

Assignment 2 Bayesian Networks

Structure Learning and Causal Inference in the dataset of the World Happiness Report

Marie-Sophie Simon

December 2023

1 Introduction

In our developed world, we often measure the welfare of a country and the success of governmental policies through the gross domestic product, or GDP. If the GDP increased, we assume policies were successful and the quality of life within a country increased. Even though it is nowadays well understood that the GDP is not an ideal measurement of social welfare and has thus been criticised as such, see for example [1, 2, 3, 4], it continues to be used as a dominating indicator of welfare in the Western world.

One of the first ones to question the use of the GDP as a measurement of social welfare was Easterlin [5], leading his findings to be named the Easterlin paradox. According to the Easterlin paradox, there is indeed a strong effect of income on happiness, if we look at one point in time, both within as also between nations. Looking at this relation over a long period of time, we see that growth rates of happiness and income are in fact *not* related [5].

One of the datasets that looks to compare countries according to life satisfaction and understand what life satisfaction means, in terms of other variables is the data collected for the World Happiness Report [6]. This report provides data on different countries on various variables every year. In our first assignment [7], we created a Bayesian network to reason about causal relationships within this structure. The goal of this assignment is to analyse both our structure and a new structure that will be learned through the dataset itself. To compare the two, we will once again estimate the causal coefficients in both structures. Additionally, we will use covariate adjustment as a causal inference analysis on the focal relationship between GDP and life satisfaction. In the following, we will first describe the methods that we use, including our dataset. After that, we will review our results and finally, we will discuss them and think about consequences for future work.

2 Methodology

Before we investigate the results to our research question, we will describe the methods that we chose in more detail in the upcoming section. We will start by describing the data that we used, then move on to the structure learning algorithm used and finally explain the causal inference analysis that we decided to use.

2.1 Data

The data we will use stems from the world happiness reports from 2005-2022 and contains 156 countries [6]¹. There are 2199 entries of data points (rows) over different countries and years with 11 variables (columns). See Table 1 for a description or Box 2.1 and Box 2.2 in [6] for more details. We decided to not use the variables country and year as those should not have any causal effects on life satisfaction, or at least not causal effects that we would be interested in as it is impossible to influence them.

Variable	Description	Range
Country	The country of which the measurements are taken	165 countries
Year	The year in which the measurements were taken	2005-2022
GDP	Natural log of the GDP per capita	$0 - \infty$
Health	Healthy Life Expectancy at Birth	$0 - \infty$
Social Support	National average of the binary response to the question of social support, where 1=yes and 0=no	0-1
Freedom	National average of the binary response to the question of freedom of life choices	0-1
Generosity	Residual of regressing the national average to the donation question on log GDP per capita	0-1
Corruption	National average of the binary response to the question of perceived corruption	0-1
Positive Affect	Average of previous-day affect measures for laughter, enjoyment and interest	0-1
Negative Affect	Average of previous-day affect measures for worry, sadness and anger	0-1
Life Ladder	Life satisfaction measured by asking respondents to evaluate their current life as a whole using the image of a ladder (10=best possible life for them, 0=worst possible life for them)	0-10

Table 1: Description of the variables present in our dataset

2.2 PC - Structure Learning Algorithm

In the first assignment, we have created a bayesian network according to the conditional independencies we could find in our dataset and previous research results that indicated relations and their directions between variables [7]. In this assignment, we want to compare our hand-made model with one that was learned through the data that was provided, using the PC-algorithm.

The PC algorithm finds a bayesian structure by starting with an undirected graph that contains edges between all pairs of variables. It then continuously tests for conditional independence between pairs of variables X and Y with a set Z , starting with the smallest possible sets (which can be the

¹Data can be found at <https://worldhappiness.report/ed/2023/#appendices-and-data>

empty set). To create the v-structures, the algorithm looks at pairs of variables X, Y that are not connected but do have a common neighbour W , such that $X - W - Y$, whether $W \in Z_{XY}$. If that is not the case, a v-structure is build, i.e. $X \rightarrow W \leftarrow Y$. Once that is done, the final step is to orient the edges that can only be oriented in a certain way without causing new v-structures and thus create a completed partial DAG (CPDAG).

We chose to use the PC-stable algorithm as implemented in the `pcalg` package of R, as explained in the documentation [8] because the stable option provides an order-independent skeleton for the CPDAG.

2.3 Covariate Adjustment

As explained in our introduction, we are particularly interested in the relation between life satisfaction and the GDP, to see whether the GDP is a good measurement of welfare within a country. Hence, we make this our focal relationship for this assignment with GDP being the exposure/treatment variable and Life Ladder the outcome. As we know, the GDP can also be influenced by other factors, thus we decided to use covariate adjustment to correct for possible bias in our causal relations. We additionally did a second covariate adjustment between positive affect and life satisfaction as that one is a particularly interesting relationship as it assumes that measuring life satisfaction is not the same as measuring laughter, enjoyment and interest.

