Online Appendix: Supplementary Information for “Contrastive Multiple Correspondence Analysis (cMCA): Using Contrastive Learning to Identify Latent Subgroups in Political Parties”

Takanori Fujiwara\(^1\) and Tzu-Ping Liu\(^2\)

\(^1\)Department of Computer Science, University of California at Davis, Davis, CA 95616, USA. Email: tpliu@ucdavis.edu
\(^2\)Department of Political Science, University of California at Davis, Davis, CA 95616, USA. Email: tpliu@ucdavis.edu

B Variable Coding Scheme

B.1 Variable Coding Scheme for the ESS 2018 Survey

**Left-Right Scale** \( (lrscale) \): This is an ordinal variable measuring respondents' self-identified ideologies on the left-right scale. This variable is originally coded from a scale of 0 (left) to 10 (right). We further recode this variable as a five-point scale: respondents who originally responded as 0 or 1 are recoded as 1, those who originally responded as 2 or 3 are recoded as 2, those who originally responded as 4, 5, or 6 are recoded as 3, those who originally responded as 7 or 8 are recoded as 4, and those who originally responded as 9 or 10 are recoded as 5.

**Trust in the European Parliament** \( (trstep) \): This is an ordinal variable measuring respondents' self-identified trust in the European Parliament. This variable is originally coded from a scale of 0 (not trust at all) to 10 (completely trust). We further recode this variable as a five-point scale: respondents who originally responded as 0 or 1 are recoded as 1, those who originally responded as 2 or 3 are recoded as 2, those who originally responded as 4, 5, or 6 are recoded as 3, those who originally responded as 7 or 8 are recoded as 4, and those who originally responded as 9 or 10 are recoded as 5.

**Trust in the United Nations** \( (trstun) \): This is an ordinal variable measuring respondents' self-identified trust in the United Nations. This variable is originally coded from a scale of 0 (not trust at all) to 10 (completely trust). We further recode this variable as a five-point scale: respondents who originally responded as 0 or 1 are recoded as 1, those who originally responded as 2 or 3 are recoded as 2, those who originally responded as 4, 5, or 6 are recoded as 3, those who originally responded as 7 or 8 are recoded as 4, and those who originally responded as 9 or 10 are recoded as 5.

**Democracy Satisfaction** \( (stfdem) \): This is an ordinal variable measuring respondents' self-identified...
satisfaction with how democracy works in the U.K. This variable is originally coded from a scale of 0 (extremely dissatisfied) to 10 (extremely satisfied). We further recode this variable as a five-point scale: respondents who originally responded as 0 or 1 are recoded as 1, those who originally responded as 2 or 3 are recoded as 2, those who originally responded as 4, 5, or 6 are recoded as 3, those who originally responded as 7 or 8 are recoded as 4, and those who originally responded as 9 or 10 are recoded as 5.

**State of Health in the U.K.** (*stfhlt*): This is an ordinal variable measuring respondents' self-identified feeling about the state of health services in the U.K. nowadays. This variable is originally coded from a scale of 0 (extremely dissatisfied) to 10 (extremely satisfied). We further recode this variable as a five-point scale: respondents who originally responded as 0 or 1 are recoded as 1, those who originally responded as 2 or 3 are recoded as 2, those who originally responded as 4, 5, or 6 are recoded as 3, those who originally responded as 7 or 8 are recoded as 4, and those who originally responded as 9 or 10 are recoded as 5.

**Income Inequality** (*gincdif*): This is an ordinal variable measuring respondents' self-identified agreement with that governments should reduce differences in income levels. This variable is originally coded from a scale of 1 (agree strongly) to 5 (disagree strongly).

**Gay Life** (*freehms*): This is an ordinal variable measuring respondents' self-identified agreement with that gays and lesbians free to live life as they wish. This variable is originally coded from a scale of 1 (agree strongly) to 5 (disagree strongly).

**Gay Shame** (*hmsfmlsh*): This is an ordinal variable measuring respondents' self-identified agreement with that respondents are shamed if close family member is gay or lesbian. This variable is originally coded from a scale of 1 (agree strongly) to 5 (disagree strongly).

**Gay Adoption** (*hmsacld*): This is an ordinal variable measuring respondents' self-identified agreement with that gay and lesbian couples adopt children. This variable is originally coded from a scale of 1 (agree strongly) to 5 (disagree strongly).

**European Unification** (*euftf*): This is an ordinal variable measuring respondents' self-identified feelings about European Union Unification between it has gone too far and it should go further. This variable is originally coded from a scale of 0 (gone too far) to 10 (going further). We further recode this variable as a five-point scale: respondents who originally responded as 0 or 1 are recoded as 1, those who originally responded as 2 or 3 are recoded as 2, those who originally responded as 4, 5, or 6 are recoded as 3, those who originally responded as 7 or 8 are recoded as 4, and those who originally responded as 9 or 10 are recoded as 5.

