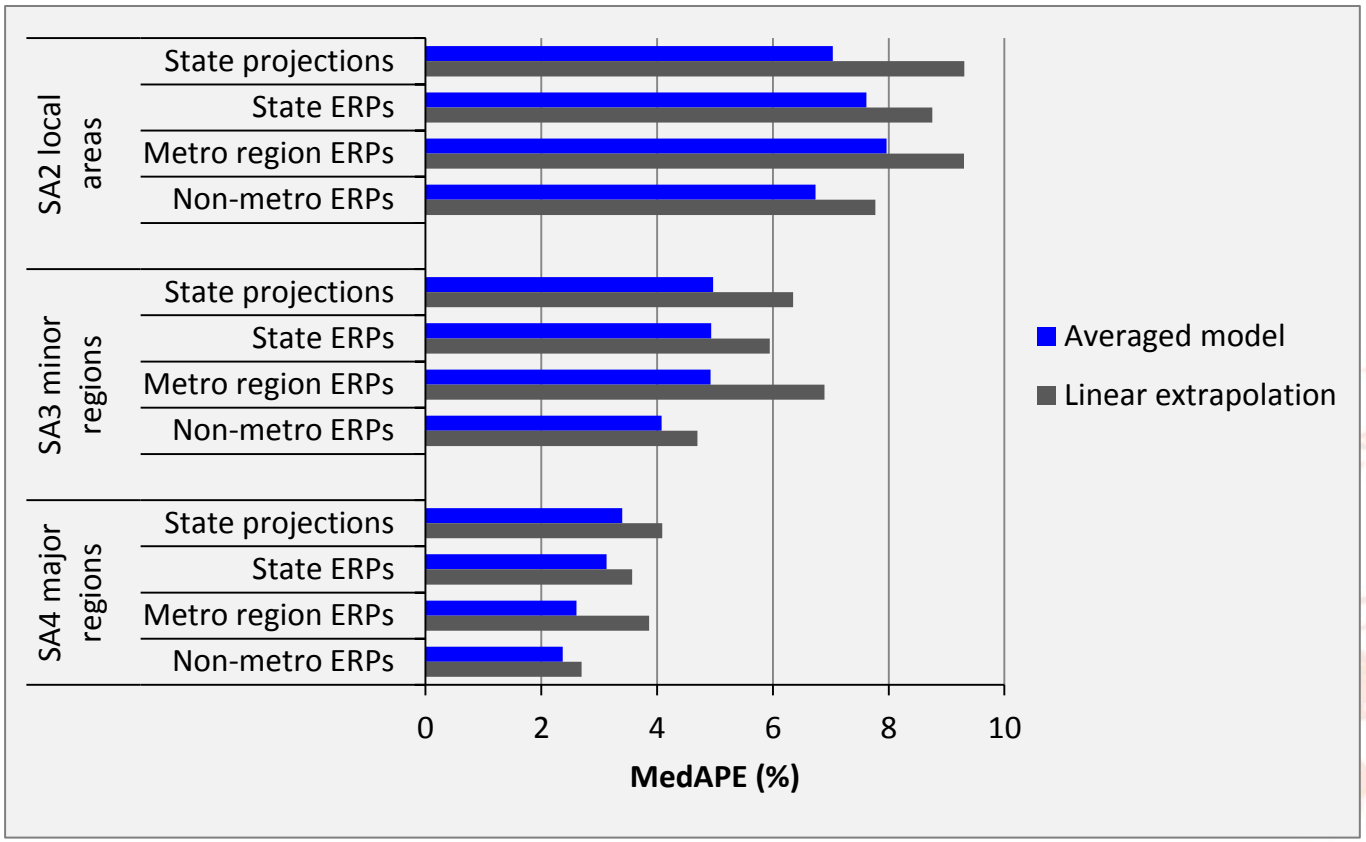
**Linear extrapolation**
Comparative naïve forecasts

**State and Territory population forecasts**
Medium series population projections produced by the Australian Bureau of Statistics used

