Modelled Territorial Authority Gross Domestic Product – MTAGDP

2024 METHODOLOGY

MINISTRY OF BUSINESS, INNOVATION & EMPLOYMENT
HĪKINA WHAKATUTUKI

Te Kāwanatanga o Aotearoa
New Zealand Government

MINISTRY OF BUSINESS,
INNOVATION & EMPLOYMENT
HĪKINA WHAKATUTUKI

Ministry of Business, Innovation and Employment (MBIE)

Hīkina Whakatutuki – Lifting to make successful.

MBIE develops and delivers policy, services, advice, and regulation to support economic growth and the prosperity and well-being of New Zealanders.

MBIE combines the former Ministries of Economic Development, Science, Innovation, and the Departments of Labour and Building and Housing.

More information

www.mbie.govt.nz

0800 20 90 20

Information, examples, and answers to your questions about the topics covered here can be found on our website www.mbie.govt.nz or by calling us free on 0800 20 90 20.

ORCID iDs

https://orcid.org/0009-0005-4321-9154
https://orcid.org/0000-0001-8920-0452
https://orcid.org/0000-0002-1243-0679

May 13, 2024

Contents

1. Summary 3
2. RAS Methodology 4
3. Hierarchical Forecasting 6
4. Creating real GDP estimates through deflating 8
5. Compound annual growth rates and per capita measures 9
6. International approaches to production of sub-national GDP 10
7. Conclusions and Future directions 11
8. References 12

1. Summary

Due to suppression constraints to ensure privacy, statistical uncertainty and the principles that apply to Tier 1 statistics, Stats NZ only disaggregates annual Gross Domestic Product (GDP) into regions by industry groups (rather than a more granular area unit). This is not unusual, with other OECD countries such as Canada and Australia Stats only publishing GDP estimates by State or Province without disaggregation into smaller geographies.

To better understand local economies and industries, MBIE scales-down the totals into Territorial Authorities and disaggregated industry groups. Bi-proportional matrix balancing (RAS) is used to estimate the disaggregation. The results, which we call the experimental Modelled Territorial Authority Gross Domestic Product dataset (MTAGDP), is a time series stretching back to 2015.

As part of MBIE's efforts to continually improve the model estimates and quality, the MTAGDP methodology has been reviewed and revised for the 2024 release. The main improvements being further exploration into forecasting methodologies and additional clarity around the components of the proportional fitting process.

Since the MTAGDP model exclusively uses returns to labour (wages and salaries) the Gross Domestic Product measures it delivers is an income-approach based model. Other approaches, such as the expenses and production, are out of scope under the current methodology.

This document describes the new methodology and outlines changes from previous releases.

Due to the experimental nature of this product, it should be used with caution. While care and diligence have been used in developing the data for this product, MBIE cannot warranty it is error-free and will not be liable for any loss or damage suffered by the use, directly or indirectly, of it.

2. RAS Methodology

The previous MTAGDP methodology release (MBIE, 2018) describes the use of a "custom Annual LEED earnings" dataset – being aggregated wages and salaries by Territorial Authority (TA) and ANZSIC06 industry group – to estimate Gross Domestic Product for each TA. This custom data is scaled-up (or 'raked') to the GDP totals for the Regional Councils and industry divisions published by Stats NZ in the Regional GDP series using iterative proportional fitting (IPF). To achieve improved estimates, the data is again re-scaled using a custom version of the Regional GDP series which was supplied by Stats NZ and contains a finer industry group disaggregation not available in the public dataset due to suppression constraints.

For this current release, a bi-proportional balancing ("RAS") method is used, replacing the two-stage raking methodology used in previous releases, following standard matrix balancing methodology (Holt, 2017). In this method, the earnings from the custom LEED data, weighted by a land coverage index, are iteratively forced to sum to the correct margin totals by Region and Industry.

The inclusion of a land coverage index is needed as many TA span regional boundaries (i.e., the Waitomo, Rotorua, Stratford, Rangitikei, Tararua, and Waitaki Districts span two regions, while the Taupō District spans four). Land coverage is preferred over population coverage for its relative simplicity, and as the population of each territorial authority crossed with region is not published.

Both the proposed method bi-proportional balancing (RAS) and Iterative Proportional Fitting Procedure (IPFP) follow the same iterative algorithm, as highlighted by Martin Idel (Idel, 2016) in a review of generalisations of matrix scaling, with the primary difference being the wider reception that RAS has had in economics and accounting.

