

Causal Discovery on Precarity and Depression

Kyuri Park¹, Leonie K. Elsenburg², Mary Nicolao², Karien Stronks², Vítor V. Vasconcelos^{1,3}

¹*Computational Science Lab, Informatics Institute, University of Amsterdam, PO Box 94323, Amsterdam, 1090GH, the Netherlands*

²*Department of Public and Occupational Health, Amsterdam Public Health Research Institute, Amsterdam UMC, University of Amsterdam, Amsterdam, the Netherland*

³*Institute for Advanced Study, University of Amsterdam, Oude Turfmarkt 147, Amsterdam, 1012GC, the Netherland*

2025-02-11

Abstract

Understanding the causal mechanisms linking precariousness and depression is critical for developing effective interventions. This study utilizes data from the HELIUS cohort study to explore these relationships using advanced causal discovery methods. By applying algorithms such as FCI, and CCI, and combining traditional Gaussian CI tests with non-parametric approaches like RCoT, we investigate how different indicators of precariousness—including factors related to employment, social relations and relational stress, financial situation, and housing—affect depressed mood, both as a sum score and at the individual symptom level. Our findings reveal that relational stress consistently emerges as a potential causal factor for depressed mood, while symptoms such as sleep disturbances, guilt, and anhedonia are particularly sensitive to external stressors, acting as potential early warning signals or intervention points for prevention. Moreover, the results highlight complexities in the data, including the influence of latent confounders and the challenges of capturing cyclic relationships. Despite some limitations, such as unresolved ambiguities in causal directions and challenges with mixed data distributions, this study demonstrates the utility of causal discovery tools in disentangling the intricate interplay between social and mental health dynamics. By mapping these causal structures into computational models, future research can simulate intervention effects, providing actionable insights to mitigate the impact of precariousness on mental health. This study serves as a foundational effort, offering both methodological advancements and practical implications for addressing depression at a population level.

Table of contents

1	Introduction	3
2	Methods	3
2.1	Data	3
2.2	Causal Discovery	5
2.3	Analysis	7
3	Results	9
3.1	Depression as sum score	9
3.2	Individual depression symptom	10
3.3	Precarity as sum score	13
4	Discussion	15
5	References	19
6	Appendix	20
6.1	Precariousness factors by Leonie	20
6.2	Exploratory Factor Analysis (EFA)	22
6.2.1	Factor Loadings (Pattern Matrix)	23
6.2.2	Variance Explained	23
6.2.3	Factor Intercorrelations	23
6.2.4	Model Fit Statistics	23
6.2.5	Summary	24
6.3	PCA	24
6.3.1	Explained variance (contributions) of variables	25
6.3.2	Cos ² Values	25
6.4	ICA	27
6.4.1	Dominant Variables per Component:	27
6.5	Hierarchical clustering	27
6.5.1	Using Euclidean distance	27
6.5.2	Using Mutual Information	30
6.6	Conclusions on Precariousness factors	31
6.7	Results from CCI algorithm	32
6.8	Results from PC algorithm	36
6.9	Randomized Conditional Independence / Correlation Test (RCIT & RCoT)	38
6.9.1	Kernel-Based Conditional Independence Testing	38
6.9.2	Random Fourier Features (RFFs)	39
6.9.3	Differences Between RCIT and RCoT	40

1 Introduction

Mental health problems in urban areas are on the rise, yet the complexity of mental health systems poses significant challenges in understanding the underlying mechanisms driving these issues — let alone designing effective interventions. Research has long sought to identify underlying factors contributing to mental health problems. Recent research looked into the association between mental health and indicators of precariousness in various dimensions of life, indicating a high level of uncertainty and instability in people's lives, such as factors related to employment, social relations, finances, housing, and culture (ref Leonie's paper). This comprehensive perspective on precariousness helped to highlight how different aspects of life may be interconnected. While this research has advanced our understanding of which indicators of precariousness are related to mental health, a key question remains unanswered: how do these factors influence mental health? Specifically, the lack of directional insight — knowing what influences what — limits our ability to identify and prioritize effective intervention targets.

This study aims to investigate the causal relationships between different dimensions of precariousness and mental health outcomes, to with a specific focus on depression, the most prevalent mental health issue. Using causal discovery methods, we explore how various aspects of precariousness contribute to depression and whether depression, in turn, exacerbates precariousness, potentially creating a feedback loop. Furthermore, we zoom in at the symptom level, examining individual depressive symptoms to identify those that may act as initiators — triggering the activation of the symptom network — by displaying heightened sensitivity to specific precariousness factors. We also investigate symptoms that may function as bridging elements, linking the dynamics of depression with precariousness by feeding back into precarious conditions, potentially reinforcing the cycle of mental health deterioration and precarious conditions. Through this analysis, we aim to uncover the causal mechanisms driving mental health challenges related to precariousness and establish a foundation for more targeted and effective interventions by identifying the key driving forces within this system.

2 Methods

2.1 Data

We use data from the HELIUS study, a multi-ethnic cohort that includes participants of Dutch, Turkish, Moroccan, Surinamese and Ghanaian origin (Snijder et al., 2017). To operationalize precariousness factors, we draw on the framework outlined in previous research (i.e., Leonie's paper) and select a set of relevant variables.

To ensure a robust representation of each precariousness factor, we conducted

various exploratory analyses to identify consistent and meaningful factor structures. Based on these analyses, we identified five precariousness factors, including two related to recent stressors, each comprising multiple variables as outlined below. Detailed information on the exploratory analyses can be found in the [Appendix](#).

- Employment precariousness: `emp_stat`, `work_sit`.
- Social precariousness: `soc_freq`, `soc_adq`.
- Housing precariousness: `nb_safe`, `nb_res`, `nb_rent`, `cul_rec`.
- Recent relational stressors: `frd_brk12`, `conf12`.
- Recent financial stressors: `fincr12`, `inc_diff`.

After preprocessing, the HELIUS dataset comprises 21,628 samples. Along with the five precariousness factors, we also compute an overall precarity score as the combined value of these factors. Additionally, PHQ-9 scores are included to represent depression, both as a total sum score and as individual symptom scores. In the subsequent causal discovery analysis, we examine the relationship between depression and precarity using both aggregated sum scores and their individual-level representations. Refer to Figure 1 for the overall distributions of the variables used in the analysis.

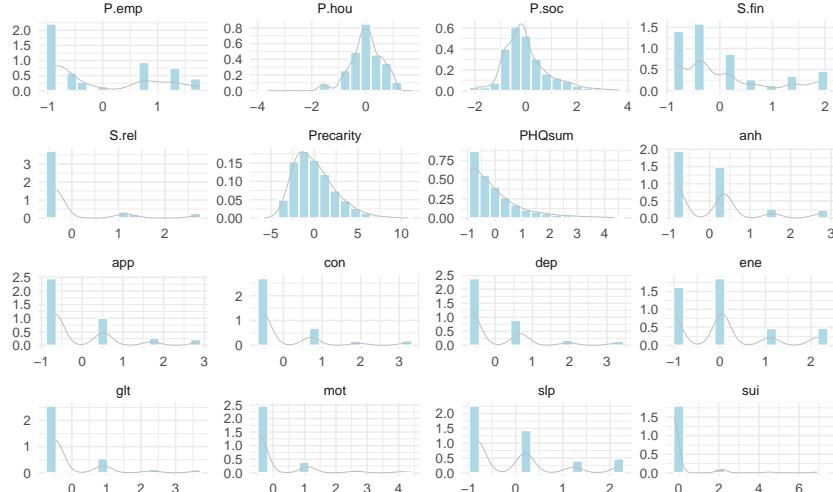


Figure 1: Distributions of variables with density overlay. `P.emp` = employment precariousness; `P.hou` = housing precariousness; `P.soc` = social precariousness; `S.fin` = recent financial stressors; `S.rel` = recent relational stressors; `Precarity` = overall precarity; `PHQsum` = PHQ-9 sum score; `anh` = anhedonia; `app` = appetite; `con` = concentration; `dep` = depressed mood; `ene` = energy; `glt` = guilty; `mot` = motor; `sui` = suicidal

2.2 Causal Discovery

There are numerous causal discovery algorithms available; however, in this study, we focus on algorithms suited to the potential cyclic relationships within our system. Specifically, we use *FCI* (Fast Causal Inference) and *CCI* (Cyclic Causal Inference), both capable of accounting for such cycles (Mooij & Claassen, 2020; Strobl, 2019). Additionally, we include the *PC* algorithm as a reference, given its simplicity and prominence as one of the most widely known causal discovery methods (Spirtes et al., 2001).

Table 1: Assumptions of causal discovery algorithms

Algorithm	Acylicity	Causal sufficiency	Absence of selection bias	Linearity	Output
PC	✓	✓	✓	✓	CPDAG
FCI	— ^a	✓	— ^a	— ^a	PAG
CCI	x	x	x	✓	(partially oriented) MAAG

Note. ^aThe FCI algorithm, introduced by Spirtes (1995), is a constraint-based causal discovery method for DAGs that accounts for latent confounding and selection bias. Mooij & Claassen (2020) later showed its applicability to cyclic causal discovery with latent confounding under general faithfulness and Markov conditions, assuming non-linear causal relationships.

As shown in Table 1, the resulting graphs from FCI and CCI differ slightly (*PAG*: partial ancestral graph; *MAAG*: maximal almost ancestral graph) due to their reliance on different underlying assumptions. Despite these differences, both graphs belong to the class of *ancestral graphs*, which are designed to encode causal relationships between variables, where the presence of an edge indicates causal *ancestry*. In these graphs, directed edges, $A \xrightarrow{*} B$, indicate that B is not an ancestor of A in every graph within the Markov equivalence class, $\text{Equiv}(G)$. The Markov equivalence class represents a set of graphs that encode the same conditional independence relationships, ensuring that the same *d-separation* conditions hold across all graphs in the class (Spirtes et al., 2001). Conversely, an edge marked as $A \xrightarrow{*} B$ indicates that B is an ancestor of A across all graphs in $\text{Equiv}(G)$. Circle endpoints, $A \xrightarrow{*} \circ B$, represent ambiguity in the ancestral relationship, meaning B 's ancestral status relative to A varies across graphs in $\text{Equiv}(G)$.¹ Finally, when an edge is represented as $A \leftrightarrow B$, it implies that neither A nor B is an ancestor of the other, suggesting the presence of a latent confounder influencing both variables.

The algorithms also differ in how they detect cycles and represent cyclic relationships in their resulting graphs. In CCI's MAAG, $A — B$ indicates that A is an ancestor of B , and simultaneously B is an ancestor of A , referring to a cyclic relationship between $A \xleftarrow{*} B$. FCI, on the other hand, identifies potential cycles

¹* serves as a *meta-symbol*, representing one of the three possible edge-endpoints. For instance, $A \xrightarrow{*} B$ can indicate any of the following edges: $A — B$, $A \rightarrow B$, or $A \circ B$ (Park et al., 2024).

more subtly. In its PAG, fully-connected nodes with circle endpoints ($\circ\text{---}\circ$) may suggest the presence of cyclic structures. FCI, therefore, provides a sufficient condition to distinguish variables that are not part of a cycle, offering a more nuanced approach to handling cyclic relationships (Mooij & Claassen, 2020).

Table 1 highlights further differences in the assumptions underlying these algorithms. CCI operates under the assumption of a linear system, whereas FCI, particularly when used to infer cyclic relationships, assumes a non-linear system without selection bias and adheres to more general faithfulness and Markov conditions (i.e., σ -separation and σ -faithfulness setting) (Forré & Mooij, 2018). When the respective assumptions of each algorithm are met, their inferred cyclic relationships align with those illustrated in Figure 2.

While CCI’s MAAG provides an explicit representation of cyclic relationships, it has certain theoretical limitations, as the MAAG may not always fully preserve d-separation relations from the original graph, \mathcal{G} (Strobl, 2019). Given these considerations, we examine and compare results from both algorithms, placing more weight on the FCI results, which are presented in the main results section, while discussing CCI results where relevant, with detailed findings provided in the [Appendix](#). A full discussion of causal discovery concepts and algorithmic details is beyond the scope of this paper; however, readers seeking a more in-depth understanding can refer to Park et al. (2024) for a comprehensive exploration of these methods and their applications.