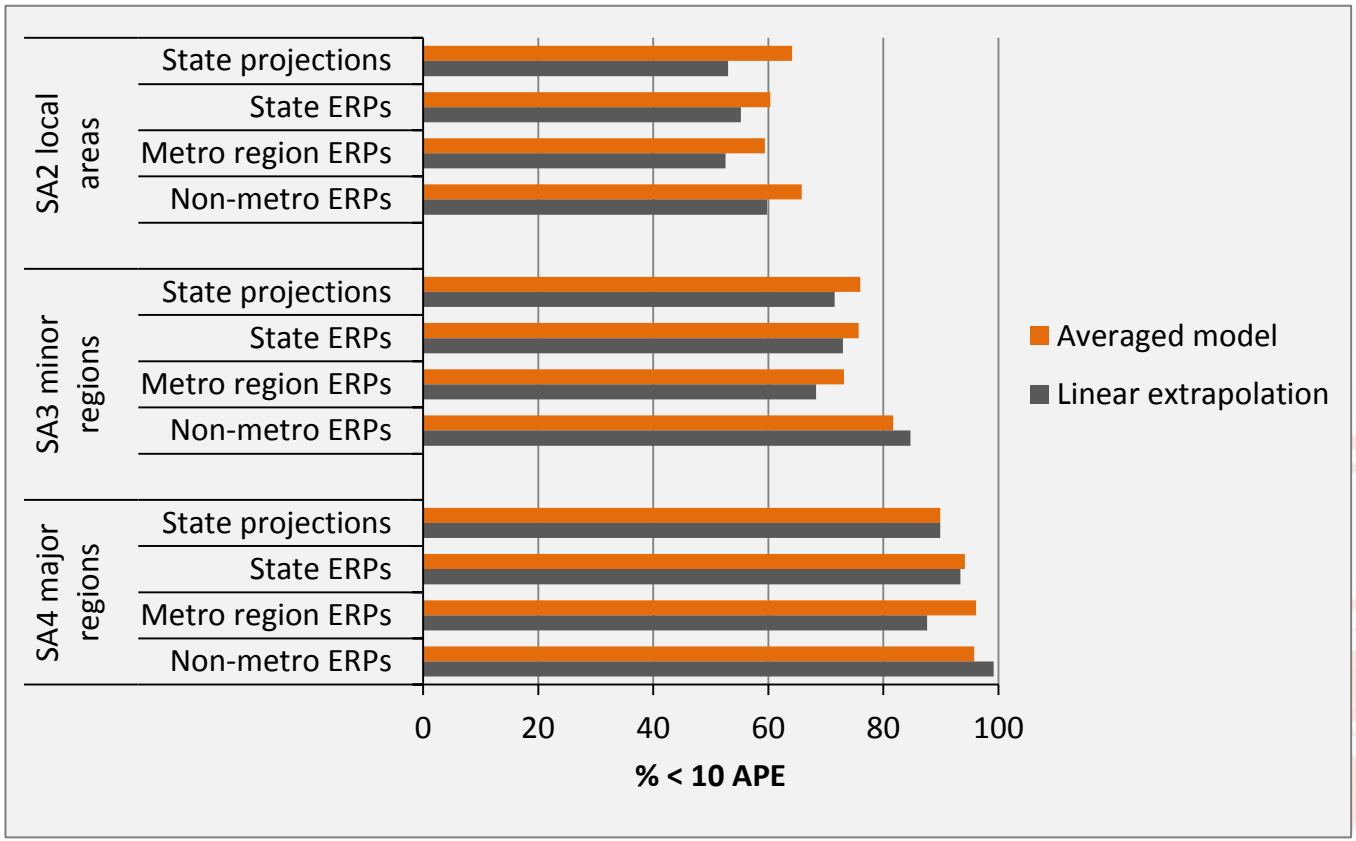
**Estimated Resident Populations (ERPs)**
Forecasts compared with ERPs (official population estimates)

**Error measures**
Absolute Percentage Error
Median Absolute Percentage Error (MedAPE)
% of areas forecast with less than 10% APE