**Same Race Immigrants** (*imsmetn*): This is an ordinal variable measuring respondents' self-identified feelings about allowing immigrants of the same race as the majority. This variable is originally
Different Race Immigrants (imdfetr): This is an ordinal variable measuring respondents' self-identified feelings about allowing immigrants of the different race as non-majority. This variable is originally coded from a scale of 1 (allow many) to 5 (allow none).

Non-European & Poor Immigrants (impcntr): This is an ordinal variable measuring respondents' self-identified feelings about allowing immigrants who are poor and from non-European countries. This variable is originally coded from a scale of 1 (allow many) to 5 (allow none).

Immigration to Economy (imbgeco): This is an ordinal variable measuring respondents' opinions about immigration's impacts on local economy. This variable is originally coded from a scale of 0 (bad for economy) to 10 (good for economy). We further recode this variable as a five-point scale: respondents who originally responded as 0 or 1 are recoded as 1, those who originally responded as 2 or 3 are recoded as 2, those who originally responded as 4, 5, or 6 are recoded as 3, those who originally responded as 7 or 8 are recoded as 4, and those who originally responded as 9 or 10 are recoded as 5.

Immigration to Culture (imueclt): This is an ordinal variable measuring respondents' opinions about whether immigration undermines or enriches local cultural life. This variable is originally coded from a scale of 0 (undermining) to 10 (enriching). We further recode this variable as a five-point scale: respondents who originally responded as 0 or 1 are recoded as 1, those who originally responded as 2 or 3 are recoded as 2, those who originally responded as 4, 5, or 6 are recoded as 3, those who originally responded as 7 or 8 are recoded as 4, and those who originally responded as 9 or 10 are recoded as 5.

Immigration to Living Quality (imwbcnt): This is an ordinal variable measuring respondents' opinions about whether immigration makes the U.K. a worse or better place to live. This variable is originally coded from a scale of 0 (worse) to 10 (better). We further recode this variable as a five-point scale: respondents who originally responded as 0 or 1 are recoded as 1, those who originally responded as 2 or 3 are recoded as 2, those who originally responded as 4, 5, or 6 are recoded as 3, those who originally responded as 7 or 8 are recoded as 4, and those who originally responded as 9 or 10 are recoded as 5.

Emotionally Attached to Europe (atcherp): This is an ordinal variable measuring respondents' feelings about how emotionally they are attached to Europe. This variable is originally coded from a scale of 0 (not at all) to 10 (very). We further recode this variable as a five-point scale: respondents who originally responded as 0 or 1 are recoded as 1, those who originally responded as 2 or 3 are recoded as 2, those who originally responded as 4, 5, or 6 are recoded as 3, those who originally responded as 7 or 8 are recoded as 4, and those who originally responded as 9 or 10 are recoded as 5.
5.

**Religiosity** ($\text{rlgdgr}$): This is an ordinal variable measuring respondents’ religious levels. This variable is originally coded from a scale of 0 (not at all) to 10 (very). We further recode this variable as a five-point scale: respondents who originally responded as 0 or 1 are recoded as 1, those who originally responded as 2 or 3 are recoded as 2, those who originally responded as 4, 5, or 6 are recoded as 3, those who originally responded as 7 or 8 are recoded as 4, and those who originally responded as 9 or 10 are recoded as 5.

**Equality** ($\text{ipeqopt}$): This is an ordinal variable measuring respondents’ opinions about that it is important that people are treated equally and have equal opportunities. This variable is originally coded from a scale of 1 (very much like me) to 6 (not like me at all). We further recode this variable as a three-point scale: respondents who originally responded as 1 or 2 are recoded as 1, those who originally responded as 3 or 4 are recoded as 2, and those who originally responded as 5 or 6 are recoded as 3.

**Understanding** ($\text{ipudrst}$): This is an ordinal variable measuring respondents’ opinions about that it is important to understand different people. This variable is originally coded from a scale of 1 (very much like me) to 6 (not like me at all). We further recode this variable as a three-point scale: respondents who originally responded as 1 or 2 are recoded as 1, those who originally responded as 3 or 4 are recoded as 2, and those who originally responded as 5 or 6 are recoded as 3.

**Strong Government** ($\text{ipstrgv}$): This is an ordinal variable measuring respondents’ opinions about that it is important that government is strong and ensures safety. This variable is originally coded from a scale of 1 (very much like me) to 6 (not like me at all). We further recode this variable as a three-point scale: respondents who originally responded as 1 or 2 are recoded as 1, those who originally responded as 3 or 4 are recoded as 2, and those who originally responded as 5 or 6 are recoded as 3.