Imputing censored data using RAS

The RAS process starts with an $n \times m$ matrix M (where in our application m is a number of industries and n is a number of regions). The RAS algorithm seeks to find diagonal matrices R and S so that:

$$x^T = 1^T RMS$$

$$y = RMS\,1$$

Where x^T, and y are the vectors of known column and row margins, and 1^T and 1 are vectors (of size n and m, respectively) whose entries are all one.

In our application, x is the total nominal GDP by industries – the column sums; and, y is the total nominal GDP by regions – the row sums.

We represent M as the sum of two matrices: one holding the non-empty values N; and, one holding the empty values K (non-existent or zero values), i.e., where $M = N + K$

Let $\overline{K}$ be obtained from K by replacing the missing values with 1:

$$\overline{K}_{jp} = \begin{cases} 0 & if \quad K_{jp} = 0 \\ 1 & if \quad K_{jp} = NA \end{cases}$$

While RAS is well known, for completeness we present the algorithm here. Proceeding iteratively:

0) Set the initial matrix M^0.
1) If k is odd, scale the rows:

$$M^{(2k-1)} = \Delta\left(\frac{y}{M^{(2k-2)}1^T}\right) M^{(2k-2)}$$

2) If k is even, scale the columns:

$$M^{(2k)} = M^{(2k-1)}\Delta\left(\frac{x}{1^T M^{(2k-1)^T}}\right)$$

Which can be summarized as:

$$M^k = \hat{r}_k M^0 \widehat{s_k}$$

where $\hat{r}_k$ is the product of the row scaling, and $\hat{s}_k$ is the product of the column scaling.

The process is repeated until either the error ϵ^k is below a set tolerance threshold, or a set number of iterations is reached, in which case the method fails to converge, where:

$$\epsilon^k = \sum_{i=1}^{m}\sum_{j=1}^{n}(M_{jp}^k - M_{jp}^0)^2$$

In our application, we apply the RAS method to $\overline{K}$ to estimate the censored cells without altering the values of the uncensored cells. The resulting matrix of GDP estimates by region and national accounts industry (denoted G_{jp}) is obtained by adding N to the matrix obtained by balancing $\overline{K}$.

Scaling-down

We seek weights w_{jp}^i for Territorial Authority i, region j, and industry p with which to scale-down G_{jp}.

We choose the following for these weights:

$$w_{jp}^{i} = \frac{LC_{ij}E_{ijp}}{\sum_i LC_{ij}E_{ijp}} \ , \quad where \quad LC_{ij} = \frac{TA_{ij}}{R_j}$$

such that LC_{ij} is the Land Coverage Index for the Territorial Authority i in the region j; E_{ipj} is the Earnings of a specific industry p on the Territorial Authority i inside the region j; and, TA_{ij} is the area of the part of TA i in region j and R_j is the area of region j.

The estimate of the Gross Domestic Product for the industry p within the TA i and region j, $\bar{G}_{ipj}$ is constructed as:

$$\bar{G}_{ipj} = w_{jp}^{i} G_{jp}$$

while constraining:

$$\sum_i \bar{G}_{ipj} = \sum_i w_{jp}^{i} G_{jp} = G_{jp}$$

Since these weights are consistent with our constraints on the totals for industry and region, with this, we have our estimates of nominal Territorial Authority GDP for each year *t* for which Stats NZ has produced GDP estimates broken down by region and industry – which, as we note below, is one year behind the publication of regional GDP.

$$GDP_{ip}^{t} = \sum_j \bar{G}_{ipj}^{t}$$

and

$$\overline{GDP}_{i}^{t} = \sum_p GDP_{ip}^{t}$$

3. Hierarchical Forecasting

There is a one-year lag between the provision of total national GDP and GDP reported at the regional level, and a further year before decomposition by industry is available. As the MTAGDP model exclusively uses returns to labour, it is reasonable to use current earnings data to forecast one year out the industry GDP at the TA level, i.e., the year in which regional GDP is yet to be produced.

For this release, we review both the hierarchical time series terms, and the performance of different forecasting methods using the hts package (Hyndman, Athanasopoulos, & Shang, 2014).

Non-hierarchical approaches were discarded due to potential forecast reconciliation issues (Sefton & Weale, 2009) between sub-national TA totals and national totals where independent forecasts are not constrained to add up to the national forecast's total.