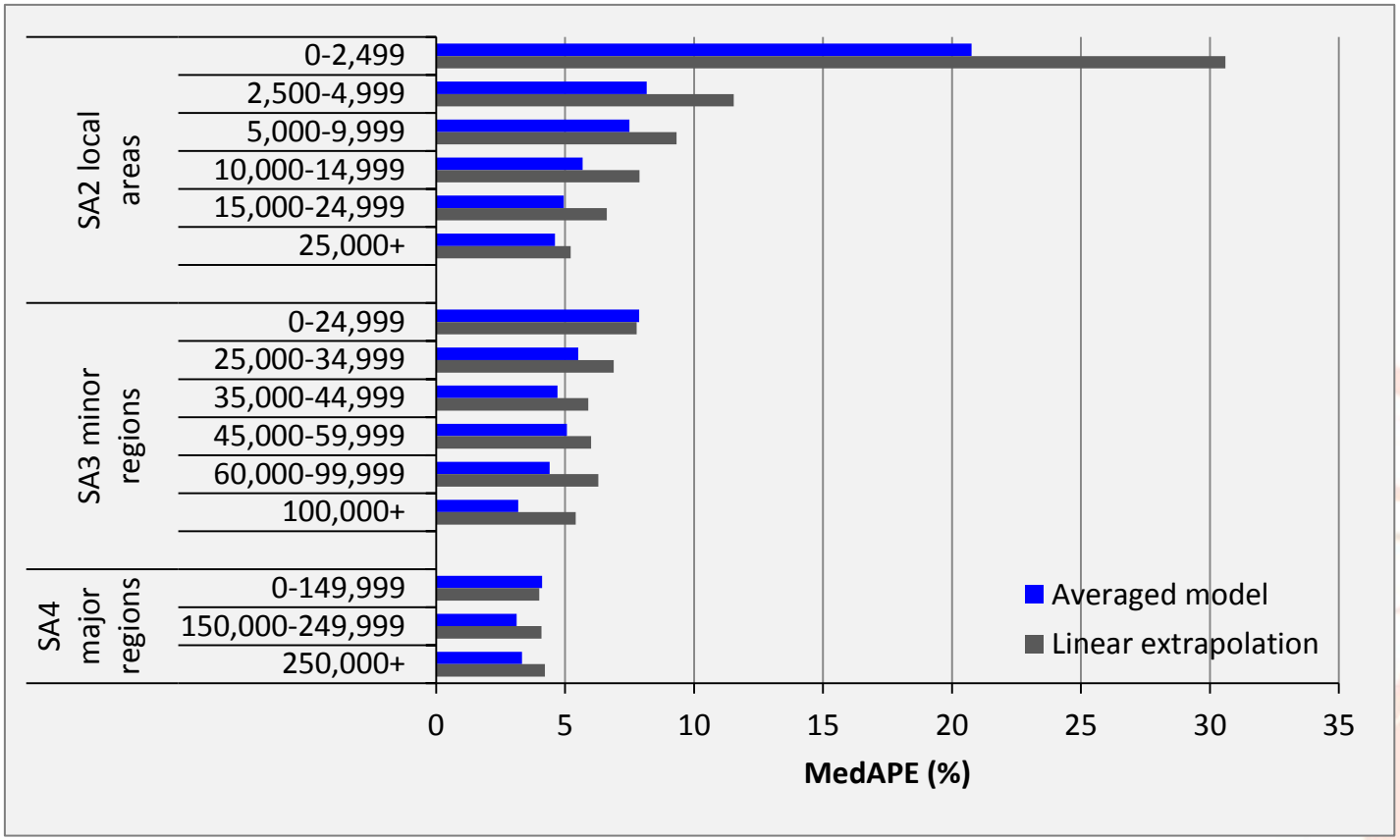
# Results: average errors



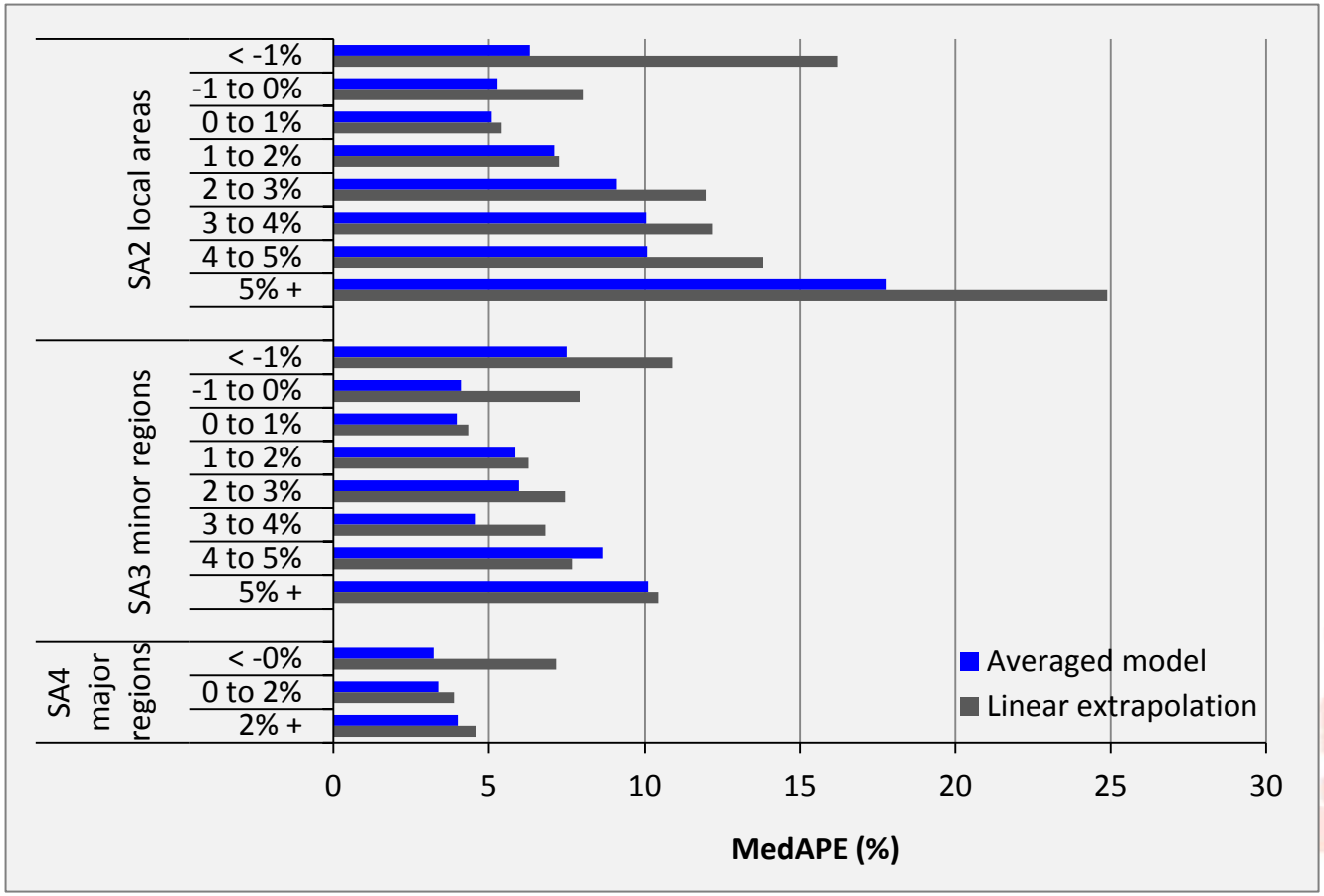
# Results: % areas with < 10% APE



# Results: average errors by population size

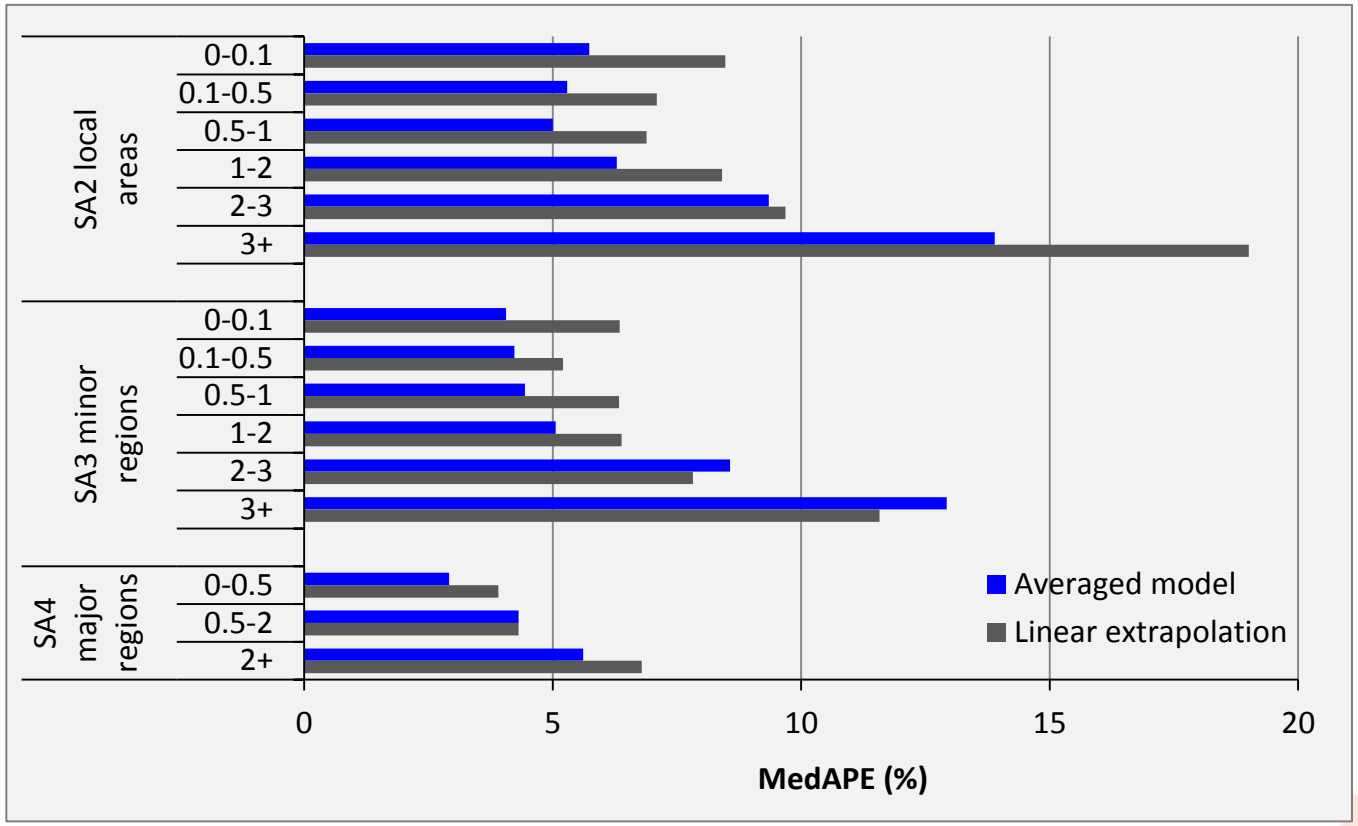


# Results: averaged errors by base period growth



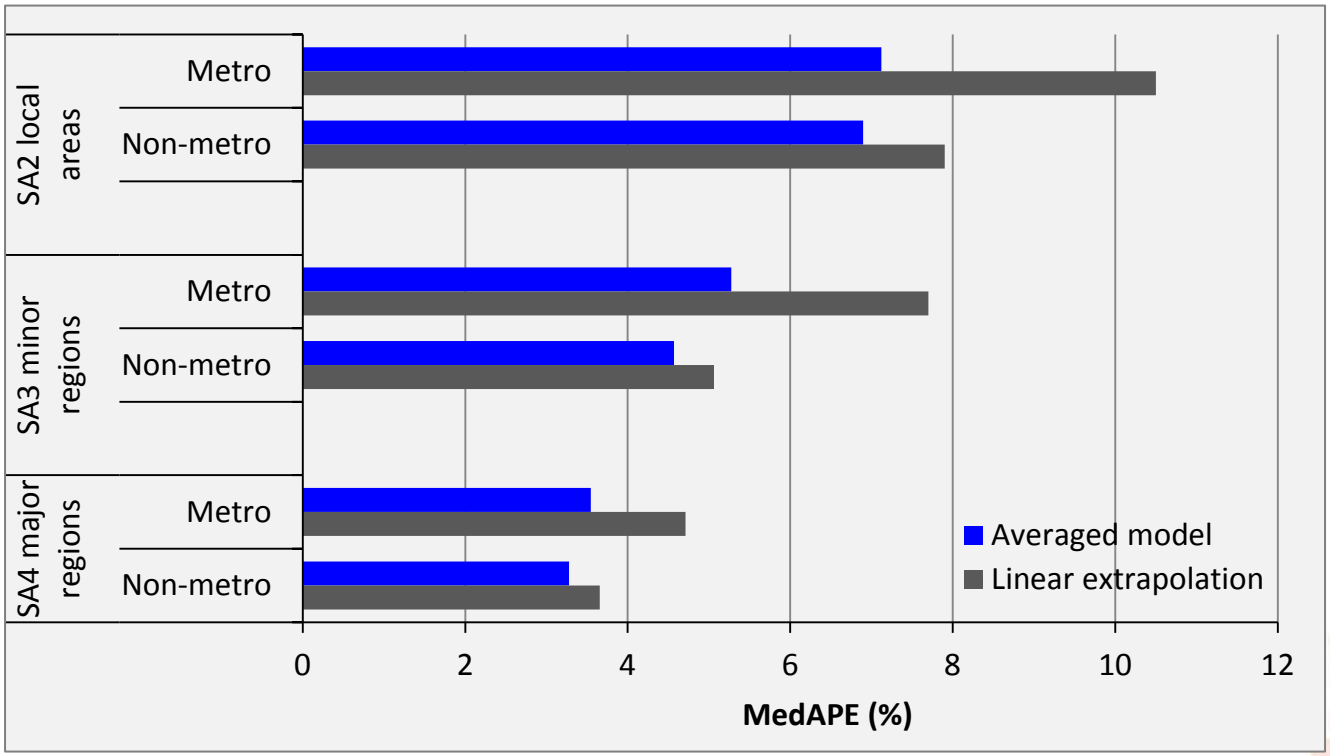

Annual average growth rate (%) of 10 year base period

# Results: average errors by base period growth volatility




Absolute difference of annual average growth rate of 1st 5 years of base period & annual average growth rate of 2nd 5 years of base period

# Results: average errors by metro/non-metro



# Results: comparison with Queensland projections

MedAPEs after 10 years by population size category

| Jump-off population | CSP-VSG averaged model | Linear extrapolation | Queensland projections |
