**Methodological appendix**

# Probability distribution for main variables

## Population

The following methodology was proposed to incorporate the uncertainty with the information available for Chile.

For the deterministic model, population was obtained from the national official projections (Instituto Nacional de Estadistinas, 2006) and the World Population Prospects (United Nations, 2015). The key variable in this population estimate is the fertility rate, which generates different scenarios that can be seen in *Fig.*1, given its uncertainty. As observed in *Fig.*1, the different scenarios include the boundaries for different probabilities, in this case an 80% and 95% probability. These are obtained from a Bayesian hierarchical model (United Nations, 2015).

****

**Fig. 1.** **Chile Total Population: Official projections World Population Prospects. Source: (United Nations, 2016)**

As in the normal distribution, 95.44% of the values are found in the interval [*μ* - 2*σ*, *μ* + 2*σ*]. It is possible to build a 95% confidence scenario (dotted-red line) by estimating the standard deviation (*σ*) using this property (approximating 95.44% to 95%) for the period 2015 - 2030. Then using the mean value of the projected population with the standard deviation in each year, it is possible to construct the probability distribution for this variable. As an example, we present the distributions for 2020 and 2030.



**Fig. 2. Population: Normal Distribution. Source: Own elaboration**

## GDP growth rate

To incorporate uncertainty into the GDP growth rate, historical data of the observed variations of GDP growth from 1976 to 2014 is used (Table 1).

***Table 1* Historical GDP growth rate data**

|  |  |  |  |
| --- | --- | --- | --- |
| Year | GDP growth rate | Year | GDP growth rate |
| 1976 | 3.7 | 1996 | 6.8 |
| 1977 | 10.3 | 1997 | 7.1 |
| 1978 | 7.8 | 1998 | 4.1 |
| 1979 | 8.6 | 1999 | -0.5 |
| 1980 | 8.2 | 2000 | 5.1 |
| 1981 | 6.7 | 2001 | 3.3 |
| 1982 | -11.1 | 2002 | 2.7 |
| 1983 | -5.4 | 2003 | 3.8 |
| 1984 | 4.0 | 2004 | 7.0 |
| 1985 | 4.3 | 2005 | 6.2 |
| 1986 | 5.4 | 2006 | 5.7 |
| 1987 | 6.5 | 2007 | 5.2 |
| 1988 | 7.3 | 2008 | 3.3 |
| 1989 | 10.0 | 2009 | -1.0 |
| 1990 | 3.4 | 2010 | 5.8 |
| 1991 | 7.8 | 2011 | 5.8 |
| 1992 | 11.2 | 2012 | 5.5 |
| 1993 | 6.8 | 2013 | 4.2 |
| 1994 | 5.0 | 2014 | 1.9 |
| 1995 | 9.0 |

Source: Chilean Central Bank Statistics

A probability distribution is obtained to create a random GDP variable for each year, around the more probable MAPS-Chile growth rate. For this, the goodness of fit is determined from the data using the Kolmogórov-Smirnovel test with Oracle Crystal Ball software (see following figure), eliminating the data above 7% and those below -6%. In this way the distribution cannot deliver rates of GDP growth outside a range according to the assumptions of the deterministic model, which consider a stable and decreasing growth rate, similar to the trend in more developed countries like Spain and the US. The probability distribution selected is a Logistic given the goodness of fit. As observed in the following figure, the two statistics presented, P-value and D (K-S), are appropriate.



**Fig. 3. Fit to GDP historical growth data. Source: Own elaboration**

P-value: 0.059

D (K-S): 0.1471

Mean: 4.36

Scale: 1.39

From the fit, the Scale value = 1.39 is obtained from the Logistic curve, that defines the future rates over the base of the average expected by MAPS-Chile.

## Copper production

Copper production each year in the deterministic model is obtained from projections up to 2030 from the Chilean Copper Commission (COCHILCO 2015a, b). These data are based on investment projects declared by mining companies for the period. This same Commission has analyzed the uncertainty associated with these projections based on historical evidence of how projects materialize in different stages. To do this, COCHILCO initially defines a set of deterministic weights to calculate the most probable yearly future production. This is the best source of uncertainty available for the current analysis.

Fifty-one mining projects are expected to be developed in the next 12 years (COCHIlLCO, 2015a). Based on this information, three deterministic production scenarios are obtained: Maximum production scenario, Most probable production scenario, and Minimum production scenario.