Covariate adjustment focuses always only on one causal relationship at a time, which is different from a Bayesian network structure. To do covariate adjustment, it is necessary to find the variables that one needs to adjust for during a variance analysis, which is called an adjustment set Z . An adjustment set contains no descendants of the exposure variable and d-separates all paths that go from exposure to outcome, starting with $X \leftarrow$. In a continuous dataset such as the one, we are using, the variance analysis that is most often used is a linear regression model [9]. This means, for the unadjusted estimate, we will simply look at the linear regression between the two variables of our focal relationship. For the adjusted estimate, we will look at the outcome in relation to the exposure and one of the adjustment sets that we have computed as described before.

As said in the introduction, we are particularly interested in the causal relationship of the GDP towards Life Satisfaction, to investigate whether the GDP is an appropriate measure of a nation's welfare. Therefore, we define the GDP to be our exposure variable and Life Satisfaction to be our outcome variable.

3 Results

In the following, we will look at the results that our analysis gives. Below is the final model that we created manually, for a detailed explanation of methodology and justification of results, please refer to that assignment [7].

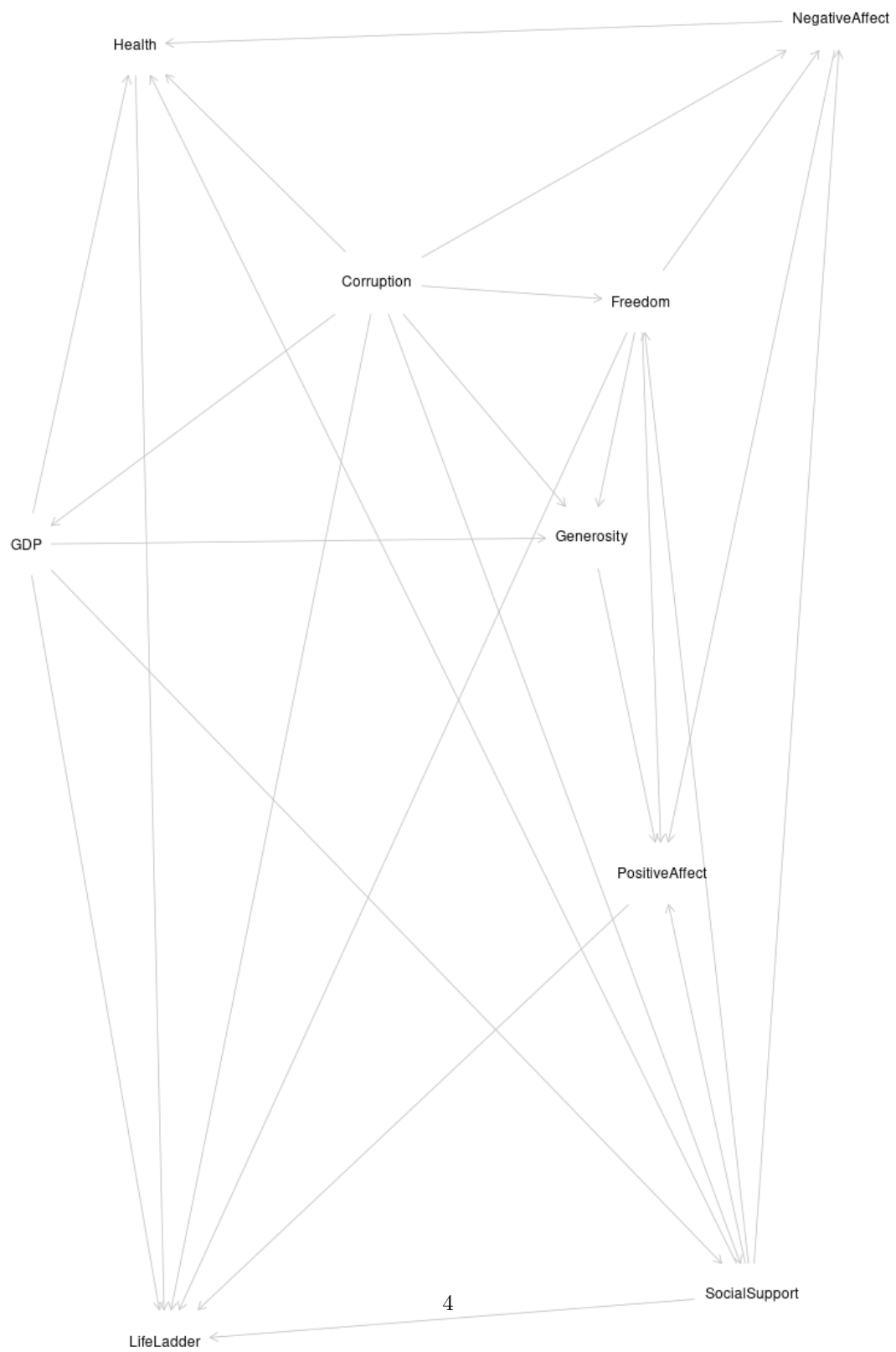


Figure 1: The final structure we created manually

3.1 The Learned Network Structure

Using `pc.stable` we learn the structure of our Bayesian network from the data we have. We started without whitelisting any connections, to see what the resulting graph would look like. As the resulting graph immediately contained all the connections that we would have possibly whitelisted (i.e. our focal relationships between GDP - Life Ladder and Positive Affect - Life Ladder), we continue with this result which can be seen below.