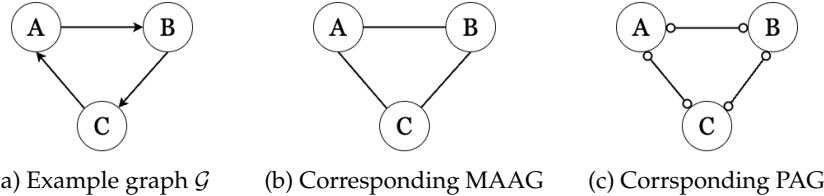


Figure 2: Example graph \mathcal{G} featuring a cycle and the corresponding MAAG from CCI and PAG from FCI.

The graph produced by the PC algorithm is a CPDAG (completed partially directed acyclic graph), where directed edges ($A \rightarrow B$) indicate that A is a direct cause (parent) of B . Unlike FCI and CCI, the CPDAG does not include circle symbols. Instead, when the PC algorithm cannot determine the direction of causality, it represents this uncertainty with bidirectional arrows. While the PC algorithm serves as a useful reference, its strict assumptions —acyclicity and the absence of latent confounders—limit its applicability in more complex scenarios. For this reason, our primary focus remains on the results obtained from FCI and CCI, with all PC algorithm results provided in the [Appendix](#) for completeness.

One practical challenge in applying these algorithms to the HELIUS dataset is that most variables do not follow a Gaussian distribution, and the relationships

among variables are unlikely to be strictly linear. To address this, we supplement the commonly used Gaussian conditional independence (CI) test, which relies on partial correlations, with a non-parametric CI test based on reproducing kernel methods (Zhang et al., 2012). While kernel-based conditional independence tests (KCIT) are theoretically effective in capturing non-linear dependencies and subtle interactions in data with non-Gaussian noise, they are computationally demanding for large datasets like HELIUS. Their quadratic scaling with sample size is primarily due to the need to invert large kernel matrices, making them impractical for large-scale analysis (Rahimi & Recht, 2007). To mitigate this issue, we use the Randomized Conditional Correlation Test (RCoT), which approximates kernel methods using random Fourier features. This approach reduces computational complexity from quadratic to linear scaling with sample size, significantly lowering computational costs while maintaining robustness (Strobl et al., 2019). For a more detailed explanation of RCoT, refer to Section 6.9.

2.3 Analysis

We analyze the causal structure using three complementary approaches: (1) Examining the relationship between the five individual precariousness factors and the PHQ sum score, which represents overall depression severity. (2) Investigating the relationship between individual depressive symptoms and individual precariousness factors. (3) Assessing the relationship between individual depressive symptoms and overall precariousness (the sum of all precariousness factors). Using the PHQ sum score as a measure of depression simplifies the analysis by reducing dimensionality, improving computational efficiency, and providing a broad, interpretable perspective on how precariousness factors relate to overall depression severity. However, this aggregated approach may overlook granular symptom-level relationships.

By zooming in on individual symptoms, we capture a more detailed picture of how specific depressive symptoms interact with different precariousness factors. This approach allows us to map individual causal pathways, identifying which symptoms are particularly sensitive to certain precarious conditions and vice versa. However, this level of detail introduces methodological challenges, including non-standard distributions of symptom variables and the high-dimensional nature of the analysis, which can reduce statistical power and significantly increase computational demands—especially when using non-parametric conditional independence (CI) tests.

Finally, analyzing individual symptoms in relation to overall precariousness provides an integrative perspective, revealing how depression symptoms collectively relate to precarious conditions. This approach complements the fully disaggregated analysis by aggregating precariousness factors, potentially uncovering relationships that may be too weak to detect when examined separately. This is particularly relevant when a symptom is associated with multiple precariousness factors, but these associations are not strong enough to emerge in an individual-level analysis. By integrating these three approaches — ranging

from partially aggregated to fully disaggregated — we aim to achieve a balance between capturing detailed granularity and gaining a broader understanding of the relationship between precarity and depression.

Furthermore, to address sensitivity to parameter values and enhance the robustness of our results, we implement multiple settings combined with bootstrapping. For each condition, we generate 100 bootstrap samples, estimate causal graphs, and retain only the edge-endpoints that exceed predefined thresholds. If no edge-endpoint type surpasses the threshold, we assign a circle (\circ) to indicate uncertainty in the edge direction.

The analyses are conducted under the following conditions:

- Significance levels (α): 0.01 and 0.05
- Thresholds: 0.5, 0.6, 0.7, and 0.8
- CI test: Gaussian CI test, RCoT
- Algorithms: FCI, CCI, and PC

This setup yields 16 combinations (2 significance levels \times 4 thresholds \times 2 CI tests), applied across three algorithms. Each combination is repeated for 100 bootstrap samples, yielding a total of 1,600 graphs per algorithm. For analyses involving individual symptom variables, we simplify the setup by focusing on thresholds of 0.6 and 0.7 and reducing the number of bootstrap samples to 30, easing computational demands. To further enhance efficiency, we fix the skeleton of the symptom network using a common structure derived from PC, FCI, and CCI algorithms ($\alpha = 0.01$, RCoT CI test). This step constrains which edges can be present among symptoms but does not restrict the estimation of causal directions within the symptom network. Additionally, the entire causal structure between symptoms and precariousness factors remains fully estimable, ensuring that key relationships are still inferred. Beyond its computational benefits — particularly in reducing the time required for skeleton estimation — this fixed structure also aligns well with commonly reported symptom network structures in the literature (Park et al., 2025), ensuring consistency with existing findings.

To obtain a final, stable summary of causal relationships, we identify the most frequently occurring edge-endpoints across different experimental setups. First, within each condition, we determine the most frequent edge-endpoint from all bootstrap samples. Then, across different conditions, we aggregate these selected symbols and determine the overall most common edge-endpoint for each graph entry.

Since we have an even number of experimental setups, ties may occasionally occur. In such cases, we explicitly denote them in the resulting graph using dashed lines. To further help with the interpretation and understanding of the resulting graph, we provide a corresponding table for each graph, showing the proportion of each edge-endpoint type to clarify how dominant a given symbol is in each case.

By double summarizing the results in this way, we generate one final graph per algorithm for each analysis approach: the partially aggregated analysis, which uses either the sum of precariousness factors or the sum of depression factors to capture overall trends, and the fully disaggregated analysis, which examines individual precariousness and depression factors separately. This approach ensures that only stable and consistent edges — those that persist across different parameter settings — are retained, leading to a robust and reliable understanding of the relationships between precariousness factors and depression. See Figure 3 for an overview of the analysis workflow.

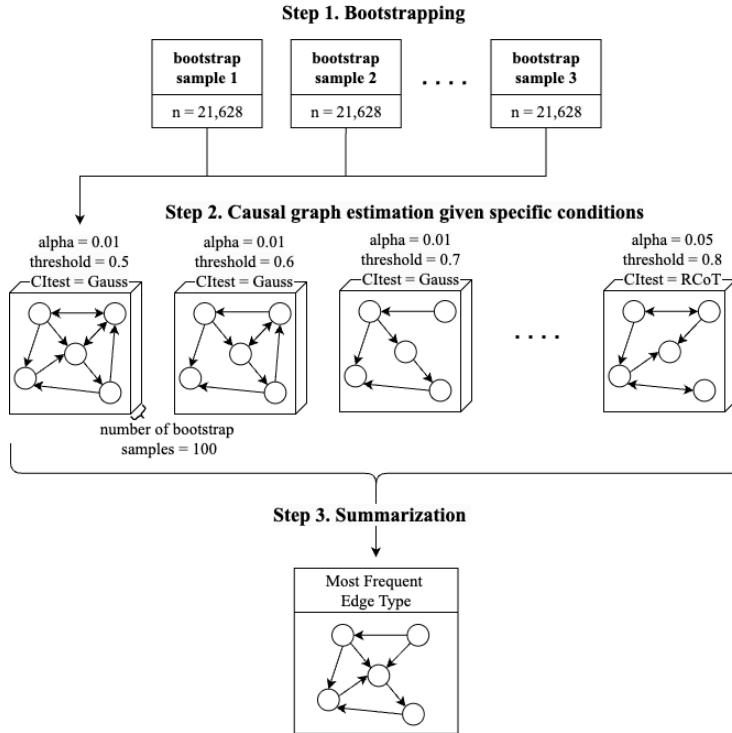


Figure 3: Analysis workflow applied across all three algorithms.

3 Results

3.1 Depression as sum score

The sum score graphs provide a high-level summary of how precariousness factors collectively influence overall depression severity, focusing on aggregated relationships. Figure 4 illustrates the causal relationships between precariousness factors (P_{hou} , P_{emp} , P_{soc} , S_{rel} , S_{fin}) and the depression sum score ($PHQsum$), as identified by the FCI algorithm. As described in Section 2.3, dashed edges

indicate ties, with the exact edge endpoint where the tie occurs is represented by a bold horizontal bar.

The causal graph in panel (a) highlights key pathways in the relationships between precarious factors and depression (PHQsum). Employment precariousness (P.emp) and social precariousness (P.soc) are not identified as causes of depression, whereas financial stress (S.fin) appears to play a causal role. While P.emp and P.soc are generally not recognized as causes of other precariousness factors, S.fin emerges as a potential cause, as indicated by its circle edge endpoint. Supporting this, the proportion matrix plot in panel (b) shows that S.fin has some probability of causally influencing either P.emp or P.soc.

The matrix plot can be interpreted such that the symbol in $\text{matrix}[i, j]$ represents the relationship $i - [\text{symbol}] j$. For example, if $\text{matrix}[i, j] >$ and $\text{matrix}[j, i] = \circ$, then the inferred relationship is $i \rightarrow j$. In the table, different edge types are represented by distinct colors: light gray for the absence of an edge, blue for circles, green for arrowheads, and coral for arrowtails. The colors are blended based on the proportion of each edge type, with higher proportions increasing the opacity of the corresponding color, making dominant symbols more visually prominent.

Relational stress (S.rel) plays a more nuanced role, interacting with depression and S.fin through a latent confounder. Its relationship with P.soc remains less certain, as the proportion table suggests a fair probability that this connection is absent, uncertain, or bidirectional. Meanwhile, housing precariousness (P.hou) is not directly causally related to depression but is linked to P.emp and S.fin. While P.emp and S.fin are identified as non-causes of P.hou, it remains unclear whether P.hou causally influences P.emp or if their relationship is mediated by an unobserved confounder.

The findings from the CCI algorithm (see Figure 7) are largely consistent with those from the FCI algorithm, with one exception: CCI does not identify S.fin as a cause of depression. Additionally, CCI introduces greater uncertainty in edge directions, with more ties mainly between circle and arrowhead endpoints, and tends to favor arrowheads more frequently than FCI. Despite these differences, the skeleton structure and overall causal directions derived from CCI align well with the results of the FCI algorithm, supporting the key pathways.

3.2 Individual depression symptom

Moving from the sum score representation to the disaggregated symptom-level graph provides a more granular perspective on the causal relationships between precariousness factors and depression symptoms. Unlike the sum score graph, which aggregates all symptoms into a single measure—potentially obscuring nuanced relationships—the symptom-level graph highlights heterogeneity in how precariousness factors influence individual depressive symptoms: con (concentration), slp (sleep), ene (energy), app (appetite), mot (motor), sui (suicidal), anh (anhedonia), glt (guilt), and dep (depressed mood), and vice versa.

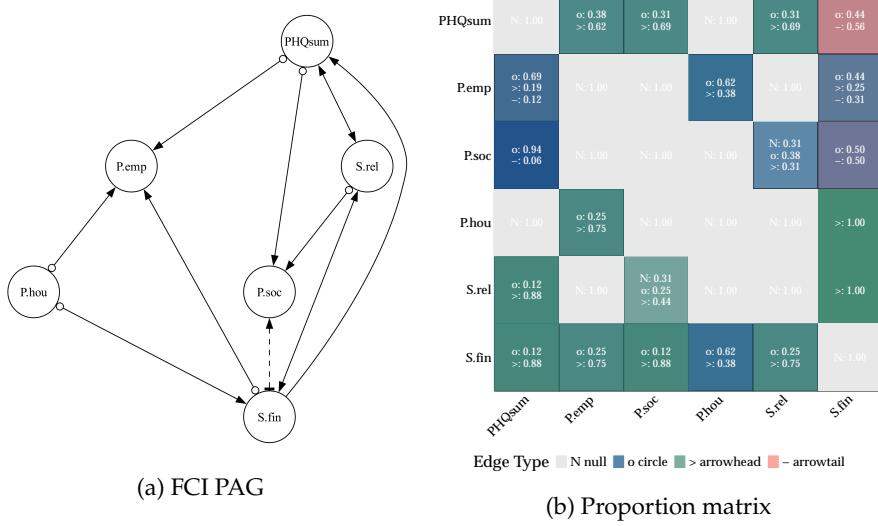


Figure 4: Resulting graph of precarious factors and depression sum score using FCI and proportion of edge endpoint types.

