Leadership quality trade-off and vote buying: Evidence from the world's third largest democracy
Pre-Analysis Plan

Naila M. Shofia & Piero Stanig

January 2024

1   Introduction

This document set out the pre-analysis plan for the project in the title. The project is conducted in three waves of surveys in conjunction with the 2024 Indonesia General Election which will be held on the 14th February 2024. There are four separate analysis we plan to conduct using the data collected in the surveys, namely 1) A Conjoint analysis, 2) Turnout analysis 3) List Experiment 4) Political dynasty analysis.

The following sections elaborate the issues of measurements using different set of survey questions and their coding and interpretation. The subsequent sections elaborate the the main analysis, including the main sets of questions to be explored together with the plan of their analysis.

2   Measurement Issues

2.1   Measuring individual ideological positions

To measure the position of individual respondents on the divisive issues that are also randomized in the candidate profiles in the conjoint module discussed below, we rely on a set of items asked in the baseline. The respondets are asked about their opinion on the following issues:

1. The military takes over if the government is not competent.

2. Three items on enforcement of traditional morality

    (a) The state criminalizes those involved in abortion activities.

    (b) The state criminalizes individuals involved in prostitution.

    (c) The state criminalizes individuals involved in adultery.

3. The government dissolves organizations deemed as extremists.

4. The government enacts laws to facilitate legal processes against the defamation of public officials on social media.

5. The government opens the local labor market by importing skilled foreign workers.

All these are answered on a 4-point scale (Strongly agree, Agree, Disagree, Strongly disagree). We treat the "unsure" option as missing. We assign numerical values -1, -.5, .5, 1 to the four possible answers.

In the multi-item case of traditional morality, the item asked to respondents is randomized, so that each respondent only answers one of the three questions. We treat the answer as a measure of the respondent's stance on traditional morality issue, regardless of which question out of the three they received, i.e. we treat these responses as the measure of respondent i's stance on issue k, x_i^k in the random utility analysis below.

2.2 Political knowledge

Respondent political knowledge is measured in two ways: 1) By asking about factual knowledge about politics, and 2) By using a categorization game to gauge the respondent's understanding of democratic accountability.

1. Factual knowledge is measured with three items:

   - "Who is the current vice president of Indonesia?"
   - "How many years is the tenure of member of parliament of Republic of Indonesia (DPR RI)?"
   - "Which party holds the second largest seat in the Indonesian parliament (DPR RI) today?"

   These are multiple choice questions with four options. The order of the answer options is randomised. Only one option is correct and will be coded as 1 if picked, 0 is assigned if any other option is picked. We then fit a two-parameter logistic IRT model through the four questions. All non-missing items in an observation are included in the likelihood calculation; only missing items are excluded.[1]

2. Understanding of democratic election is measured using a drag-and-drop categorization game: we ask the respondents to move objects from one box to another box based on the category that they think is correct.

   The wording of the question is: "From the three objects on the left, there is one object that is similar to the object in each right box. Move (drag and release) one object from the left box to each corresponding right box. (There is only one correct object in each of the right boxes)"

---

1. This entails **not** specifying `listwise` as an option in Stata `irt 2pl` command.

The first two items in this block are practice questions, asking respondents to categorize fruits and working tools. We do not use this question as filtering, but only as an exercise and acclimation to the actual questions.

There are two substantive questions following the practice questions, and the instructions are the same:

"Same as before, we kindly ask for your assistance in matching one object in the left box with the category in the right box that you find most suitable."

(a) First substantive item

- Items to be placed (left box): Subjects (of a king); giver of authority (principal); receiver of authority (agent).
- Filled box 1 header : Voter
- Filled box 2 header: Political candidate
- Scoring rule:
    - we assign full credit (1) to answers that match "voter" with "principal" and match "political candidate" with "agent"
    - we score -1 answers that invert the principal-agent relationship by matching voter with agent, and political candidate with principal
    - zero otherwise.