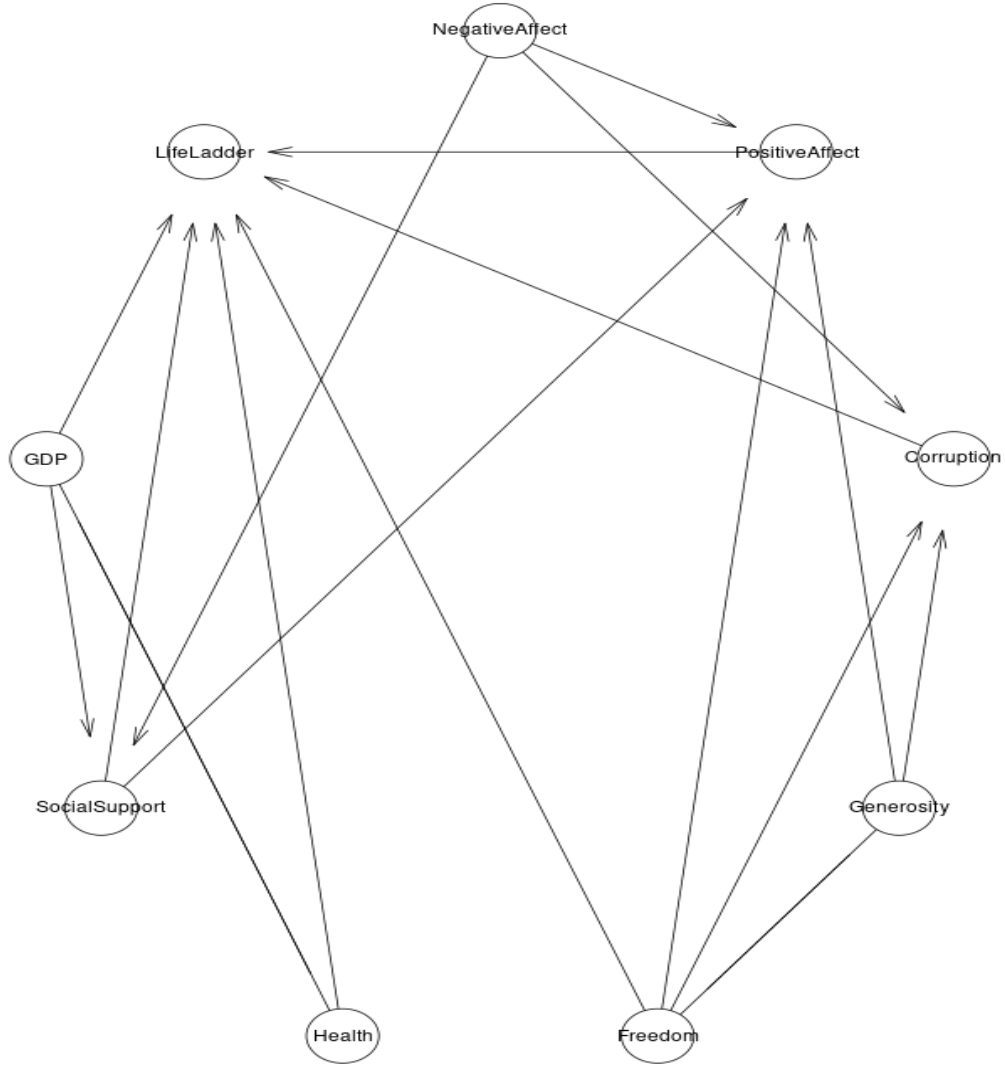


Figure 2: The initial structure of the learned bayesian network

A few simple sanity checks no immediate problems. For example, our outcome variable Life Ladder, does not have any outgoing arrows. Similarly, we can see that all the direct correlations

that contribute towards Life Satisfaction, according to the World Happiness Report [6], are present in the graph. In the following, we will compare simply the structure of the learned graph with our hand-made graph from assignment one [7]. Only after that, will we calculate the coefficients of the graphs to investigate more differences.

3.1.1 Comparison of structure

We can easily see that the learned model and our hand-made model are not Markov equivalent as our hand-made model has significantly more connections, meaning the skeleton is already not the same, which will make it impossible to be equivalent to the learned structure. Unsurprisingly, this also leads to the fact that none of the v-structures are the same, which would be another criterion for equivalence. In the following, we will look at a few of the most obvious differences between the learned and the hand-made graph.