The symptom-level graph in Figure 5 panel (a) reveals a more complex and interconnected structure than the sum score graph, reflecting the strong interdependence among symptoms and suggesting much presence of latent confounders, as indicated by numerous bidirectional edges. As before, dashed edges denote ties, with specific ties marked by bold horizontal bars. In Figure 5 panel (b), the proportion matrix provides more insight into these directionalities. Here, opaque green dominates most symptom-to-symptom connections, indicating that arrowheads are the most frequently inferred edge type. However, blue regions, which correspond to circle endpoints, are particularly common for anh, slp, ene, and sui, reflecting uncertainty in the causal direction among these symptoms. Notably, the only arrowtail connection appears between dep and anh, suggesting that anhedonia (anh) is a potential cause of depressed mood (dep).

Looking at the symptom-precariousness connections, one of the key patterns is that ties are most frequently found in the relationships between individual symptoms and precariousness factors. This suggests that these causal links may be less stable across different conditions. Upon closer examination, most ties arise due to discrepancies between the Gaussian CI test and the RCoT test—with Gaussian CI favoring arrowheads and RCoT more often predicting the absence of an edge.

This discrepancy likely arises from a combination of factors. First, as the analysis shifts from aggregated to disaggregated variables, the number of conditional independence (CI) tests increases, leading to a loss of statistical power in high-

dimensional causal discovery. With more variables being conditioned on, the data becomes sparser, making it harder to detect weaker dependencies. Additionally, over-conditioning—controlling for too many variables—can artificially remove statistical associations, leading to false negatives. Beyond this general statistical power issue, the fundamental differences in how these CI tests operate contribute to the inconsistency. RCoT, as a nonparametric test, does not assume linearity and can capture both linear and nonlinear dependencies. However, it is more conservative and requires larger sample sizes, making it more prone to false negatives in high-dimensional settings. In contrast, partial correlation (Gaussian CI test) assumes linear Gaussian relationships, making it more permissive and sometimes detecting weak dependencies that may not be statistically meaningful. As a result, when Gaussian CI detects an edge while RCoT does not, it could indicate that the relationship is weak, strictly linear, or requires more data for reliable nonparametric detection. This explains why many ties in the proportion matrix show a 50-50 split between symbols—where half of the conditions with RCoT suggest the absence of an edge, while the other half using Gaussian CI test favor an arrowhead. This systematic divergence highlights the challenges of causal discovery in high-dimensional settings, where varying statistical assumptions can lead to differing conclusions about the presence and direction of causal relationships.

Despite these differences, some consistent patterns emerge across both CI tests, particularly in the case of financial stress (S.fin), which appears to be connected to nearly all other variables—though many of these connections are marked by ties, reflecting uncertainty in directionality. One exception is the stronger evidence suggesting that S.fin may cause changes in app (appetite), while its relationships with other symptoms, such as ene, mot, and dep, remain ambiguous. Similarly, social precarity (P.soc) exhibits numerous connections with depressive symptoms, with most edges pointing toward P.soc rather than outward from it. This suggests that depressive symptoms, particularly dep and glt, may contribute to worsening social precarity rather than the other way around. Additionally, anh also shows some probability of causally influencing P.soc, reinforcing the idea that social precarity is more often a consequence rather than a driver of depressive symptoms.

Other precariousness factors exhibit more uncertain but still notable relationships. Relational stress (S.rel) shows weak but existing connections with slp, glt, and app, though directionality remains unclear in many cases. Employment precarity (P.emp) also connects to symptoms like slp, mot, and con, but, like S.rel, these relationships exhibit a 50/50 split in directionality, reflecting uncertainty in the inferred causal paths. In line with the aggregated graph in Figure 4, housing precarity (P.hou) does not appear to have any direct relationship with depression symptoms but consistently shows associations with employment precarity (P.emp) and financial stress (S.fin). The causal relationships among precariousness factors remain largely unchanged from the aggregated analysis, with financial stress (S.fin) exhibiting the strongest tendency to influence other precariousness factors.

Among depressive symptoms, dep, glt, and slp appear to be the most connected to precariousness factors, while P.soc has the most connections with symptoms, predominantly as a recipient rather than a driver of influence. Within the symptom network, slp, sui, and anh emerge as causally influential symptoms, as they exhibit more outgoing arrows compared to other symptoms. On the other hand, con, mot, dep, and glt, despite having high connectivity, predominantly receive incoming arrows, indicating they are more likely effects rather than causes. Considering both symptom-precariousness connections and symptom-level dynamics, sleep disturbance (slp) emerges as a central symptom, given its strong ties to precariousness factors and its influential role within the symptom network. This suggests that sleep issues could be an initiating symptom, particularly sensitive to relational stress and employment precarity. On the other hand, social precarity (P.soc) serves as a key bridge between the depression and precariousness subsystems, as depressive symptoms appear to feed back into social precarity, reinforcing a self-sustaining dynamic between depression and precarious conditions.

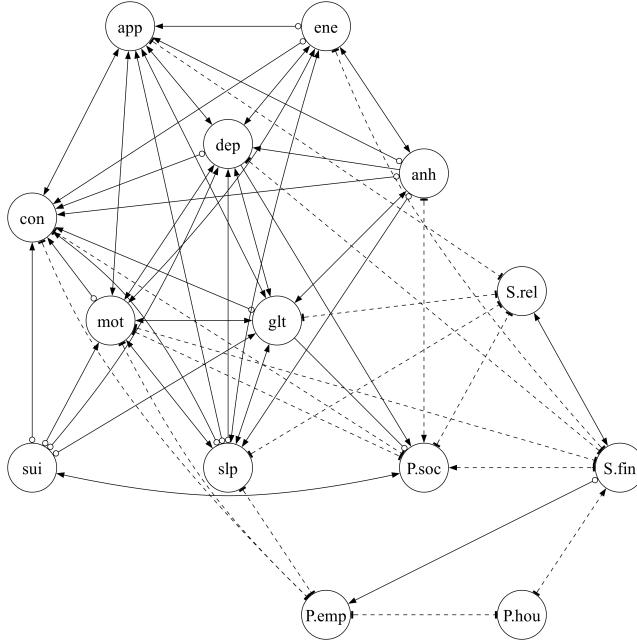
Finally, the results from CCI are largely consistent with those from FCI, but with some key differences. CCI tends to favor arrowheads more frequently, resulting in a greater number of bidirectional edges, which suggests a higher involvement of latent confounders. Additionally, CCI exhibits more tie situations, though, similar to FCI, most ties occur between absence and arrowhead edges, reflecting the discrepancy between the Gaussian CI test and RCoT—where the Gaussian test favors arrowheads, while RCoT more often suggests the absence of an edge. For a detailed visualization of the CCI-derived graph and proportion matrix, see Figure 8.

3.3 Precarity as sum score

Lastly, we examine the relationships between individual symptoms and overall precarity, an aggregated measure represented by the sum of five precariousness factors. This analysis provides a complementary perspective, capturing broader patterns that may not be evident in the disaggregated symptom-precariousness analysis. By aggregating precariousness factors into a single score, this approach may capture distributed relationships across different precarity factors that were previously overlooked in the disaggregated analysis.

Unlike the disaggregated graph in Figure 5, the aggregated symptom-precariousness graph (Figure 6 panel (a)) shows more certainty in causal directions, with only a few edges resulting in ties, all of which involve the overall precarity factor. This is even more evident in the proportion matrix (Figure 6 panel (b)), where symptom-to-symptom interactions almost fully converge to a single edge type. Additionally, all symptom interactions are connected either through bidirectional edges or a combination of an arrowhead and a circle, suggesting a strong presence of latent confounders.

Examining the symptom-precarity connections, we observe a stronger overall



(a) FCI PAG

	anh	> 1.00	> 1.00	> 1.00	> 1.00	> 1.00	> 1.00	N. 1.00	N. 1.00	> 1.00	> 1.00	N. 1.00	N. 1.00	N. 1.00
dep	$\alpha: 0.12$ -0.88	N. 1.00	$\alpha: 0.75$ -0.25	$\alpha: 0.12$ > 0.88	$\alpha: 0.38$ > 0.62	$\alpha: 0.25$ > 0.50	$\alpha: 0.12$ > 0.88	$\alpha: 0.38$ > 0.62	$\alpha: 0.62$ > 0.38	N. 1.00	$\alpha: 0.38$ > 0.50	$\alpha: 0.12$ > 0.50	N. 1.00	N. 1.00
slp	$\alpha: 1.00$	> 1.00	N. 1.00	$\alpha: 0.38$ > 0.62	$\alpha: 0.25$ > 0.75	> 1.00	> 1.00	> 1.00	N. 1.00	$\alpha: 0.50$ > 0.50	N. 1.00	N. 1.00	N. 1.00	N. 1.00
ene	> 1.00	> 1.00	> 1.00	N. 1.00	> 1.00	> 1.00	> 1.00	> 1.00	N. 1.00	N. 1.00	N. 1.00	N. 1.00	N. 1.00	N. 1.00
app	$\alpha: 0.62$ -0.38	> 1.00	$\alpha: 0.62$ -0.38	$\alpha: 0.62$ > 0.38	N. 1.00	> 1.00	> 1.00	> 1.00	N. 1.00	N. 1.00	N. 1.00	$\alpha: 0.25$ > 0.50	$\alpha: 0.50$ > 0.12	$\alpha: 0.38$
glt	> 1.00	> 1.00	> 1.00	N. 1.00	> 1.00	N. 1.00	> 1.00	> 1.00	$\alpha: 0.25$ > 0.75	N. 1.00	$\alpha: 0.25$ > 0.75	N. 1.00	N. 1.00	N. 1.00
con	$\alpha: 0.62$ -0.38	$\alpha: 0.25$ > 0.75	$\alpha: 0.75$ > 0.25	$\alpha: 0.62$ > 0.38	$\alpha: 0.12$ > 0.88	$\alpha: 0.75$ > 0.25	N. 1.00	$\alpha: 0.62$ > 0.38	$\alpha: 0.62$ > 0.50	N. 1.00				
mot	N. 1.00	> 1.00	> 1.00	> 1.00	> 1.00	> 1.00	> 1.00	N. 1.00	$\alpha: 0.88$ > 0.12	N. 1.00	$\alpha: 0.50$ > 0.50	N. 1.00	N. 1.00	N. 1.00
sui	N. 1.00	> 1.00	N. 1.00	N. 1.00	N. 1.00	> 1.00	> 1.00	N. 1.00	N. 1.00	N. 1.00	$\alpha: 0.25$ > 0.75	N. 1.00	$\alpha: 0.50$ > 0.50	N. 1.00
P.emp	N. 1.00	$\alpha: 0.50$ > 0.38	$\alpha: 0.12$ > 0.50	N. 1.00	N. 1.00	N. 1.00	N. 1.00	$\alpha: 0.50$ > 0.50	N. 1.00	N. 1.00	$\alpha: 0.88$ > 0.12	N. 1.00	N. 1.00	$\alpha: 0.62$ > 0.38
P.soc	$\alpha: 0.50$ -0.50	$\alpha: 0.38$ > 0.12	$\alpha: 0.12$ > 0.50	N. 1.00	N. 1.00	N. 1.00	$\alpha: 0.25$ > 0.50	N. 1.00	$\alpha: 0.25$ > 0.62	N. 1.00	$\alpha: 0.88$ > 0.12	N. 1.00	N. 1.00	$\alpha: 0.50$ > 0.50
P.hou	$\alpha: 0.50$ > 0.25	N. 1.00	N. 1.00	N. 1.00	$\alpha: 0.75$ > 0.25	N. 1.00	N. 1.00	N. 1.00	N. 1.00	$\alpha: 0.50$ > 0.50	N. 1.00	N. 1.00	N. 1.00	$\alpha: 0.38$ > 0.62
S.rel	N. 1.00	N. 1.00	$\alpha: 0.50$ > 0.50	N. 1.00	$\alpha: 0.50$ > 0.50	$\alpha: 0.50$ > 0.50	N. 1.00	N. 1.00	$\alpha: 0.50$ > 0.38	N. 1.00	$\alpha: 0.50$ > 0.50	N. 1.00	$\alpha: 0.25$ > 0.12	$\alpha: 0.12$ > 0.62
S.fin	$\alpha: 0.50$ -0.12	$\alpha: 0.38$ > 0.50	$\alpha: 0.12$ > 0.50	$\alpha: 0.88$ > 0.50	$\alpha: 0.50$ > 0.50	$\alpha: 0.50$ > 0.50	$\alpha: 0.50$ > 0.50	$\alpha: 0.12$ > 0.38	N. 1.00	$\alpha: 0.38$ > 0.62	$\alpha: 0.12$ > 0.88	$\alpha: 0.38$ > 0.88	$\alpha: 0.25$ > 0.25	$\alpha: 0.25$ > 0.50

Edge Type ■ N null ■ circle ■ > arrowhead ■ – arrowtail

(b) Proportion matrix

Figure 5: Resulting graph of precarity factors and individual depression symptoms using FCI and proportion of edge endpoint types.