**Environmental Protection** ($\text{impenv}$): This is an ordinal variable measuring respondents’ opinions about that it is important to care for nature and environment. This variable is originally coded from a scale of 1 (very much like me) to 6 (not like me at all). We further recode this variable as a three-point scale: respondents who originally responded as 1 or 2 are recoded as 1, those who originally responded as 3 or 4 are recoded as 2, and those who originally responded as 5 or 6 are recoded as 3.

**Tradition** ($\text{imptrad}$): This is an ordinal variable measuring respondents’ opinions about that it is important to follow traditions and costumes. This variable is originally coded from a scale of 1 (very much like me) to 6 (not like me at all). We further recode this variable as a three-point scale: respondents who originally responded as 1 or 2 are recoded as 1, those who originally responded as 3 or 4 are recoded as 2, and those who originally responded as 5 or 6 are recoded as 3.
as 3 or 4 are recoded as 2, and those who originally responded as 5 or 6 are recoded as 3.

B.2 Variable Coding Scheme for the UTAS 2012 Survey

**PartyID (psup_short):** This is a categorical variable measuring respondents’ self-identified partisanship. This variable is originally coded by parties. We further recode this variable through only keeping respondents who self-identify themselves as supporters of LDP, DPJ, and JRP.

**Ideology (policy00):** This is an ordinal variable measuring respondents’ self-identified ideologies. This variable is originally coded from a scale of 0 (left) to 10 (right).

**Electoral System (policy01):** This is an ordinal variable measuring respondents’ agreement with going back to the single non-transferable vote (SNTV) system. This variable is originally coded from a scale of 1 (support) to 5 (oppose).

**HoR over HoC (policy04):** This is an ordinal variable measuring respondents’ agreement with strengthening the superiority of the House of Representatives (HoR) over the House of Councilors (HoC). This variable is originally coded from a scale of 1 (support) to 5 (oppose).

**Constitution Change (policy06):** This is an ordinal variable measuring respondents’ agreement with changing the constitution. This variable is originally coded from a scale of 1 (support) to 5 (oppose).

**Defense Capacity (policy07):** This is an ordinal variable measuring respondents’ agreement with strengthening Japan’s defense capacity. This variable is originally coded from a scale of 1 (support) to 5 (oppose).

**Preemptive Attack (policy09):** This is an ordinal variable measuring respondents’ agreement with that Japan should not hesitate preemptive attack if the attack from another country is expected. This variable is originally coded from a scale of 1 (support) to 5 (oppose).

**UN Security Council (policy10):** This is an ordinal variable measuring respondents’ agreement with that Japan should become a permanent member of the UN Security council and fulfill its international duty. This variable is originally coded from a scale of 1 (support) to 5 (oppose).

**Pressure to North Korea (policy11):** This is an ordinal variable measuring respondents’ agreement with prioritizing pressure over dialogue for North Korea. This variable is originally coded from a scale of 1 (support) to 5 (oppose).

**Collective Defense (policy12):** This is an ordinal variable measuring respondents’ agreement with that Japan should exercise collective defense right. This variable is originally coded from a scale of 1 (support) to 5 (oppose).

**Small Government (policy14):** This is an ordinal variable measuring respondents’ agreement with that small government is better even when public service such as social welfare may get worse.
This variable is originally coded from a scale of 1 (support) to 5 (oppose).

**Public Work Security** (policy16): This is an ordinal variable measuring respondents’ agreement with securing employment of public works. This variable is originally coded from a scale of 1 (support) to 5 (oppose).

**Fiscal Policy** (policy17): This is an ordinal variable measuring respondents’ agreement with that the government should conduct fiscal mobilization instead of reducing the budget for now. This variable is originally coded from a scale of 1 (support) to 5 (oppose).

**Consumption Tax Raise** (policy19): This is an ordinal variable measuring respondents’ agreement with the unavoidable consumption tax raise within 5 years. This variable is originally coded from a scale of 1 (support) to 5 (oppose).

**Permanent Foreign Residents with Local Suffrage** (policy21): This is an ordinal variable measuring respondents’ agreement with allowing permanent resident foreigners to have local suffrage. This variable is originally coded from a scale of 1 (support) to 5 (oppose).

**Foreign Workers** (policy22): This is an ordinal variable measuring respondents’ agreement with increasing the acceptance of foreign workers. This variable is originally coded from a scale of 1 (support) to 5 (oppose).

**Limiting Individual Rights** (policy23): This is an ordinal variable measuring respondents’ agreement with that it is fine to limit privacy and individual rights for the sake of public safety. This variable is originally coded from a scale of 1 (support) to 5 (oppose).

**Consumption Tax Long Run** (policy26): This is an ordinal variable measuring respondents’ agreement with unavoidable consumption tax rate going above 10% in the long run. This variable is originally coded from a scale of 1 (support) to 5 (oppose).