The learned model has only 17 connections, while our hand-made model had 26 connections. A lot of the ones that are not part of the model are indeed the ones that were found to be very weak once we calculated the causality. However, this is not the case for all of them. For example, Corruption and GDP were found to have a rather strong causality in our hand-made model (0.35) and are now left out completely. Similarly, a few of the relations we had added very early in our process are now not a part of the model anymore, such as Health to Negative Affect (0.27) which we added due to a correlation found in the World Happiness Report itself [6] or Social Support to Health due to previous research [10, 11]. The fact that the learned model has significantly fewer connections implies more conditional independencies, which we have not identified in our hand-made model. As already mentioned, corruption is one of these variables. Another interesting one is health, which is thought to be conditionally independent of more variables in the learned model than in our hand-made model. This is counter to some research we have found when building our model in the first assignment and we will discuss some of the implications of this in section 4 [7].

Furthermore, it is interesting to look at the direction of causal relations that the structure learning algorithm returns. Firstly, there are two relations for which the algorithm cannot distinguish the direction of a causal relationship, namely GDP and Health, and Freedom and Generosity. Specifically for GDP and Health, this seems interesting to us as we had also found scientific literature that said the relation can go both ways, depending on what kind of circumstances were considered [12, 13, 14]. While this is not the way the PC-algorithm orients relations, intuitively it seems unsurprising that if research cannot agree on a direction of causality, the conditional independence in data will also not provide such insights.

Next to the two relations of undefined direction, there are a few causal relations that the structure learning algorithm identified to be the other way around compared to our hand-made model. The most outstanding one here is the relation between social support and negative affect. We had assumed a relation between social support to negative affect due to the way the correlation was presented in the World Happiness Report [6]. The structure learning algorithm, however, has found it to be the other way around. For a list of all differences, see the two tables below, where Table 2a compares the connections of our hand-made model in the order that we added connections to our model and Table 2b compares the connections in regards to the causal strength of our hand-made model with the relations found by the structure learning algorithm. The reason we show both orders

is that each of them gives a different impression about the “mistakes”² that we made. In both, we see that the further down we are, the more connections we have that the learned structure is missing, hence we may have been rather sensitive in our model-testing phase. In Table 2a we see that the mistakes we made about direction are more so in the lower part of the table, whereas in Table 2b there are a few directional mistakes in the top part of the table. Interestingly, we see that corruption is one of the variables that is part of a lot of the relations where our direction is the opposite of the learned direction. In our worldview, corruption is thus more so something that people are exposed to and has a causal effect, while the learned structure leads us to think that corruption is more so an outcome of other causes.

Causal relationship	New	Causal relationship	New
GDP → Life Ladder		GDP → Health	↔
Social Support → Life Ladder		GDP → Social Support	
Social Support → Positive Affect		Freedom → Positive Affect	
Social Support → Negative Affect	←	Corruption → Freedom	←
Health → Life Ladder		Social Support → Negative Affect	←
Health → Negative Affect		GDP → Life Ladder	
Freedom → Life Ladder		Corruption → GDP	
Freedom → Positive Affect		Social Support → Freedom	
Freedom → Negative Affect		Freedom → Generosity	↔
Generosity → Positive Affect		Positive Affect → Life Ladder	
Positive Affect → Life Ladder		Social Support → Positive Affect	
Corruption → Life Ladder		Corruption → Generosity	←
Corruption → Negative Affect	←	Social Support → Life Ladder	
GDP → Health	↔	GDP → Generosity	
Social Support → Health		Health → Life Ladder	
GDP → Social Support		Corruption → Negative Affect	←
Corruption → GDP		Generosity → Positive Affect	
Corruption → Freedom	←	Health → Negative Affect	
Social Support → Freedom		Corruption → Life Ladder	
Corruption → Generosity	←	Social Support → Health	
Corruption → Health		Negative Affect → Positive Affect	
Freedom → Generosity	↔	Freedom → Life Ladder	
Corruption → Social Support		Corruption → Health	
GDP → Generosity		Freedom → Negative Affect	
Negative Affect → Positive Affect		Corruption → Social Support	

(a) Comparison of relations in the order that we added the relations in our hand-made model. Arrows indicate a difference of direction or undefined direction, empty space indicates it is the same and blacked out space means the relation does not exist in the learned model.

(b) Comparison of relations in order of strength found in hand-made model. Arrows indicate a difference of direction or undefined direction, empty space indicates it is the same and blacked out space means the relation does not exist in the learned model.

²As a model cannot be proven correct, but can only be disproven, clearly it cannot truly be called mistakes. It would be more correct to describe it as superfluous connections or simply the differences between our hand-made model and the learned structure. We call it mistakes in this paragraph for readability of grouping together all the differences.