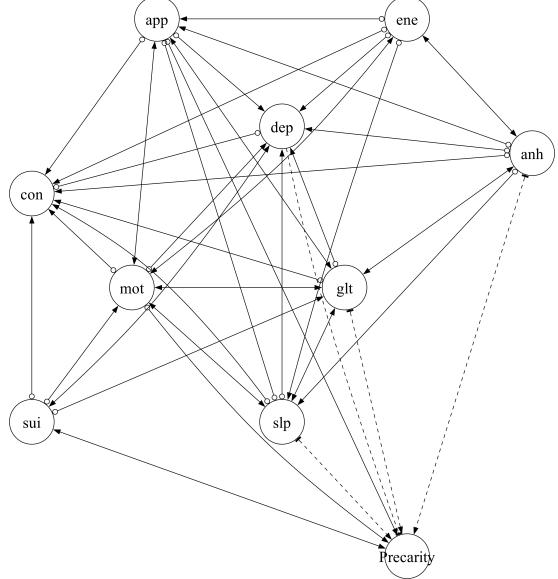
trend of depressive symptoms influencing precariousness, rather than the other way around. Specifically, anh, dep, and slp show a high probability of causing precariousness, whereas app, glt, and mot exhibit more uncertainty in directionality, often resulting in circle endpoints. This pattern suggests that when precariousness factors are aggregated, the dominant causal flow is from depressive symptoms to precariousness, rather than precariousness driving depression. Particularly, dep emerges as the strongest predictor of precariousness, exhibiting the highest proportion of arrowtails, consistent with findings from the disaggregated analysis.

Comparing this with the CCI-derived graph (Figure 9), several key differences stand out. The CCI results reveal that most symptoms are interconnected through bidirectional edges, suggesting a strong presence of latent confounders. However, dep appears to be a distinct exception, as it is predominantly caused by nearly all other symptoms, with one exception sui, which does not contribute to dep. Similarly, con is influenced by multiple symptoms, albeit with weaker support compared to dep. Most interestingly, the CCI results highlight a distinct feedback loop between dep and overall precariousness, represented by a tail-tail edge (—). This suggests a possible reinforcing cycle between depression and overall precariousness, where depression not only arises from precarious conditions but also contributes to their persistence, creating a self-sustaining dynamic.

Overall, the aggregated precariousness graph supports the key findings from both the disaggregated analysis and the depression sum score analysis, providing greater certainty in several important patterns. The results highlight the significant role of latent confounders in symptom-to-symptom interactions. While certain stressor-related precariousness factors, particularly S.fin and S.rel, appear to causally influence depressive symptoms, other precariousness factors are more likely driven by depression rather than acting as primary causes, suggesting that precariousness may function more as a consequence of depression than its source. Some symptoms, particularly glt, slp, and dep, show greater sensitivity to precarious conditions, while also playing a central role in symptom dynamics. Among them, slp and glt emerge as influential symptoms within the depression network, whereas con and dep, despite their high connectivity, behave more as outcomes rather than primary drivers. Together, these findings suggest that depression—especially symptoms like sleep disturbances and guilt—may contribute to sustaining precarious conditions, possibly forming a reciprocal cycle between depression and precariousness.

4 Discussion

This study examined the causal relationships between precariousness factors and depression using causal discovery algorithms, integrating both sum score and symptom-level analyses. By analyzing PHQ sum scores alongside individual symptom scores, and comparing aggregated precariousness scores with individual precariousness factors, we gained a comprehensive perspective on how different aspects of precariousness influence depression. This dual approach revealed



(a) FCI PAG

	anh	> 1.00	> 1.00	> 1.00	> 1.00	> 1.00	> 1.00	N. 1.00	N. 1.00	> 1.00
anh	N. 1.00							N. 1.00	N. 1.00	
dep	o: 1.00	N. 1.00	o: 1.00	o: 1.00	o: 1.00	o: 0.62 > 0.38	o: 1.00	o: 0.88 > 0.12	o: 0.38 > 0.62	o: 0.50 > 0.50
slp	o: 1.00	> 1.00	N. 1.00	o: 1.00	o: 1.00	> 1.00	> 1.00	> 1.00	N. 1.00	> 1.00
ene	> 1.00	> 1.00	> 1.00	N. 1.00	> 1.00	N. 1.00	> 1.00	> 1.00	N. 1.00	N. 1.00
app	o: 1.00	> 1.00	o: 1.00	o: 1.00	N. 1.00	> 1.00	> 1.00	> 1.00	N. 1.00	> 1.00
glt	> 1.00	> 1.00	> 1.00	N. 1.00	> 1.00	N. 1.00	> 1.00	> 1.00	o: 1.00	> 1.00
con	o: 1.00	o: 1.00	o: 0.88 > 0.12	o: 1.00	o: 1.00	o: 1.00	N. 1.00	o: 0.88 > 0.12	o: 0.62 > 0.38	N. 0.62 > 0.38
mot	N. 1.00	> 1.00	> 1.00	> 1.00	> 1.00	> 1.00	> 1.00	N. 1.00	o: 0.75 > 0.25	> 1.00
sui	N. 1.00	> 1.00	N. 1.00	N. 1.00	N. 1.00	> 1.00	> 1.00	> 1.00	N. 1.00	> 1.00
precarity	o: 0.50 - 0.50	o: 0.38 - 0.50	o: 0.50 - 0.50	N. 1.00	o: 0.62 - 0.38	o: 0.50 - 0.50	N. 0.62 - 0.38	o: 0.88 - 0.12	o: 0.12 - 0.88	N. 1.00

Edge Type ■ N null ■ o circle ■ > arrowhead ■ - arrowtail

(b) Proportion matrix

Figure 6: Resulting graph of precarity sum score and individual depression symptoms using FCI and proportion of edge endpoint types.

nuanced pathways that may be overlooked when relying solely on aggregate measures, while also capturing broader interactions that might remain undetected in analyses focused on individual factors.

Our findings highlight the significant role of recent financial stress (S.fin) as a potential causal factor for depression, consistently observed across both sum score and symptom-level analyses. This consistency underscores the profound impact of financial stressors on mental health. The symptom-level analysis further identified specific symptoms—particularly sleep disturbances (slp), depressed mood (dep), guilt (glt), and anhedonia (anh)—as being especially sensitive to external precarious conditions.

Among these symptoms, dep stands out as a distinct causal factor for precarity in both aggregated and disaggregated analyses, with CCI even suggesting a possible cyclic relationship between them. This implies that depression and precariousness may reinforce one another, potentially leading to a worsening cycle over time. Within the symptom network, dep also exhibits a high degree of connectivity, but with mostly incoming arrows, suggesting that as symptom dynamics intensify, dep is quickly impacted by multiple factors—particularly P.soc (social precarity), further fueling the depression-precariousness cycle.

Beyond dep, slp emerges as a critical symptom in this dynamic. It connects with both social and employment precariousness, while also playing a central role in symptom interactions, affecting other symptoms more than most others. This suggests that slp may act as an initiator within the depressive symptom system, potentially triggering or amplifying symptom cascades.

Among the precariousness factors, P.soc appears to be one of the most pivotal—not as a driver of symptoms, but rather as a consequence of depression. It maintains the strongest connectivity with symptoms, yet its directionality indicates that it is largely shaped by depressive symptoms rather than acting as a primary cause. Together, these patterns suggest a broader directional dynamic in which financial stress (S.fin) exacerbates depressive symptoms, which in turn feed into social precarity, perpetuating the cycle.

This directional insight suggests that interventions targeting specific depressive symptoms, particularly dep, may help prevent downstream effects on worsening precarious conditions. Alternatively, early intervention focusing on slp may be a promising strategy, as slp appears to act as a potential initiator or activator within the depressive symptom network. This positioning suggests that slp could serve as an early warning sign, offering a valuable target for preventive interventions aimed at breaking the depression-precariousness cycle before it intensifies.

Several limitations should be acknowledged when interpreting these findings. The prevalence of bidirectional arrows and circle-marked endpoints in the graphs reflects unresolved ambiguities in the causal relationships suggested by the data. These uncertainties highlight the need for further research, particularly with higher-resolution datasets, such as time-series data. Incorporating

time-series data could capture temporal dynamics between precarity factors and depressive symptoms, providing time-specific insights and revealing how these relationships evolve over time. Methods like *PCMCI* (Runge et al., 2019), *tsFCI* (Entner & Hoyer, 2010), and other time-series adaptations of causal discovery algorithms could offer a more robust framework for addressing temporal dependencies and refining the analysis.

Another notable limitation is the absence of clear evidence for cycles within the symptom network, despite employing algorithms designed to account for cyclic relationships. The CCI algorithm primarily produced bidirectional arrows, while FCI predominantly generated directional arrows, yet neither demonstrated patterns indicative of definitive cyclic structures. This suggests that addressing cyclic relationships is particularly challenging when relying solely on observational datasets. Future research could benefit from refined datasets that incorporate intervention data. Methods such as *LLC* (Hyttinen et al., 2012), *NODGAS-Flow* (Sethuraman et al., 2023), and the recently developed *Bicycle* algorithm (Rohbeck et al., 2024), which utilize both observational and intervention data, may uncover potential cycles more effectively.

Lastly, regarding conditional independence (CI) testing, the differences between the graphs generated by Gaussian CI and RCoT underscore the methodological sensitivities in detecting causal relationships. Gaussian CI produced denser graphs, which is somewhat counterintuitive, as the Gaussian CI's strict linearity assumption would typically result in fewer detected relationships, not more. A possible explanation for this discrepancy is that Gaussian CI's reliance on partial correlations may overestimate relationships when specific non-linear dependencies exist in the data. In contrast, RCoT, free from such assumptions, might better capture independence patterns under such conditions. However, RCoT is not without its limitations. While technically non-parametric, its performance can be influenced by the distributional characteristics of the data. The Gaussian RBF (radial basis function) kernel, optimized for continuous data with smooth transitions, may struggle with discrete or mixed datasets, where distances between discrete points may fail to convey meaningful similarity (Howlett, 2001). This limitation could explain why RCoT might miss certain dependencies, particularly in the HELIUS dataset, where some variables lack smooth continuity. Future research could tackle these challenges by developing hybrid kernels designed for mixed datasets, integrating both continuous and discrete variables into RCoT-like methods. This would have the potential to improve the reliability of findings, especially when working with datasets characterized by complex and heterogeneous structures.

Despite its limitations, this study marks a meaningful step toward understanding the mechanisms linking precarity factors and depression. By applying causal discovery methods, it moves beyond traditional association-based analyses, providing insights that can inform more precise and targeted interventions. While the resulting graphs are preliminary and contain unresolved ambiguities, they offer a valuable starting point for leveraging causal discovery tools to

investigate the causal interplay between depression and precarity factors. A promising next step would involve integrating these causal structures into computational models, such as the symptom dynamic model proposed in Park et al. (2025). By simulating intervention effects, such models could provide more realistic insights into how targeted actions might influence symptom networks and precarity factors over time. For example, interventions focused on improving sleep hygiene or alleviating guilt could be evaluated for their cascading effects on employment and social relationships, offering actionable guidance for designing population-level mental health strategies. As one of the early applications of causal discovery tools to the complex dynamics of depression and precarity factors, this study lays a foundation for future research. We hope it inspires further refinement of these methods and ultimately contribute to more effective solutions for alleviating depression and improving societal well-being.