The probability function for copper production used for simulation is a *Beta* (α, β), which has been suggested by COCHILCO in its study of production projections and energy consumption (COCHILCO, 2015b), and is thus aligned with the official projections of the country. This allows the distribution of production to be estimated based on the ranges of value that copper production can take based principally on the delays or advances projects exhibit over time, indicating a minimum value, most probable value, and a maximum value.





|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Distribution** | **A-D** | **P value A-D** | **K-S** | **P value K-S** | **Chi-square** | **P value Chi-square** | **Parameters**  |
| Beta | .1361 | --- | .0153 | --- | 18.0160 | 0.802 | Minimum = 5,196, Maximum = 6,175, Alpha = 67.71832, Beta = 33.41576 |
| Weibull | .2302 | 0.674 | .0183 | 0.462 | 10.6500 | 0.994 | Location = 5,679, Scale = 190, Shape = 4.28359 |
| Normal | .3496 | 0.477 | .0170 | 0.720 | 29.9640 | 0.269 | Mean = 5,852, Est Dev = 46 |
| Normal Logarithmic | .3505 | 0.384 | .0170 | 0.657 | 29.9640 | 0.226 | Location = -412.164, Mean = 5,852, Est Dev = 46 |
| Gamma | .5739 | 0.067 | .0202 | 0.334 | 24.8020 | 0.474 | Location = 4,406, Scale = 1, Shape = 999 |
| Logistics | 1.1171 | 0.000 | .0248 | 0.060 | 30.0800 | 0.264 | Mean = 5,852, Scale = 26 |
| Minimum Extreme | 8.2455 | 0.000 | .0574 | 0.000 | 89.4140 | 0.000 | Most likely = 5,874, Scale = 44 |
| Maximum Extreme | 15.1473 | 0.000 | .0775 | 0.000 | 148.5740 | 0.000 | Most likely = 5,829, Scale = 47 |
| t Student | 24.0026 | --- | .1062 | --- | 443.1560 | 0.000 | Midpoint = 5,852, Scale = 19, Degrees of freedom = 1 |
| Beta PERT | 29.8726 | --- | .0994 | --- | 196.8300 | 0.000 | Minimum = 5,680, Most likely = 5,856, Maximum = 6,007 |
| Triangular | 38.8227 | --- | .1265 | --- | 245.6660 | 0.000 | Minimum = 5,680, Most likely = 5,856, Maximum = 6,007 |
| Uniform | 129.7772 | 0.000 | .2385 | 0.000 | 952.8600 | 0.000 | Minimum = 5,687, Maximum = 6,001 |
| Pareto | 241.6760 | --- | .3771 | --- | 1,883.1220 | 0.000 | Location = 5,687, Form = 35,06049 |
| Exponential | 451.7451 | 0.000 | .6221 | 0.000 | 28,000.0000 | 0.000 | Rate = 0 |

**Fig. 4. Copper Production delivered by COCHILCO, 2015. Source: Own elaboration**

## Copper ore grade

For the copper ore grade, the model considers an average copper grade of all deposits in Chile that have an inherent uncertainty generally reflected in a lognormal variable, according to (Singer, D. 2013) among others. As a result, the average for the country copper ore grade fits a normal distribution, according to the Central Limit Theorem.

In the context of the MAPS-Chile project, scenarios are projected for copper ore grade for sulfides and oxides at the national level. It has been assumed that the most probable scenario will correspond to the average of the probability distribution for each year.

To define the Standard Deviation parameter required for the distribution, Chebyshev's theorem[[1]](#footnote-1) is used. In 95% of the cases, the standard deviation is within the maximum and minimum range defined by the scenarios (Depool Rivero and Monasterio, 2013). In this way, the following Chebyshev’s inequality for k can be obtained (*Ec. 1*).

|  |  |
| --- | --- |
|  | *Ec. 1* |

From the value of k, the standard deviation can be obtained from *Ec. 2*.

|  |  |
| --- | --- |
|   | *Ec. 2* |

Using the average value of the projected copper ore-grade (for sulfides and oxides) and the standard deviation in each year, the normal probability distribution can be obtained for this variable. In *Fig. 5* we show the probability distributions for 2015 of the sulfides, as an example.



**Fig. 5. Copper ore grade (Sulfides) year 2015: Normal Distribution. Source: Own elaboration**

1. For normal distributions, about 68% of results will fall between +1 and -1 standard deviations from the mean. About 95% will fall between +2 and -2 standard deviations. Chebyshev’s Theorem allows you to use this idea for any distribution, even if that distribution isn’t normal. The theorem states that for a population or sample, the proportion of observations is no less than (1 – (1 / k2 )), as long as the z score’s absolute value is less than or equal to k. You can only use Chebyshev’s Theorem to get results for standard deviations more than 1. It cannot be used to find results for smaller values like 0.1 or 0.9. Technically, you could use it and get some kind of a result, but those results would not be valid. [↑](#footnote-ref-1)