3.1.2 Comparison of Causal Relationships

To further look into differences between our hand-made model from assignment one and the one that was learned through the PC-algorithm, we again use `sem` as we did in the first assignment [7]. Both full graphs can be found in Appendix A to save space and we instead included the tables with the sorted estimates below in Table 3a and Table 3b. Because we used the `sem`-function, which can handle bidirectional arrows (which also represents implicit latent variables), we did not need to decide on which way to orient the two relations between GDP - Health and Freedom - Generosity. However, we did again do a test where we directed the relations and used `bnlearn` to fit the coefficients, which indeed did not yield different results.

Causal relationship	Effect		
GDP → Health	0.77		
GDP → Social Support	0.70		
Freedom → Positive Affect	0.42		
Corruption → Freedom	-0.41		
Social Support → Negative Affect	-0.40		
GDP → Life Ladder	0.39		
Corruption → GDP	-0.35		
Social Support → Freedom	0.31		
Freedom → Generosity	0.29		
Positive Affect → Life Ladder	0.23		
Social Support → Positive Affect	0.21		
Corruption → Generosity	-0.21		
Social Support → Life Ladder	0.20		
GDP → Generosity	-0.20		
Health → Life Ladder	0.17		
Corruption → Negative Affect	0.17		
Generosity → Positive Affect	0.16		
Health → Negative Affect	0.12		
Corruption → Life Ladder	-0.12		
Social Support → Health	0.11		
Negative Affect → Positive Affect	-0.11		
Freedom → Life Ladder	0.05		
Corruption → Health	-0.04		
Freedom → Negative Affect	-0.03		
Corruption → Social Support	0.02		
(a) Causal relationship sorted in descending order by strength of effect (i.e. absolute value of the coefficient)			
Causal relationship	Effect		
GDP ↔ Health	0.83		
GDP → Social Support	0.62		
Freedom → Positive Affect	0.42		
GDP → Life Ladder	0.39		
Freedom → Corruption	-0.39		
Freedom ↔ Generosity	0.32		
Negative Affect → Social Support	-0.29		
Positive Affect → Life Ladder	0.23		
Social Support → Life Ladder	0.20		
Health → Life Ladder	0.17		
Negative Affect → Corruption	0.16		
Generosity → Positive Affect	0.16		
Generosity → Corruption	-0.14		
Corruption → Life Ladder	-0.12		
Negative Affect → Positive Affect	-0.11		
Corruption → Health	-0.02		
(b) Causal relationships in the learned structure, sorted in descending order by strength of effect (i.e. absolute value of the coefficient)			

One of the interesting things to note are the three strongest causal estimates in both our models (i.e. GDP-Health, GDP-Social Support, Freedom-Positive Affect), which are the same. Similarly, the last three in the learned model, are also in our hand-made model very low (i.e. Corruption-Life Ladder, Negative Affect-Positive Affect, Corruption - Health). Thus qualitatively, there are quite some similarities between the two models. Similarly, the causal relation between GDP and life satisfaction is also comparable in both models (i.e. quantitatively the same, qualitatively similar). As before, it is still interesting to see that the learned model has no causal relation between Corruption

and GDP (or Social Support and Freedom) as those are quite strong in our hand-made model. We will go further into this in section 4.

3.2 Focal Relationships - Covariate adjustment

As said in our methodology, we chose two focal relationships, namely the one between GDP and life satisfaction and the one between positive affect and life satisfaction. We choose to do a covariate adjustment on this relation, to investigate the causal effect and whether we have estimated this correctly in the above models. To do covariate adjustment, it is necessary to find adjustment sets, which depend on the direction of the arrows. Therefore, we will now specify the direction of the CPDAG that we retrieved from our PC-algorithm. We chose the same direction as we had in our hand-made model, as this was based on earlier research done [15, 16, 14], thus $\text{GDP} \rightarrow \text{Health and Freedom} \rightarrow \text{Generosity}$ ³.

3.2.1 GDP \rightarrow Life Satisfaction

We will begin with the relationship between GDP and life satisfaction and compare our hand-made model with the one that we retrieved through the structure learning algorithm. To do so, we calculate the adjustment sets for both our models and then use this in a regression model to find the effect size. In our hand-made model, the only confounder between GDP and life satisfaction is corruption. In the model that we retrieved through the structure learning algorithm, there is no confounder biasing the causal path between GDP and the variable life ladder. As our data is continuous, we calculate the effect size of GDP on life satisfaction using a linear regression model on our scaled data. Unadjusted that effect size is returned to be 0.79 (which is thus also the true effect size if we assume the learned structure to be correct), while when we adjust for corruption, we get 0.72 as a coefficient⁴.