Forecasting was performed on GDP nominal values to respect the requirement that each hierarchy of forecast at every level must sum to the forecast of the levels above. After these forecasts were obtained, growth rates were calculated to ease interpretation and be in alignment with the GDP forecast errors comparisons published by the Reserve Bank – RBNZ (McCaw, 2002) (Ranchhod, 2002) and Bank for International Settlements – BIS (Richardson, Van Florenstein Mulder, & Vehbi, 2018).

Note that the latest year of provisional estimates assumes future values will behave similarly to past history, unexpected events (such as COVID and natural hazards) will not be reflected and it is advised to be use these estimates with caution.

Forecasting MTAGDP at the level of industry and TA using hts

The hierarchical time series was set up as a three-factor model with a single interaction term as seen in **Figure 1**. The model respects that territorial authorities are nested within Regions, and contains the non-nested Industries factor, with a single interaction term for the non-nested high level factors of Regions and Industries:

$$y_{ijp} = \mu + R_i + D_{ij} + I_p + R_i I_p + \epsilon_{ijp}$$

Where R_i is the effect of the regions, D_{ij} is the effect of the territorial authority/districts i intersecting a region j, and I_p is the effect of the industry p.

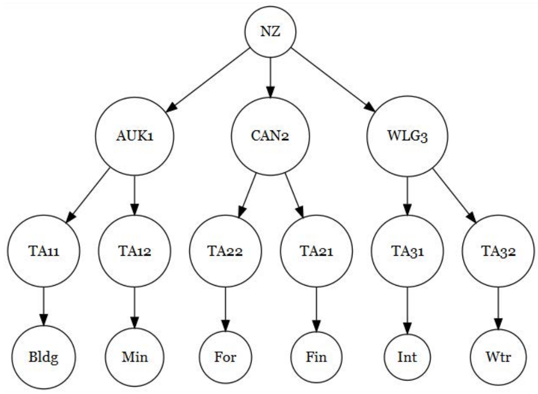

Figure 1. Hierarchy levels of the MTAGDP time series

In total, across all levels of disaggregation in the hierarchical timeseries there were 1575 individual forecast series that were the result of the combinations between the 15 regions, 75 territorial authorities and 21 industries divisions; where the 66 TA become a total of 75 as many territorial authorities extend across regions and as a consequence are represented within each region of which they are a member.

Two additional methodologies were explored: exponential smoothing, i.e, of form $y_t = \mu + y_{t-1} + \theta_1 e_{t-1}$; and, random walk, i.e., of form $y_t = \mu + y_{t-1}$.

These methods were contrasted with autoregressive models, $y_t = \mu + \varphi_1 y_{t-1} + \varphi_0$, fitted for each individual series as a benchmark.

The following tables show the performance statistics for the different forecasting methodologies, when used to forecast horizon years for which estimates were already available.

Table 1. Comparison of forecasting methodologies MAE

Forecast horizon (-h)	year	ARIMA	ETS	RW
-1	2022	0.1395330	0.1432910	0.1432910
-2	2021	0.1121459	0.1146872	0.1146872
-3	2020	0.1120884	0.1150136	0.1150136
-4	2019	0.1285088	0.1333602	0.1333602

Table 2. Comparison of forecasting methodologies RMSE

Forecast horizon (-h)	year	ARIMA	ETS	RW
-1	2022	0.00601150	0.00585983	0.00603911
-2	2021	0.00776406	0.00760255	0.00772119
-3	2020	0.01106441	0.01114870	0.01114584
-4	2019	0.01039547	0.01109380	0.01042400

All forecasts had a median variance of 0.01591, while the median difference between the provisional forecast values and the estimated values was 0.08682994, and 0.02637053 for all prior releases.

4. Creating real GDP estimates through deflating

An output of this work is to create estimates of real GDP, which are subsequently used in our estimation of compound annual growth of GDP at the TA level. To deliver this it is necessary to deflate the nominal GDP estimates.

The MTAGDP nominal estimates are adjusted by a deflator factor calculated as the ratio between the real GDP values and the nominal ones for each of the required years and across all industry groups for industry p and year t.:

$$Deflator\ Factor_{p,t} = DF_{pt} := \frac{Real\ GDP_{p,t}}{Nominal\ GDP_{p,t}}$$

As industry breakdown of real GDP by industry was not available at the time of publication, for the last forecast provisional year the deflator factors were assumed to be the same than the year in which the forecast was based.

5. Compound annual growth rates and per capita measures

Compound Annual Growth Rates (CAGR) for each territorial authority were calculated to assess the pace of economic growth. CAGR is reported as it presents the GDP in Purchasing Power Parity terms and enables both national and international comparisons while avoiding the distortions of values due to the currency changes across time.