5 References

- Dojer, N. (2016). Learning bayesian networks from datasets joining continuous and discrete variables. *International Journal of Approximate Reasoning*, 78, 116–124.
- Entner, D., & Hoyer, P. O. (2010). On causal discovery from time series data using FCI. *Probabilistic Graphical Models*, 16.
- Forré, P., & Mooij, J. M. (2018). Constraint-based causal discovery for non-linear structural causal models with cycles and latent confounders. *arXiv Preprint arXiv:1807.03024*.
- Howlett, R. J. (2001). *Radial basis function networks 1: Recent developments in theory and applications*.
- Hyttilinen, A., Eberhardt, F., & Hoyer, P. O. (2012). Learning linear cyclic causal models with latent variables. *The Journal of Machine Learning Research*, 13(1), 3387–3439.
- Lindsay, B. G., Pilla, R. S., & Basak, P. (2000). Moment-based approximations of distributions using mixtures: Theory and applications. *Annals of the Institute of Statistical Mathematics*, 52, 215–230.
- Mooij, J. M., & Claassen, T. (2020). Constraint-based causal discovery using partial ancestral graphs in the presence of cycles. *Conference on Uncertainty in Artificial Intelligence*, 1159–1168.
- Neapolitan, R. E. et al. (2004). *Learning bayesian networks* (Vol. 38). Pearson Prentice Hall Upper Saddle River.
- Park, K., Waldorp, L. J., & Ryan, O. (2024). Discovering cyclic causal models in psychological research. *Advances in Psychology*, 2, e72425.
- Park, K., Waldorp, L., & Vasconcelos, V. V. (2025). *The individual-and population-level mechanistic implications of statistical networks of symptoms*.
- Rahimi, A., & Recht, B. (2007). Random features for large-scale kernel machines. *Advances in Neural Information Processing Systems*, 20.
- Rohbeck, M., Clarke, B., Mikulik, K., Pettet, A., Stegle, O., & Ueltzhöffer, K. (2024). Bicycle: Intervention-based causal discovery with cycles. In F. Lo-

- catello & V. Didelez (Eds.), *Proceedings of the third conference on causal learning and reasoning* (Vol. 236, pp. 209–242). PMLR.
- Runge, J., Nowack, P., Kretschmer, M., Flaxman, S., & Sejdinovic, D. (2019). Detecting and quantifying causal associations in large nonlinear time series datasets. *Science Advances*, 5(11), eaau4996.
- Sethuraman, M. G., Lopez, R., Mohan, R., Fekri, F., Biancalani, T., & Hüttler, J.-C. (2023). NODAGS-flow: Nonlinear cyclic causal structure learning. *International Conference on Artificial Intelligence and Statistics*, 6371–6387.
- Snijder, M. B., Galenkamp, H., Prins, M., Derkx, E. M., Peters, R. J., Zwinterman, A. H., & Stronks, K. (2017). Cohort profile: The healthy life in an urban setting (HELIUS) study in amsterdam, the netherlands. *BMJ Open*, 7(12), e017873.
- Spirites, P., Glymour, C., & Scheines, R. (2001). *Causation, prediction, and search*. MIT press.
- Spirites, P., Meek, C., & Richardson, T. (1995). Causal inference in the presence of latent variables and selection bias. *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, 499–506.
- Strobl, E. V. (2019). A constraint-based algorithm for causal discovery with cycles, latent variables and selection bias. *International Journal of Data Science and Analytics*, 8(1), 33–56. <https://doi.org/10.1007/s41060-018-0158-2>
- Strobl, E. V., Zhang, K., & Visweswaran, S. (2019). Approximate kernel-based conditional independence tests for fast non-parametric causal discovery. *Journal of Causal Inference*, 7(1), 20180017.
- Zhang, K., Peters, J., Janzing, D., & Schölkopf, B. (2012). Kernel-based conditional independence test and application in causal discovery. *arXiv Preprint arXiv:1202.3775*.

6 Appendix

6.1 Precariousness factors by Leonie

1. EMPLOYMENT PRECARIOUSNESS
 - H1_Arbeidsparticipatie: Working status
 - H1_WerkSit: Which work situation most applies to you?
 - H1_RecentErv8: Experiences past 12 months: h. You were sacked from your job or became unemployed (*reverse*)
2. FINANCIAL PRECARIOUSNESS
 - H1_InkHhMoeite: During the past year, did you have problems managing your household income?
 - H1_RecentErv9: Experiences past 12 months: i. You had a major financial crisis (*reverse*)
3. HOUSING PRECARIOUSNESS
 - veilig_2012: Score safety (veiligheid) in 2012 (*reverse*)

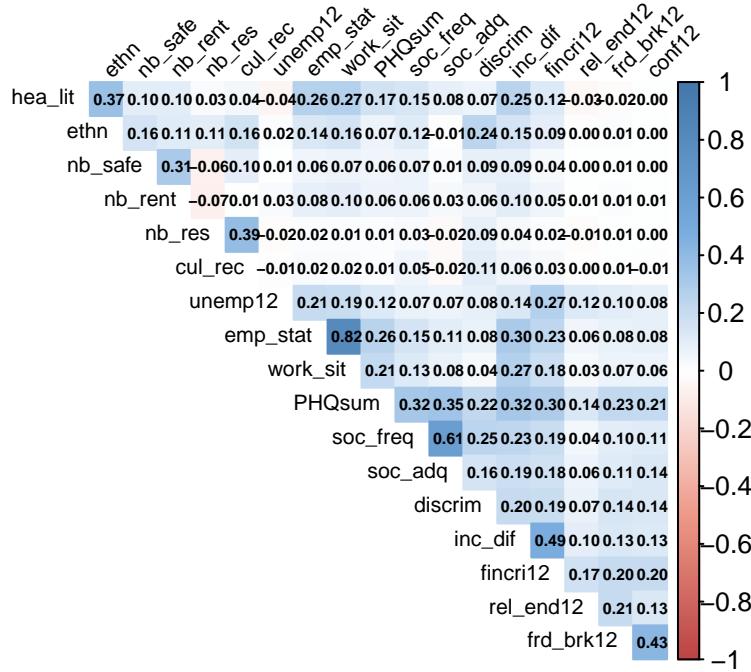
- vrz_2012: Score level of resources (niveau voorzieningen) in 2012 (*reverse*)
- P_HUURWON: Percentage Huurwoningen

4. CULTURAL PRECARIOUSNESS

- H1_Discr_sumscore: Perceived discrimination: sum score of 9 items (range 9-45)
- H1_SBSQ_meanscore: Health literacy: SBSQ meanscore (range 1-5) (*reverse*)
- A_BED_RU: Aantal bedrijfsvestigingen; cultuur, recreatie, overige diensten (*reverse*)

5. SOCIAL PRECARIOUSNESS

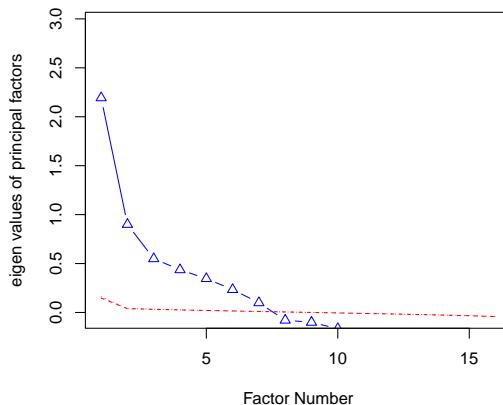
- H1_RecentErv5: Experiences past 12 months: e. Your steady relationship ended (*reverse*)
- H1_RecentErv6: Experiences past 12 months: f. A long-term friendship with a good friend or family member was broken off (*reverse*)
- H1_RecentErv7: Experiences past 12 months: g. You had a serious problem with a good friend or family member, or neighbour (*reverse*)
- H1_SSQT: SSQT (frequency of social contact): sum score of 5 items (range 5-20) (*reverse*)
- H1_SSQSa: SSQS (adequacy of social contact): sum score of 5 items, category 3 and 4 not combined (range 5-20) (*reverse*)

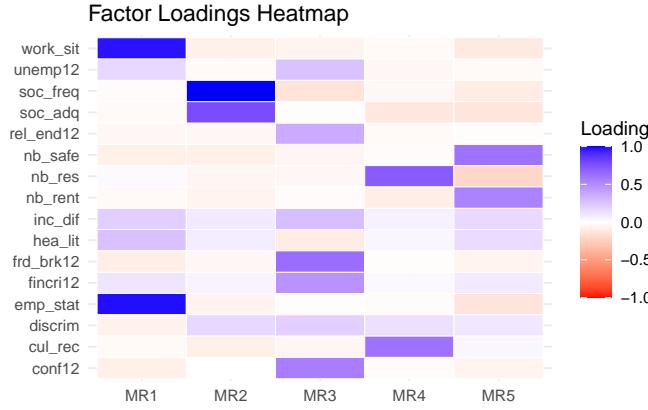


- **High** Correlations: `emp_stat` (employment status) and `work_sit` (work situation) have a strong positive correlation of 0.82. This suggests that individuals with higher employment status tend to have more secure or favorable work situations. `soc_freq` (social contact frequency) shows a strong positive correlation with `soc_adq` (social adequacy) at 0.61. This indicates that individuals with more frequent social contact also tend to have higher perceived adequacy of social interactions.
- **Moderate** Correlations: `nb_safe` (neighborhood safety) and `nb_res` (resources) have a moderate positive correlation of 0.39, suggesting that areas with higher safety also have better resources. `hea_lit` (health literacy) has moderate correlations with `emp_stat` (0.26) and `work_sit` (0.25), which could mean that higher health literacy is associated with better employment situations. `frd_brk12` (friendship breakups) and `conf12` (conflicts) have a notable correlation of 0.43, indicating a relationship between having conflicts and friendship losses.
- **Low to Moderate** Correlations in Financial Precariousness: `inc_dif` (income difficulties) has a moderate correlation with `fincris12` (financial crisis) at 0.49. This aligns with the expected relationship, where individuals who experience general income difficulties are more likely to report financial crises.
- **Low** Correlations (0.1 - 0.2): Many variables, such as `discrim` (discrimination), `unemp12` (unemployment experience), and `rel_end12` (relationship end), have low correlations with other variables, suggesting relatively independent relationships in the context of this dataset.

6.2 Exploratory Factor Analysis (EFA)

Parallel Analysis Scree Plots





6.2.1 Factor Loadings (Pattern Matrix)

- **MR1:** High loadings on `emp_stat` and `work_sit` suggest this factor captures *employment precariousness*.
- **MR2:** Strong loadings on `soc_freq` and `soc_adq` indicate *social precariousness*.
- **MR3:** Key items like `frd_brk12`, `conf12`, and `fincril2`, suggest recent *stressful events*.
- **MR4:** High loadings on `nb_res` and `cul_rec` may reflect *community resources precariousness*.
- **MR5:** Variables `nb_safe` and `nb_rent` with high loadings indicate *housing precariousness*.

6.2.2 Variance Explained

The factors cumulatively explain 38% of the variance, with MR1 being the most influential factor. Each factor contributes a smaller proportion to the total variance (MR1 at 12%, MR2 at 9%, etc.).

6.2.3 Factor Intercorrelations

Factors are moderately correlated, especially between *MR1 and MR5*, and *MR2 and MR3*. This indicates that while distinct, these factors are related—reasonable in a complex socio-economic context.

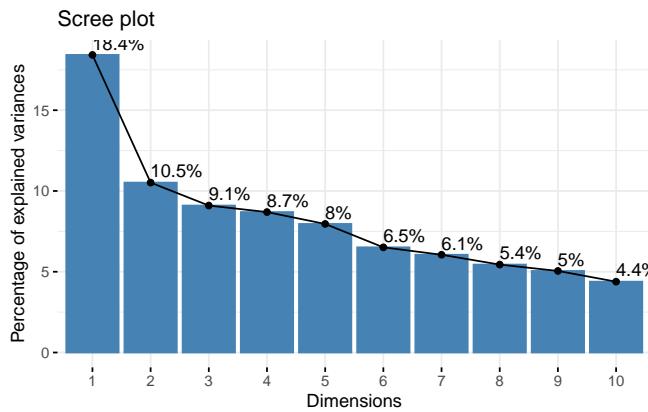
6.2.4 Model Fit Statistics

RMSEA (0.071) suggest an acceptable fit. Tucker Lewis Index (0.802) suggests moderate reliability for the model.