3.2.2 Positive Affect \rightarrow Life Satisfaction

As one of our models did not need any adjustment in the above relationship, we also investigated the relation between positive affect and life satisfaction as a second relationship. We consider this relationship to be interesting as the difference between measuring life satisfaction and measuring laughter, enjoyment and interest is sometimes doubted in literature [17, chapter 8]. Therefore, we also decided to do a covariate adjustment with the variable positive affect as the exposure. We again get two different adjustment sets for both our models. For our hand-made model, the minimal one consists of Corruption, Freedom, GDP, Health and Social Support. For the learned model, this is Corruption, Freedom, GDP and Social Support. Like above, we use our scaled data and a linear regression model to calculate the effect size of the unadjusted relation between Life Ladder and Positive Affect and then again the adjusted effect size. We use both adjustments once, to continue to compare our hand-made model with the learned model. The unadjusted coefficient that the linear regression model returns is 0.51. The adjustment of our hand-made model results in a coefficient of 0.29 and if we adjust according to the learned model, we receive an effect size of 0.24.

³An interesting note to add, due to an earlier mistake we had done all of this already with the undirected CPDAG and it turns out that most of the adjustment set results are the same as with the orientation of arrows we have chosen.

⁴Again noting on our earlier mistake: With the undirected graph the adjustment set would have been health. Adjusting for this as a confounder leads to a coefficient of 0.6 for GDP to Life Ladder.

4 Discussion and Conclusions

In the following section we will summarise our results as described above and discuss what possible consequences are and where future work is needed. Furthermore, we will try to come to some conclusions about which model we believe is better.

4.1 Implications of the differences in our models

As we found in the result section, the qualitative results of the learned structure do not differ much from the structure that we had created ourselves in the first assignment [7]. With this we mean, that a lot of the main relationships are still present and also in a similar order when we calculate the causal estimates. However, as already said in the result section, it is still interesting to see that the learned model has no causal relation between, for example, Corruption and GDP (or Social Support and Freedom) as those are quite strong in our hand-made model. Specifically for Corruption and GDP, we find this interesting when thinking about consequences for the real world. Because in our model the GDP is such a strong influence on life satisfaction and the GDP in turn is strongly negatively influenced by corruption, our model may suggest that addressing corruption in a country may be a good policy to improve overall life satisfaction. This is not a conclusion that the learned model would suggest intuitively. This similarly holds for other variables that are, according to our model, influenced by corruption, while the learned model connects corruption much less (i.e. corruption is conditionally independent of more variables). Thus, one could conclude that for the simple research question to see whether the GDP is a good welfare measurement for nations, the learned structure is more appropriate as it is simpler and largely shows the same result regarding life satisfaction. If we, however, wanted to go a step further and see how to influence life satisfaction, then our model may be a bit more appropriate as it shows more relations that may be of relevance to such a question.

While there are no immediately obvious reasons for us to dismiss the learned model as less believable, we would still believe that our hand-made model is more accurate. The reason we believe this is that the learned model seems to miss some of the more relevant relations, such as health and positive or negative affect. While we only included negative affect in our model, we did already mention that as a limitation in our first assignment [7]. We would also argue, that simply from a personal experience, it is easy to see that sickness (so decreased health) would lead to more negative emotions (i.e. affect). While this is only one of the connections, we believe that for a complex domain such as happiness, life satisfaction and other subjective measurements, the learned model may leave too many conditional independencies, that we do not find convincing enough. Therefore, we would argue that our model represents the reality more. However, it would always hold that more research is needed and for a full investigation, more variables and measurements are needed.

4.2 Covariate Adjustments

Our results for the covariate analysis of the causal effect of the GDP on life satisfaction showed that no matter which underlying structure (i.e. our hand-made or the learned model) we assume to be true, the effect size is not biased by a lot and there seems to be a strong causal relationship. In our hand-made model, we do seem to be slightly overestimating the causal relation between GDP and life satisfaction. As our research question is trying to answer whether the GDP is a good measurement of welfare, an overestimation is unfortunate and may suggest that it is an even less complete measurement than the model had indicated before. However, as said the difference is not very big and the effect size is in both cases still pretty strong. Therefore, this does probably not yet give

convincing evidence that the GDP (out of the here considered measurements) is a bad measurement for welfare.

The covariate analysis of the causal effect of positive affect on life satisfaction indicated a strong negative bias in the causal relation. Because it is a negative bias, this means that the effect of positive affect on life satisfaction is even less than our models would have indicated. This is an interesting result to investigate as it means there seems to be little causal effect for hedonism towards life satisfaction. Future work in research areas such as psychology and philosophy could use this result when considering what happiness, hedonism, life satisfaction and welfare mean [17, chapter 8]. Both of our models do not differ a lot here, so there is no real argument to be made here for which one is more appropriate.

Overall, our covariate adjustments do not seem to give us clear indications for either of the models to be more reliable than the other. Therefore, we would continue to argue, as above, that in a complex field such as subjective well-being, a model with more connections is closer to reality and gives us more insights.