**Pension-Asset Balance** (policy27): This is an ordinal variable measuring respondents’ agreement with withdrawing pension to those who have assets. This variable is originally coded from a scale of 1 (support) to 5 (oppose).

**Trans-Pacific Partnership** (policy28): This is an ordinal variable measuring respondents’ agreement with joining Trans-Pacific Partnership. This variable is originally coded from a scale of 1 (support) to 5 (oppose).

**Central Bank** (policy29): This is an ordinal variable measuring respondents’ agreement with that Japan central bank should set clear inflation target. This variable is originally coded from a scale of 1 (support) to 5 (oppose).

**Moral Education** (policy30): This is an ordinal variable measuring respondents’ agreement with providing more moral education. This variable is originally coded from a scale of 1 (support) to 5 (oppose).
Nuclear Power Plant Restart (policy31): This is an ordinal variable measuring respondents’ agreement with restating nuclear power plant that are temporarily suspended due to maintenance is unavoidable. This variable is originally coded from a scale of 1 (support) to 5 (oppose).

Local Government (policy32): This is an ordinal variable measuring respondents’ agreement with that local government outside the disaster area should actively accept debris. This variable is originally coded from a scale of 1 (support) to 5 (oppose).

Prime Minister Selection (policy33): This is an ordinal variable measuring respondents’ agreement with public selection of the prime minister. This variable is originally coded from a scale of 1 (support) to 5 (oppose).

Prefecture System (policy34): This is an ordinal variable measuring respondents’ agreement with that the current system should be replaced by prefecture. This variable is originally coded from a scale of 1 (support) to 5 (oppose).

Consumption Tax-Local Tax (policy35): This is an ordinal variable measuring respondents’ agreement with that consumption tax should become local tax and local allocation tax should be abolished. This variable is originally coded from a scale of 1 (support) to 5 (oppose).

Cutting Legislature Seats (policy36): This is an ordinal variable measuring respondents’ agreement with that the number of lawmakers should be cut in half. This variable is originally coded from a scale of 1 (support) to 5 (oppose).

Livings Standard-Environmental Protection (policy37): This is an ordinal variable measuring respondents’ feelings between living standard and environmental protection. This variable is originally coded from a scale of 1 (more to living standard) to 5 (more to environmental protection).

Family Form (policy43): This is an ordinal variable measuring respondents’ feelings between the traditional form (husband, wife, and kids) and the flexible form (including single mother and DINKs). This variable is originally coded from a scale of 1 (more to the traditional form) to 5 (more to the flexible form).

Domestic Industry-Open Market (policy45): This is an ordinal variable measuring respondents’ feelings between protecting domestic industries and opening domestic market. This variable is originally coded from a scale of 1 (more to protecting domestic industries) to 5 (more to opening domestic market).

US-Asia Priority (policy47): This is an ordinal variable measuring respondents’ feelings of foreign policies should prioritize between the U.S. and Asia. This variable is originally coded from a scale of 1 (more to the U.S.) to 5 (more to Asia).

Competition-Inequality Priority (policy48): This is an ordinal variable measuring respondents’ feelings of economic policies should prioritize between economic competition and inequality.
This variable is originally coded from a scale of 1 (more to economic competition) to 5 (more to inequality).

**Nuclear Power 2030** (policy49): This is an ordinal variable measuring respondents’ feelings between abandoning nuclear power by 2030 and continuing using nuclear power after 2030. This variable is originally coded from a scale of 1 (more to abandoning) to 5 (more to continuing).

**Campaign Contribution** (policy50): This is an ordinal variable measuring respondents’ feelings between strict campaign contribution law and loose campaign contribution law. This variable is originally coded from a scale of 1 (more to abandoning) to 5 (more to continuing).

**Diet** (policy51): This is an ordinal variable measuring whether respondents feel about cutting legislature seats due to political conflict or coordination. This variable is originally coded from a scale of 1 (more to conflict) to 5 (more to coordination).

**Party Control** (policy52): This is an ordinal variable measuring respondents’ feelings regarding party members should vote uniformly or independently during roll-call votes. This variable is originally coded from a scale of 1 (more to vote uniformly) to 5 (more to vote independently).

**Feeling Close to the U.S.** (policy53): This is an ordinal variable measuring respondents’ familiarity with the U.S. This variable is originally coded from a scale of 1 (familiar) to 5 (not familiar).

**Feeling Close to China** (policy54): This is an ordinal variable measuring respondents’ familiarity with China. This variable is originally coded from a scale of 1 (familiar) to 5 (not familiar).

**Feeling Close to Russia** (policy55): This is an ordinal variable measuring respondents’ familiarity with Russia. This variable is originally coded from a scale of 1 (familiar) to 5 (not familiar).

**Feeling Close to South Korea** (policy56): This is an ordinal variable measuring respondents’ familiarity with South Korea. This variable is originally coded from a scale of 1 (familiar) to 5 (not familiar).