6.2.5 Summary

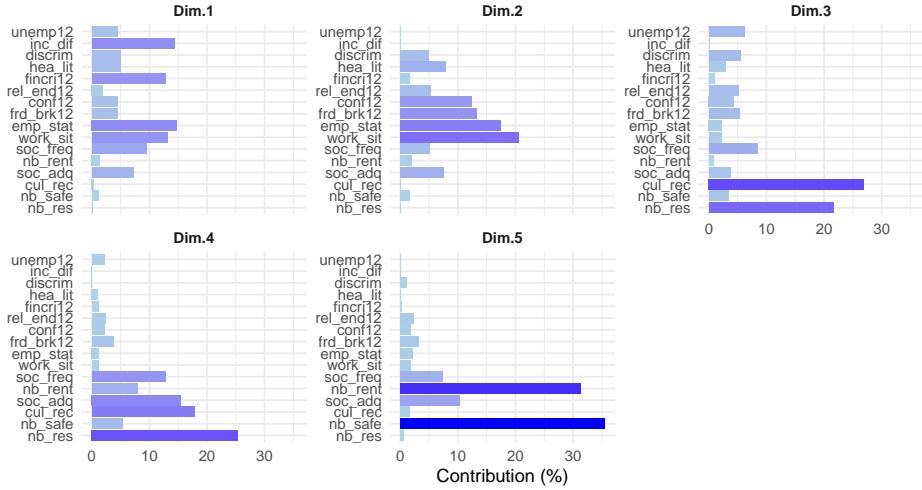
The 5-factor model appears interpretable and captures distinct dimensions of precariousness: *employment, social, stressors, community resources, and housing precariousness*. Although the overall fit and explained variance could be stronger, these factors offer insights into the underlying structure of the data, highlighting key areas of precariousness.

6.3 PCA



- Component Retention: The scree plot shows a clear “elbow” after the first component. This steep drop suggests that most variance is explained by the first component. After Dimension 5, the percentage of explained variance decreases slightly more gradually, indicating diminishing returns for adding more components. If we need to choose multiple components, retaining the first 5 components seems reasonable, as they capture most of the variance (cumulatively explaining about 54.7% of the total variance).

Contribution of Variables to 5 Principal Components



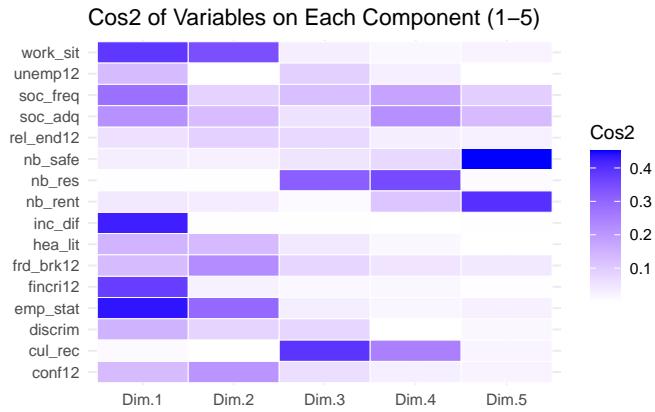
6.3.1 Explained variance (contributions) of variables

It shows the importance of variables within each component.

- **Dim1:** High contributions are observed from `emp_stat`, `work_sit`, `inc_dif`, and `fincri12`, suggesting that this dimension captures aspects of *employment and financial security*.
- **Dim2:** While `emp_stat` and `work_sit` overlap with Dim1, the strong contributions from `frd_brk12` and `rel_end12` indicate that this dimension captures a focus on *recent relationship stressors*.
- **Dim3:** `cul_rec`, `nb_res` have the highest contributions, indicating this dimension likely represents *community and cultural factors*.
- **Dim4:** `soc_freq` and `soc_adq` stand out in this dimension, suggesting an emphasis on *social precariousness*.
- **Dim5:** `nb_safe` and `nb_rent` are the top contributors, pointing to *housing security* as key themes in this component.

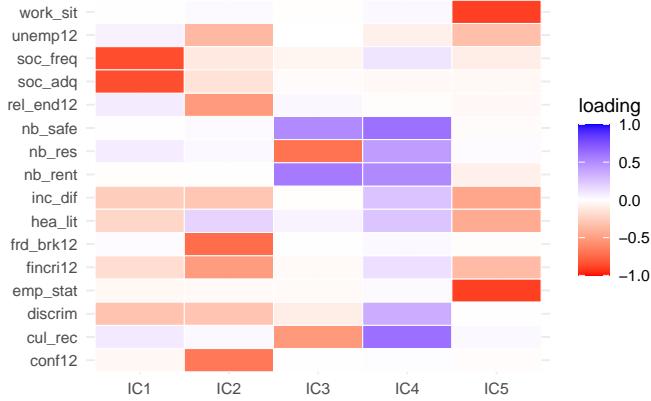
6.3.2 Cos² Values

Cos² (squared cosine) values, or the quality of representation, show how well each variable is represented by each dimension. where higher cos² values (closer to 1) indicate better representation of a variable by a component.



- **Dim.1:** Variables `emp_stat`, `work_sit`, `inc_dif`, and `fincri12` show high \cos^2 values, meaning that PC1 primarily captures variations in employment and financial difficulties. This component could represent *employment & finance* precariousness.
- **Dim.2:** Variables `work_sit`, `emp_stat`, `frd_brk12`, and `conf12` are well-represented in this component, suggesting PC2 captures aspects of *recent relationship stressors*.
- **Dim.3:** Variables `nb_res` and `cul_rec` load strongly on PC3. This may represent community or cultural resources, indicating that this component is associated with *neighborhood resources*.
- **Dim.4:** This component has high \cos^2 values for `nb_res`, `cul_rec`, `soc_freq`, and `soc_adq`. While `nb_res` and `cul_rec` are also prominent in PC3, PC4 uniquely captures nuanced differentiation in *social* precariousness.
- **Dim.5:** `nb_safe` and `nb_rent` are well-represented by PC5. This component might capture *housing* precariousness.

6.4 ICA



6.4.1 Dominant Variables per Component:

For each Independent Component (IC), we can identify variables with *high absolute* values in each column. These values indicate that the IC captures a strong, independent signal associated with these variables.

- **IC1:** soc_freq and soc_adq have strong negative loadings on this component, indicating that this component might represent *social precariousness*.
- **IC2:** frd_brk12, conf12, rel_end12, fincri12 and unemp12 have the most substantial loadings on this component, all with negative signs. This might point to a *recent relational or social stressor* component.
- **IC3:** nb_res and cul_rec show notable negative loadings, pointing to a focus on *community resource precariousness*.
- **IC4:** High loadings for nb_safe, nb_rent, nb_res, cul_rec, and discrim suggest a theme of *housing and community-based precariousness*, reflecting both safety and social challenges within the neighborhood context.
- **IC5:** emp_stat and work_sit both have strong negative loadings on this component, suggesting it captures *employment precariousness*.

6.5 Hierarchical clustering

6.5.1 Using Euclidean distance

- Ward.D's method: Minimizes the variance within clusters, producing more compact and spherical clusters.
- Single linkage: Groups clusters based on the minimum distance between points.

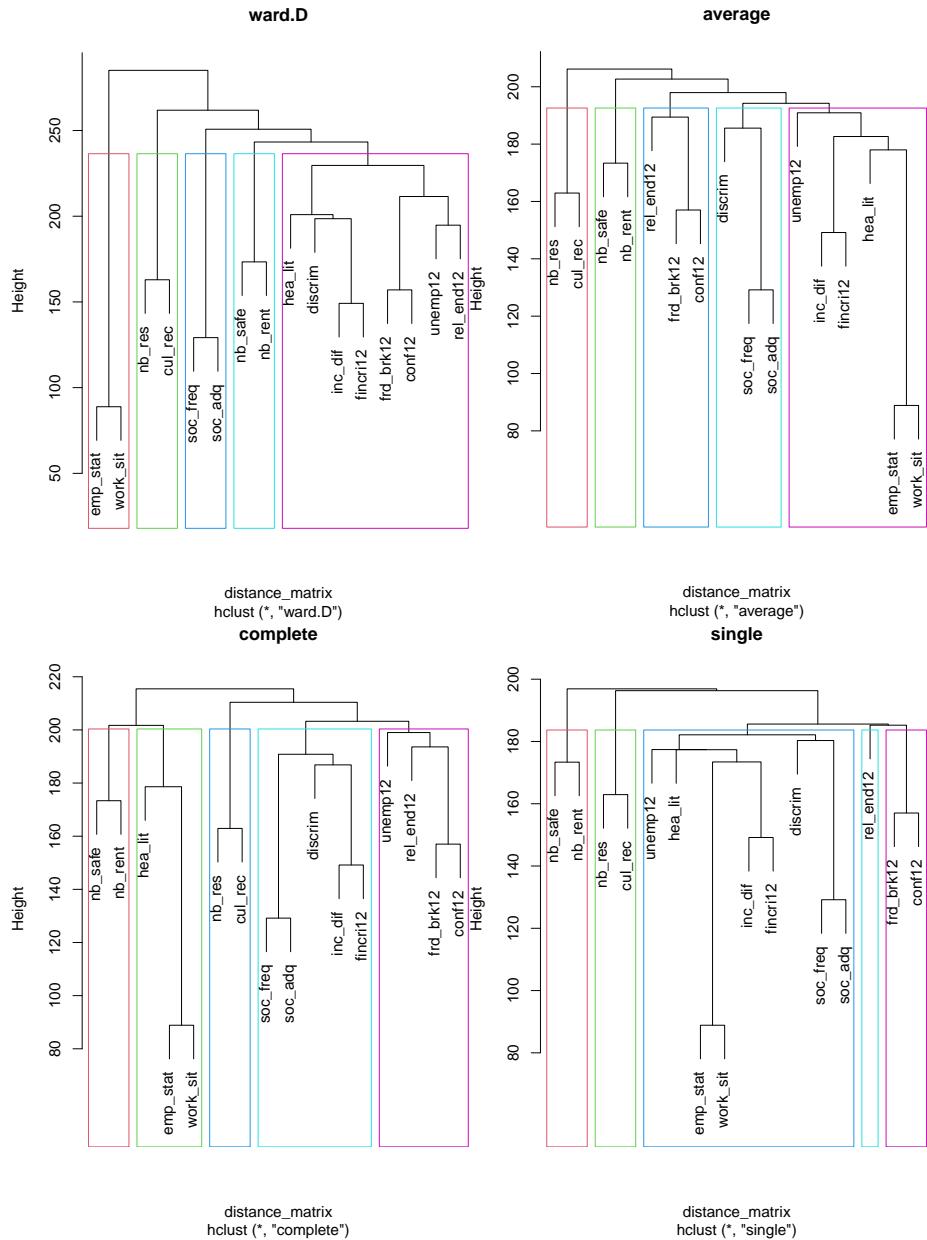
- Complete linkage: Groups clusters based on the maximum distance between points.
- Average linkage: Uses the average distance between all pairs of points in the two clusters.

6.5.1.1 Consistent Groupings (Across All or Most Methods)

- `emp_stat` and `work_sit`: This pair consistently clusters together across all linkage methods, suggesting that they are closely related variables, likely capturing a similar aspect of the data (possibly employment status or employment-related information).
- `nb_safe`, `nb_res`, and `nb_rent`: These variables are often grouped closely in several methods (especially Ward.D, average, and complete linkage). This suggests a similarity or common theme among them, potentially related to neighborhood or housing precariousness.
- `soc_freq` and `soc_adq`: These two variables frequently cluster together, indicating they likely measure aspects of social frequency and adequacy in similar ways. They appear together in Ward.D, average, and complete linkage.
- `frd_brk12` and `conf12`: These variables are often clustered closely (though they sometimes join with other variables like `rel_end12`), suggesting they may capture aspects of relationship or social conflict. This pair appears in close proximity, especially in average and Ward.D.

6.5.1.2 Inconsistent Groupings (Variability Across Methods)

- `hea_lit`: This variable shows inconsistent clustering across methods. In Ward.D, it joins with `fincril12`, while in other methods, it's often more isolated or grouped with variables that do not appear similar. This may suggest that `hea_lit` does not strongly correlate with other variables, or it has multidimensional aspects affecting its grouping across methods.
- `discrim`: This variable also shows variable groupings. In Ward.D, it is grouped with `hea_lit`, while in other methods (e.g., complete and single linkage), it clusters differently, sometimes on its own. This variability may indicate that `discrim` has weaker associations with the main clusters in the data or overlaps partially with multiple clusters.
- Social and Financial Variables (`inc_dif`, `fincril12`, `unemp12`): These variables appear together in some methods (e.g., Ward.D clusters `fincril12` and `inc_dif`), but in others, they are spread out. This inconsistency suggests that social and financial variables may not have strong or consistent ties across different methods, perhaps due to capturing different aspects of precariousness.