A Appendix: Causal relations in Graphs

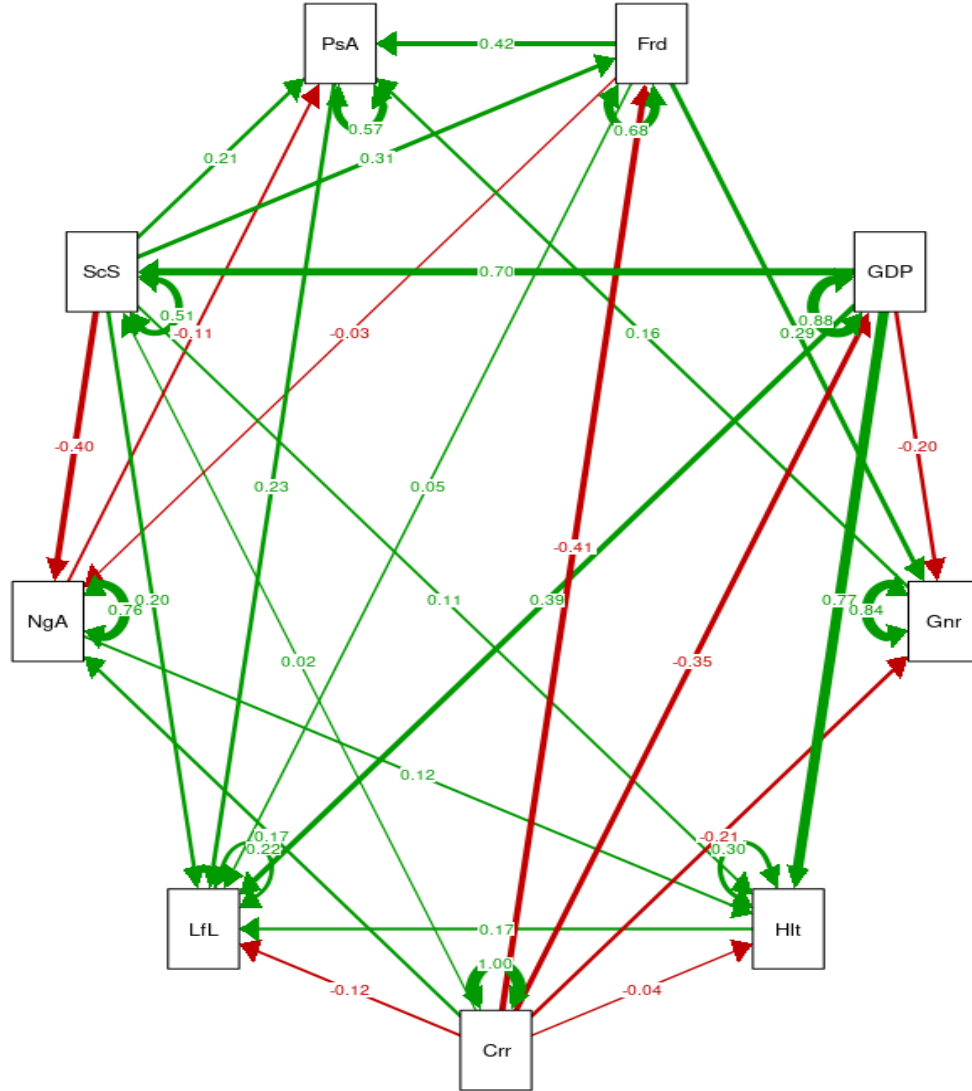


Figure 3: The causality of our handmade DAG

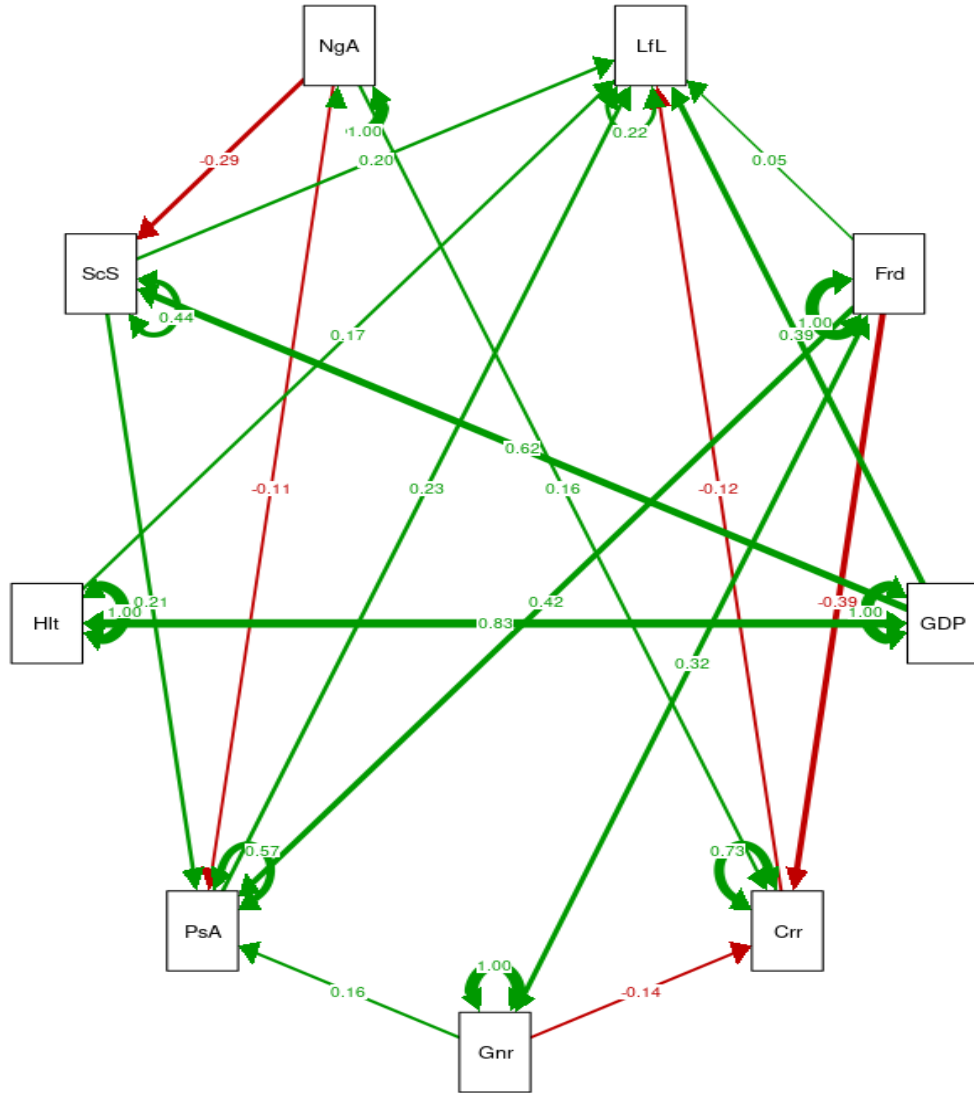


Figure 4: The causality when we keep the originally learned structure (including bidirectional relations)

References

- [1] Van Den Bergh and Jeroen C. J. M. Abolishing GDP, February 2007.
- [2] Jeroen C. J. M. van den Bergh. The GDP paradox. *Journal of Economic Psychology*, 30(2):117–135, April 2009.
- [3] R Cuijpers. Gdp and happiness: Gross national happiness, the new gdp. *Erasmus School of Economics Department of Applied Economics*, 2009.
- [4] Robert Costanza, Maureen Hart, John Talberth, and Stephen Posner. Beyond GDP: The Need for New Measures of Progress. *The Pardee Papers*, January 2009.
- [5] Richard A Easterlin and Kelsey O’Connor. The easterlin paradox. *Available at SSRN 3743147*, 2020.
- [6] Layard R. Sachs J. D. Akin L. B. De Neve J.-E. Wang S. (Eds.) Helliwell, J. F. *World Happiness Report 2023 (11th ed.)*. Sustainable Development Solutions Network, 2023. Accessed on: 2. Oct 2023.
- [7] Thijs de Jong and Marie-Sophie Simon. Assignment 1 bayesian networks - causality in the world happiness report. October 2023.
- [8] M Kalisch, A Hauser, M H Maathuis, and Martin Mächler. An Overview of the pcalg Package for R. March 2023.
- [9] Committee for proprietary medicinal products (CPMP) points to consider on adjustment for baseline covariates. *Statistics in Medicine*, 23(5):701–709, March 2004.