|---|---|---|---|
| 100-2,499 | 20.1 | 30.2 | **10.7** |
| 2,500-4,999 | 8.2 | 11.5 | **7.3** |
| 5,000-9,999 | 7.5 | 9.3 | **7.4** |
| 10,000-14,999 | 5.9 | 7.9 | 6.5 |
| 15,000-24,999 | 5.1 | 6.6 | 6.4 |
| 25,000+ | 4.3 | 5.4 | 6.2 |
| All sizes | 6.3 | 8.1 | 7.6 |

# Results: modelling errors

State projection-constrained forecasts for SA2 local areas

| Effect | Forecast period | |
|---|---|---|
| | 1996-2006 | 2001-2011 |
| Intercept | 73.218*** | 61.658*** |
| ln(population) | -7.376*** | -6.299*** |
| Base period growth rate | 0.378*** | 0.312*** |
| Base period growth rate volatility | 0.269*** | 0.960*** |
| Metropolitan / non-metropolitan | 3.873*** | 3.992*** |
| Adjusted $R^2$ | 0.220 | 0.235 |

# Key points

Relative to linear extrapolation averaged model produces

- lower average forecast errors, and
- greater proportion of areas forecast within 10% error

% point reduction in error greater for SA2 local areas

Non-metropolitan-constrained forecasts more accurate

Better than Queensland projections for 10,000+ people

Poor forecasts for some areas

Forecasts simple and easy to produce, requiring minimal data inputs, staff time and organisational resources

# Strengths & weaknesses of the averaged model

| Strengths | Weaknesses |
|---|---|
| Simple model: easy to understand | Atheoretical |
| Low input data requirements | Outputs total population only |
| Largely automated: ready-made Excel spreadsheet template available | Cannot be applied to areas with zero population at the start of the base period |
| Forecasts can be produced very quickly and cheaply for hundreds or thousands of areas | Performs poorly for some areas undergoing large-scale residential (re)development |
| Relatively low average errors demonstrated for Australia over 10 year forecast horizons | Some areas will have large errors |
| Links to an independent forecast for a State/large region or other 'parent region' | Difficult to incorporate local area-specific assumptions and alternative scenarios |
| Reduces decline of declining populations and slows growth of rapidly growing populations | |

# Possible uses of the averaged model

Averaged model could form a useful part of a subnational population forecasting system.

1. Integral part of a subnational forecasting system for all, or just the non-metropolitan part, of a State.

2. Validate forecasts from another model by providing an independent set of forecasts.

3. Benchmark set of forecasts when undertaking retrospective tests of other potential forecasting models.