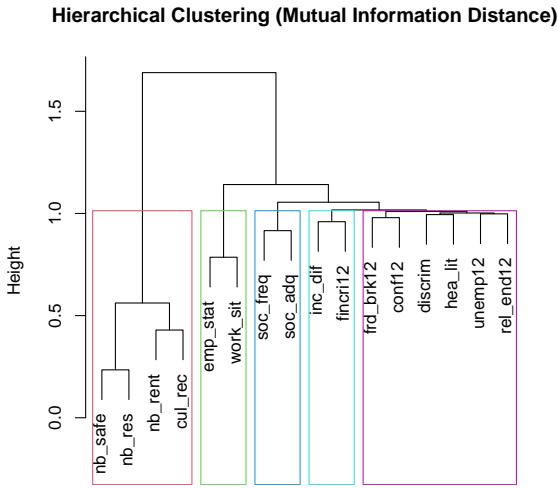


6.5.1.3 Summary

The consistent clusters are likely capturing distinct thematic dimensions of the data (e.g., employment, housing, social contact), while the inconsistent variables may reflect multifaceted or weakly correlated attributes that do not fit neatly into one cluster.

6.5.2 Using Mutual Information

Using mutual information (MI) as a basis for hierarchical clustering differs from using traditional distance measures (like Euclidean distance) in a few key ways.



6.5.2.1 Comparison to Euclidean Distance Clustering

- **Housing and Community Cluster:** The variables `nb_safe`, `nb_res`, `nb_rent`, and `cul_rec` cluster together, indicating a strong association among housing-related and community-based factors. This suggests a shared theme of housing or community precariousness. This grouping is also observed in the Euclidean-based clustering, but it appears more tightly connected here, potentially due to the non-linear relationships highlighted by mutual information.
- **Employment and Social Support Cluster:** `emp_stat` and `work_sit` form a cluster, linking employment status and work situation together as they did in Euclidean-based clustering. These remain closely associated regardless of the distance metric used. `soc_freq` and `soc_adq`, related to social contact frequency and adequacy, cluster nearby, indicating they have a stronger non-linear relationship with employment variables. This is a subtle difference as Euclidean distance might not capture this association as effectively.

- **Financial Stressor** Cluster: `inc_dif` and `fincril2`, representing income difficulties and recent financial crises, consistently cluster together in both approaches, showing a strong association, likely linear. However, mutual information-based clustering links these financial stressors with social support variables, suggesting that financial challenges may have complex dependencies with social support in this dataset.
- **Relational Stressor** Cluster: `frd_brk12`, `conf12`, `discrim`, `hea_lit`, `unemp12`, and `rel_end12` form a *looser* cluster focused on social and relational stressors (e.g., friendship breakup, conflicts, and discrimination). Compared to Euclidean clustering, `discrim` and `hea_lit` (health literacy) appear closer to relational stressors here, indicating that non-linear relationships might play a larger role in linking these variables.

6.5.2.2 Summary

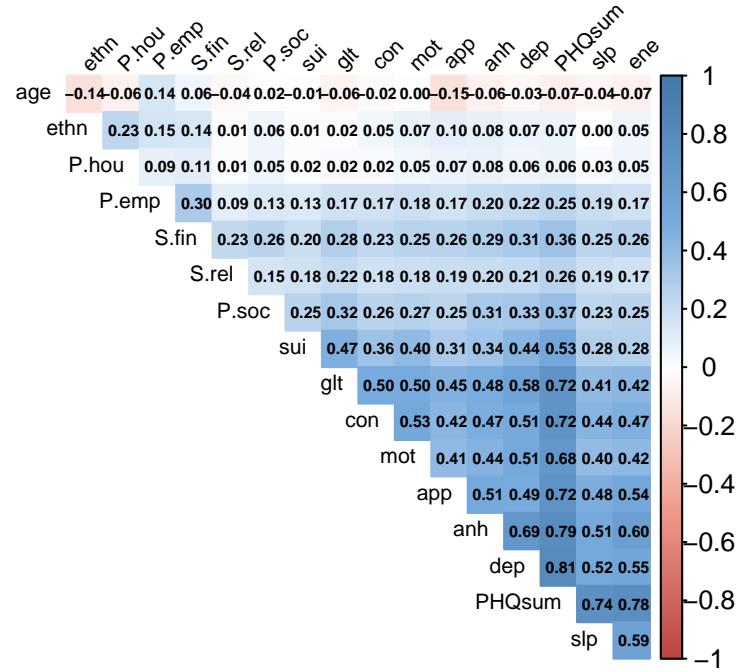
In conclusion, mutual information-based clustering provides an alternative perspective that can reveal more intricate associations between variables, especially for those with non-linear relationships. Compared to Euclidean clustering, it shows a similar high-level structure but emphasizes nuanced connections between variables, particularly around social support, employment, and financial stress.

6.6 Conclusions on Precariousness factors

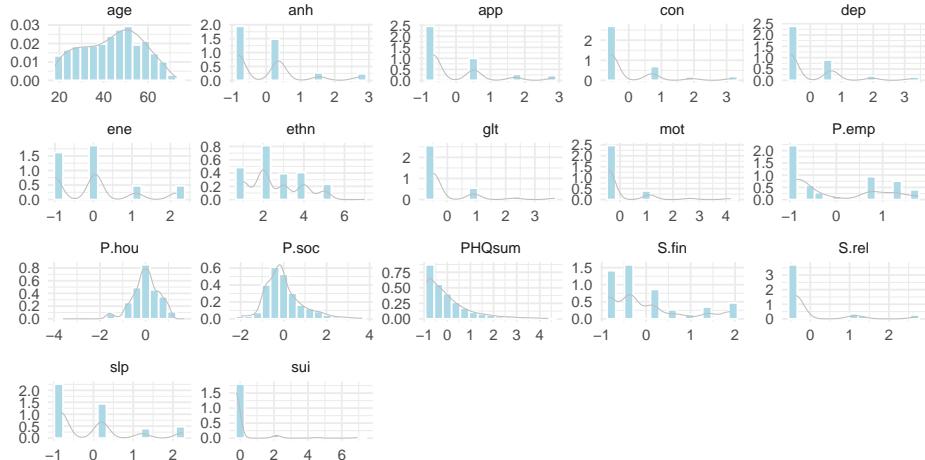
Based on the consistent findings across multiple analyses, we decided to exclude the variables `discrim`, `hea_lit`, `umemp12`, and `rel_end12`, as they do not clearly belong to any specific precariousness factor nor exhibit strong associations with depression (see the correlation table above). Therefore, we propose retaining the following key precariousness factors:

- Employment Precariousness: `emp_stat`, `work_sit`
- Social Precariousness: `soc_freq`, `soc_adq`
- Housing Precariousness: `nb_safe`, `nb_res`, `nb_rent`, `cul_rec`
- Recent Relational Stressors: `frd_brk12`, `conf12`
- Recent Financial Stressors: `fincril2`, `inc_diff`

We construct each precariousness factor by calculating the mean value of the combined variables. Below, we present the updated correlation table for the newly composed factors, along with the corresponding distributions of all variables to be used in the causal discovery analysis.



Distribution of All Variables with Density Overlay



6.7 Results from CCI algorithm

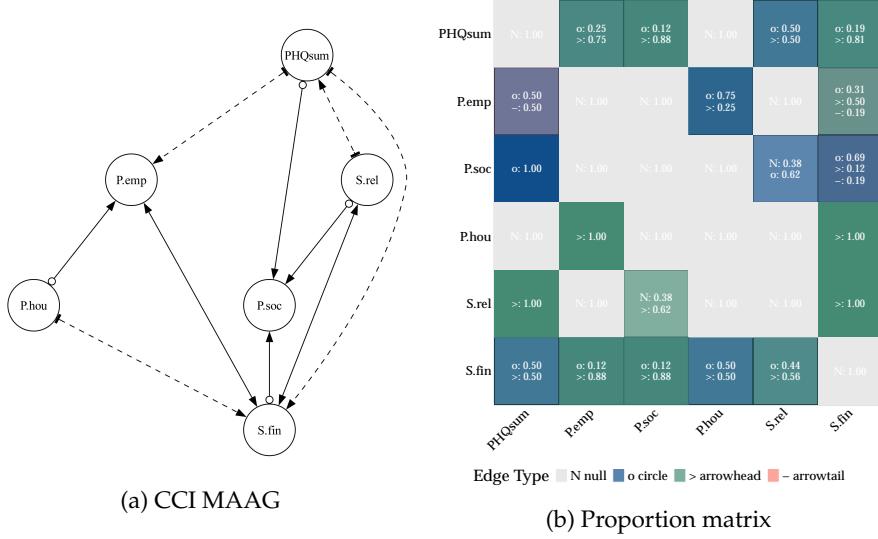


Figure 7: Resulting graph of precarity factors and depression sum score using CCI and proportion of edge endpoint types.

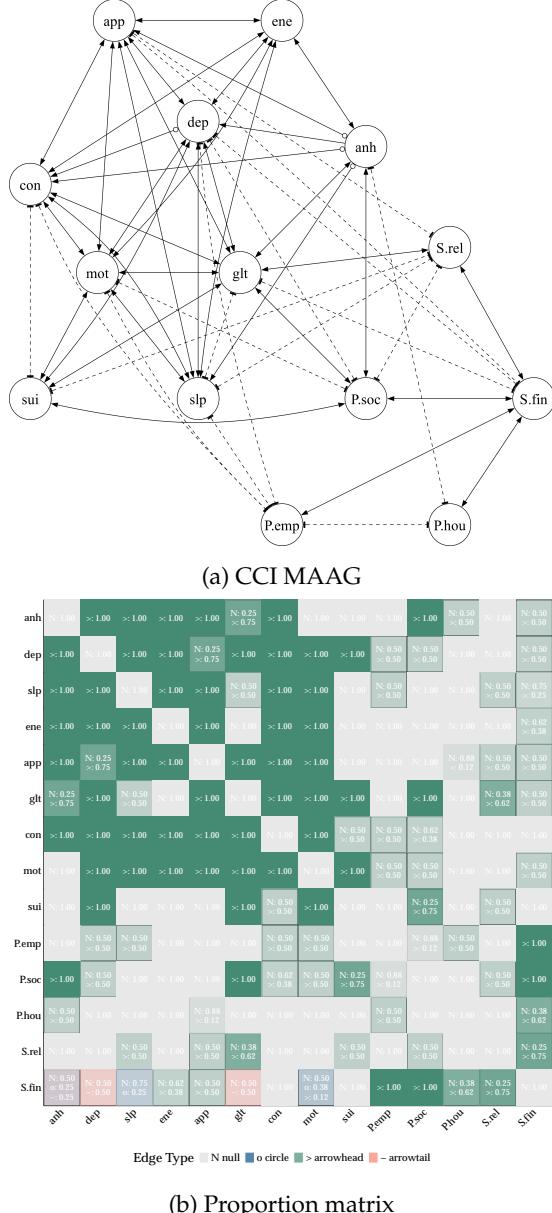
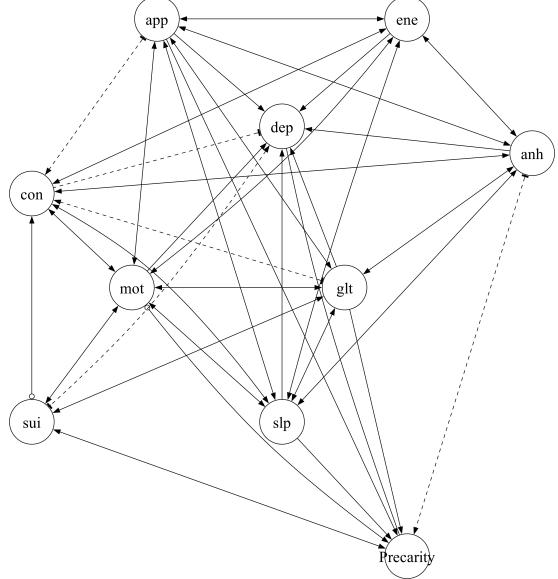


Figure 8: Resulting graph of precarity factors and individual depression symptoms using CCI and proportion of edge endpoint types.