- [10] Lisa Moore. Exploring trends and factors in the world happiness report. 2020. Publisher: Dublin Business School.
- [11] Chris Tofallis. Which formula for national happiness? *Socio-Economic Planning Sciences*, 70:100688, June 2020.
- [12] Howard Waitzkin. Report of the WHO Commission on Macroeconomics and Health: a summary and critique. *The Lancet*, 361(9356):523–526, February 2003. Publisher: Elsevier.
- [13] Quamrul H. Ashraf, Ashley Lester, and David N. Weil. When Does Improving Health Raise GDP? *NBER Macroeconomics Annual*, 23:157–204, January 2008. Publisher: The University of Chicago Press.
- [14] Michael Marmot. The Influence Of Income On Health: Views Of An Epidemiologist. *Health Affairs*, 21(2):31–46, March 2002. Publisher: Health Affairs.
- [15] Jane B. Baron. Do We Believe in Generosity: Reflections on the Relationship between Gifts and Exchanges Commentary. *Florida Law Review*, 44(3):355–364, 1992.
- [16] Hasana Sharp. Generosity as Freedom in Spinoza’s Ethics. In Jack Stetter and Charles Ramond, editors, *Spinoza in Twenty-First-Century American and French Philosophy: Metaphysics, Philosophy of Mind, Moral and Political Philosophy*, pages 277–288. Bloomsbury, 2019.

- [17] Christopher Newfield, Anna Alexandrova, and Stephen John. *Limits of the Numerical: The Abuses and Uses of Quantification*. University of Chicago Press, June 2022. Google-Books-ID: AJVuEAAAQBAJ.