For the MTAGDP model, a medium-term window was chosen by using a five-year window for CAGR:

$$CAGR_{it} = \left(\frac{\sum_p DF_{pt} GDP_{ip}^{t}}{\sum_p DF_{p(t-4)} GDP_{ip}^{t-4}}\right)^{\frac{1}{5}} - 1 \quad , where\ \ i\ \ is\ the\ TA$$

The MTAGDP model provides per capita experimental estimates for both nominal and real GDP values. Sub-national population estimates published by Stats New Zealand were used to calculate per capita values. CAGR per capita values were used to make comparisons between territories independent of their size:

$$PerCapitaCAGR_{it} = \frac{CAGR_{it}}{PopulationEstimate_{it}}$$

6. International approaches to production of sub-national GDP

Many jurisdictions produce sub-national estimates of GDP. In the current review, the methodology was compared with that of the United Kingdom, Australia, and Canada.

The United Kingdom's Office for National Statistics (ONS) publishes quarterly estimates of regional gross value added (GVA) for the nine English regions and Wales, which are based on Value Added Tax (VAT) returns of firm-level turnover. The ONS highlights that quarterly regional gross domestic product (GDP) is primarily based on firm-level estimates of VAT turnover. They benchmark their quarterly regional GVA estimates against the annual regional accounts by industry and region of the latest year published and state that this provides the most reliable estimate of the levels and change in volume GDP at the sub-national level.

The Australian Bureau of Statistics (ABS) in alignment with the ONS also publishes quarterly GDP estimates by State without further disaggregation into territories. In parallel with the ABS, the Australian Department of Infrastructure, Regional Development and Cities produces a Gross Regional Product (GRP) that measures the total value of goods and services produced within sub-State metropolitan areas.

In contrast, Statistics Canada (StatCan) has annual releases on provincial and territorial economic accounts of GDP across both income-based and expenditure-based measure. In addition of being useful in the formulation of policy, Statistics Canada argues for the importance of these sub-national GDP estimates to assist a private sector that needs to assess the risks and opportunities associated with doing business in particular industries, provinces, and territories.

7. Conclusions and Future directions

The MTAGDP model's subnational estimates are important to assist local authorities and the New Zealand private sector, which needs to assess the nature and health of the economy at local and industry levels.

For the current release, bi-proportional matrix balancing (RAS) was selected as the preferred methodology given the algorithm's simplicity, interpretability, and wide use in economics, while yielding similar results to the survey package's raking method used prior.

In terms of forecasting, non-linear ETS and RW methodologies outperform linear methodologies in the COVID years due to the distortion of the wage-subsidy on wages. Nonetheless, linear ARIMA methodologies result in consistently lower mean absolute errors for all forecast periods. As the difference of RMSE for the two last horizon periods of interest is insignificant between methods, the preferred method for forecasting was kept as ARIMA.

Combined forecasting methodologies were out of scope but are suggested to be explored in future iterations of the model.

International approaches for subnational GDP reporting were reviewed for UK, Canada, and Australia. Future developments could include extension of the model to quarterly releases such as Australia's, or alternatively the extension of the income approach to an expenditure-based approach and a value-added approach should suitable data source be identified to drive them.

8. References

Holt, J. (2017). *Approaches to Estimating Regional Input-Output tables.* NZ Transport Agency research report. Retrieved 02 19, 2024, from NZTA: https://www.nzta.govt.nz/resources/research/reports/619/

Hyndman, R. J., Athanasopoulos, G., & Shang, H. (2014). hts: An R Package for Forecasting Hierarchical or Grouped Time Series. *CRAN*.

Idel, M. (2016). A review of matrix scaling and Sinkhorn's normal form for matrices and positive maps. *arXiv:1609.06349v1*.

MBIE. (2018). *Regional Economic Development*. Retrieved from MTAGDP Methodology 2018: https://www.mbie.govt.nz/assets/d0f3c7808e/modelled-territorial-authority-gdp-methodology.pdf

McCaw, S. (2002). Analysis of bias and RMSE in our forecasts of key. *Reserve Bank of New Zealand*.

Ranchhod, S. (2002). GDP forecast errors. *Reserve Bank of New Zealand*.

Richardson, A., Van Florenstein Mulder, T., & Vehbi, T. (2018). Nowcasting New Zealand GDP. *Bank for International Settlements*.