(a) CCI MAAG

	anh	> 1.00	> 1.00	> 1.00	> 1.00	N: 0.12 > 0.88	> 1.00	N: 1.00	> 1.00
dep	$\alpha: 0.12-: 0.88$	N: 1.00	-> 1.00	$\alpha: 0.25-: 0.12-: 0.62$	-> 1.00	-> 1.00	-> 1.00	-> 1.00	$\alpha: 0.50-: 0.50$
slp	> 1.00	> 1.00	N: 1.00	> 1.00	> 1.00	N: 0.25 > 0.75	> 1.00	> 1.00	N: 1.00 > 1.00
ene	> 1.00	> 1.00	> 1.00	N: 1.00	> 1.00	N: 1.00	> 1.00	> 1.00	N: 1.00 > 1.00
app	> 1.00	> 1.00	> 1.00	> 1.00	N: 1.00	> 1.00	> 1.00	> 1.00	N: 1.00 > 1.00
glt	N: 0.12 > 0.88	> 1.00	N: 0.25 > 0.75	N: 1.00	> 1.00	N: 1.00	> 1.00	> 1.00	> 1.00
con	$\alpha: 0.12-: 0.50-: 0.38$	> 0.50 -: 0.30	$\alpha: 0.25-: 0.50-: 0.25$	$\alpha: 0.25-: 0.75$	> 0.50 -: 0.50	> 0.50 -: 0.50	N: 1.00	$\alpha: 0.25-: 0.50-: 0.25$	N: 0.25 -: 0.62 -: 0.12
mot	N: 1.00	> 1.00	> 1.00	> 1.00	> 1.00	> 1.00	> 1.00	N: 1.00	> 1.00 > 1.00
sui	N: 1.00	> 1.00	N: 1.00	N: 1.00	N: 1.00	> 1.00	N: 0.25 > 0.75	> 1.00	N: 1.00 > 1.00
precarity	$\alpha: 0.50-: 0.50$	$\alpha: 0.25-: 0.12-: 0.62$	$\alpha: 0.12-: 0.38-: 0.50$	N: 1.00	$\alpha: 0.25-: 0.25-: 0.50$	$\alpha: 0.25-: 0.12-: 0.62$	$\alpha: 0.50-: 0.12-: 0.58$	$\alpha: 0.75-: 0.25$	$\alpha: 0.38-: 0.62-: 1.00$

Edge Type ■ N null ■ o circle ■ > arrowhead ■ - arrowtail

(b) Proportion matrix

Figure 9: Resulting graph of precarity sum score and individual depression symptoms using CCI and proportion of edge endpoint types.

6.8 Results from PC algorithm

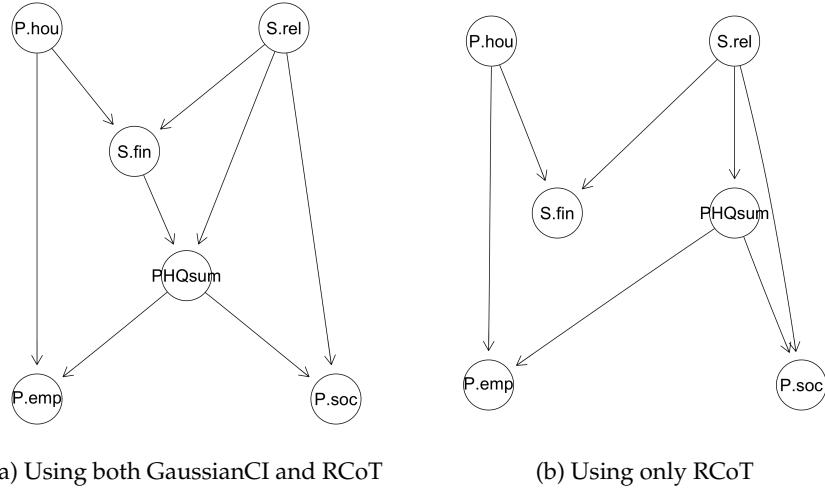


Figure 10: Resulting graphs of precarious factors and depression sum score using PC.

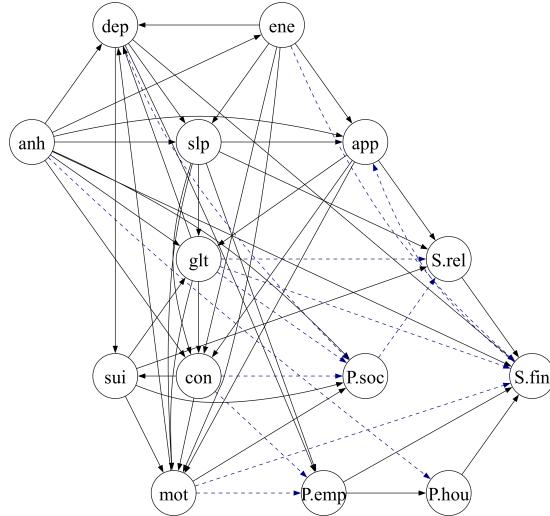


Figure 11: Resulting graphs of precarity factors and individual depression symptoms using PC.

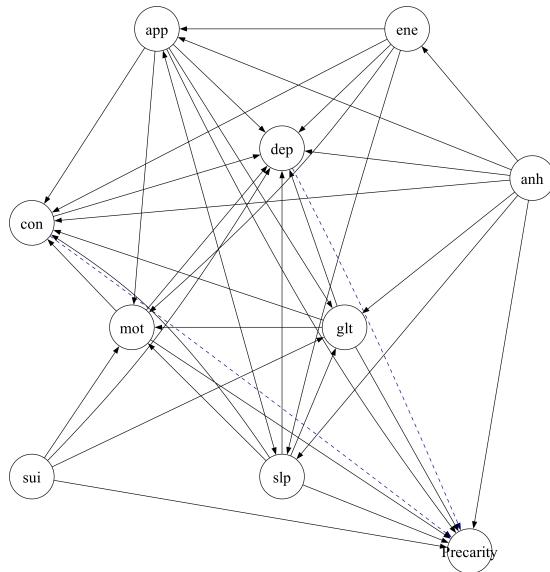


Figure 12: Resulting graphs of precarity sum score and individual depression symptoms using PC.

6.9 Randomized Conditional Independence / Correlation Test (RCIT & RCoT)

RCIT (Randomized Conditional Independence Test) and RCoT (Randomized conditional Correlation Test) are advanced methods for scalable conditional independence (CI) testing, offering computational efficiency while maintaining the accuracy of kernel-based approaches. These methods evaluate conditional independence between two variables X and Y given a third variable Z while addressing computational challenges inherent in kernel-based CI tests. In this section, we provide a high-level overview of RCIT and RCoT based on (Strobl et al., 2019).

6.9.1 Kernel-Based Conditional Independence Testing

Traditional kernel-based CI tests, such as the Kernel Conditional Independence Test (KCIT), compute dependencies using the Hilbert-Schmidt Independence Criterion (HSIC) in reproducing kernel Hilbert spaces (RKHS) (Zhang et al., 2012). KCIT uses the following hypothesis framework:

$$H_0 : X \perp\!\!\!\perp Y | Z, \quad H_1 : X \not\perp\!\!\!\perp Y | Z.$$

The core quantity in KCIT is the partial cross-covariance operator:

$$\Sigma_{XY \cdot Z} = \Sigma_{XY} - \Sigma_{XZ} \Sigma_{ZZ}^{-1} \Sigma_{ZY},$$

where Σ_{XY} represents the cross-covariance operator between X and Y , and $\Sigma_{XZ} \Sigma_{ZZ}^{-1} \Sigma_{ZY}$ removes the dependence mediated by Z .

The squared Hilbert-Schmidt (HS) norm of $\Sigma_{XY \cdot Z}$ serves as the test statistic:

$$\|\Sigma_{XY \cdot Z}\|_{HS}^2 = 0 \quad \text{if and only if} \quad X \perp\!\!\!\perp Y | Z.$$

KCIT estimates residual dependencies using kernel ridge regression:

$$f^*(z) = K_Z(K_Z + \lambda I)^{-1} f(x),$$

where K_Z is the kernel matrix for Z , $f(x)$ is the kernel feature map for X , and λ is the ridge regularization parameter. The residual function for X is:

$$f_{\text{res}}(x) = f(x) - f^*(z) = R_Z f(x),$$

with:

$$R_Z = I - K_Z(K_Z + \lambda I)^{-1}.$$

The kernel matrix for residualized X is:

$$K_{X \cdot Z} = R_Z K_X R_Z,$$

and similarly for Y , $K_{Y \cdot Z} = R_Z K_Y R_Z$.

The test statistic is computed as:

$$T_{XY \cdot Z} = \frac{1}{n^2} \text{tr}(K_{X \cdot Z} K_{Y \cdot Z}),$$

which estimates the Hilbert-Schmidt (HS) norm of the partial cross-covariance operator. To ensure convergence, KCIT scales the statistic by n :

$$S_K = n T_{XY \cdot Z}.$$

The null hypothesis H_0 is rejected if S_K exceeds a threshold determined by permutation or moment-matching-based null distribution (Lindsay et al., 2000).

6.9.2 Random Fourier Features (RFFs)

Kernel-based methods like KCIT face scalability issues, as they involve operations on $n \times n$ kernel matrices, which scale quadratically with the sample size n . RCIT and RCoT overcome this bottleneck using *Random Fourier Features (RFFs)* to approximate kernel operations efficiently.

6.9.2.1 Bochner's Theorem

Bochner's theorem provides the foundation for RFFs, stating that any continuous shift-invariant kernel $k(x, y)$ can be expressed as:

$$k(x, y) = \int_{\mathbb{R}^p} e^{i\omega^\top (x-y)} dP_\omega,$$

where P_ω is the spectral distribution of the kernel. For the widely used RBF kernel:

$$k(x, y) = \exp\left(-\frac{\|x - y\|^2}{2\sigma^2}\right),$$

P_ω follows a Gaussian distribution: $\omega \sim \mathcal{N}(0, \sigma^2 I)$.

6.9.2.2 RFF Approximation

Using Monte Carlo sampling, the kernel function is approximated as:

$$k(x, y) \approx \phi(x)^\top \phi(y),$$

where $\phi(x)$ is the random Fourier feature mapping:

$$\phi(x) = \sqrt{\frac{2}{D}} \cos(W^\top x + b),$$

with $W \sim \mathcal{N}(0, \sigma^2 I)$ and $b \sim \text{Uniform}(0, 2\pi)$. Here, D is the number of Fourier features, which balances computational efficiency and approximation accuracy.

6.9.3 Differences Between RCIT and RCoT

RCIT and RCoT differ in their test statistics, computational efficiency, and practical performance, which makes them suited for different scenarios in causal discovery. RCIT evaluates the Hilbert-Schmidt norm of the full partial cross-covariance operator, providing a general test for conditional independence but at a higher computational cost. RCoT simplifies the process by using the Frobenius norm of a finite-dimensional residualized cross-covariance matrix, significantly reducing complexity and improving scalability.

These distinctions are particularly important for large-scale datasets, where RCoT's computational efficiency makes it a practical choice for high-dimensional causal discovery tasks.

6.9.3.1 RCIT: Randomized Conditional Independence Test

RCIT tests full conditional independence by examining the squared Hilbert-Schmidt (HS) norm of the partial cross-covariance operator $\Sigma_{XY \cdot Z}$:

$$S_K = nT_{XY \cdot Z} = \frac{1}{n}\text{tr}(K_{X \cdot Z}K_{Y \cdot Z}),$$

where $T_{XY \cdot Z}$ is an empirical estimate of $\|\Sigma_{XY \cdot Z}\|_{HS}^2$. The null and alternative hypotheses are:

$$H_0 : \|\Sigma_{XY \cdot Z}\|_{HS}^2 = 0, \quad H_1 : \|\Sigma_{XY \cdot Z}\|_{HS}^2 > 0.$$

RCIT is a general test for conditional independence but becomes computationally demanding as the size of Z increases, due to the high-dimensional kernel operations required.

6.9.3.2 RCoT: Randomized Conditional Correlation Test

RCoT simplifies the testing process by using a finite-dimensional partial cross-covariance matrix, avoiding full HS norm calculations. Instead, it uses the Frobenius norm of the residualized cross-covariance matrix:

$$S' = n\|C_{AB \cdot C}\|_F^2,$$

where $C_{AB \cdot C}$ represents the residualized cross-covariance matrix. The hypotheses are:

$$H_0 : \|C_{AB \cdot C}\|_F^2 = 0, \quad H_1 : \|C_{AB \cdot C}\|_F^2 > 0.$$

RCoT is computationally efficient and well-suited for large conditioning sets ($|Z| \geq 4$). Its simplicity enables robust calibration of the null distribution and improved scalability for high-dimensional data.