(b) Second substantive item

- Items to be placed (left box): King, Shop employee, Shop owner, Subjects (of a king).
- Filled box 1 header: Voter
- Filled box 2 header : Parliament member
- Scoring rule:
    - We assign full points (2) to the respondents who correctly place the principal (shop owner) in the "voter" box, and the agent (shop employee) in the "parliament member" box;
    - We give partial credit (1) for answers that place "king" in the voter box and "subject" in the parliament member box;
    - Negative credit (-2) is assigned to answers that get the principal-agent relationship wrong –voters matched with "subjects" and politician matched with "king", or voters matched with shop manager and politician matched with shop owner
    - No credit (0) is awarded otherwise.

From these two, we create an aggregate score at the respondent level as the weighted average of the two scores (each rescaled on a -1 to 1 scale), where the second gets twice the weight of the first.

2.3 Measuring media exposure and political interest

To measure media exposure, we use the following item:
   "In the past 24 hours, have you (check all that apply):"
   Choosing multiple items from the following list is allowed:

- Used social media (such as Facebook, Youtube or Whatsapp)

- Watched TV News, Read a newspaper in print or online

- Listened to a radio news program or talk radio

- None of these."

The exposure is measured by assigning a value of 1 to each media selected, and zero if option *None of these* is selected. The sum of all the indicators for use of one medium is the media exposure score for the individual. Hence this takes values from 0 to 3.

To measure interest in politics, we use the following item:
   "In the past 24 hours, did you do any of the following on social media (such as Facebook, Youtube, Twitter, or WA)"
   Choosing multiple items from the following list is allowed:

- Posted a story or link about politics

- Posted a comment about politics

- Read a story or watched a video about politics

- Followed a political event

- Forwarded a story, photo, video or link about politics to friend

- None of the above "

The level of interest is measured by assigning a value of 1 to each action selected, and zero if option *None of the above* is selected. Again the political interest score is the sum of the binary indicators for each activity. Hence this score varies from 0 to 5.

2.4 Strength of religiosity

Two questions are used to measure respondent's level of religiosity:

- First, we ask a question on which religion is professed by the candidate. Then based on the answer, we ask the frequency of attendance in worship activity (on a 5-point scale).

- We also ask "How important is religion for you in making important life decisions? (such as marriage, changing jobs, moving residences)" on 4 point scale

As strength of religiosity we take the simple average of the two items, originally scaled respectively from 0 to 4 and 0 to 3, and rescaled dividing respectively by 4 and by 3, so they range from 0 to 1.

3   Data collection

The survey targets a representative sample of Indonesian voters who will be voting in the 2024 Indonesia General Election on 14 February 2024. There are three waves of the survey, targeting a panel of 3600 respondents at the end of the third wave. With this design we will collect more respondents in the baseline to accommodate for attrition in the subsequent stages of the survey. Under the assumption of 30 percent attrition at each subsequent survey, the number of responses collected will be around 7347 at the baseline.

The timing of the surveys is as follows

1. First wave: Baseline survey, fielded 5 - 4 weeks prior to the election.

2. Second wave: Pre-election conjoint and pre-election list survey experiment, fielded 3-2 weeks prior to the election.

3. Third wave: Post-election conjoint and post-election list survey experiment, fielded from 3 days to 2 weeks after the election

The survey is distributed online in collaboration with a local marketing polling companies which maintains a panel of respondents with an excellent reach of population across islands in Indonesia. We collect the responses via Qualtrics questionnaire designed and managed by our research team; the survey company only play the role of connecting their panel respondents to our survey links.

We instruct the company regarding quotas by province of residence, gender, age group, rural-urban, and education conditional on age group (over or under-40), based on the 2010 census, and 2024 voter distribution published by the Indonesian Electoral Commission.

4   Experimental manipulation

Respondents are randomly assigned, in the second wave, to one of two informational treatments: a short video explaining the accountability view of democracy (treatment condition in the following), a video of the same length and graphical presentation, explaining how voting operations in the general election work (control condition 1 in the following), or no video (control condition 2 in the following). The probability of being in the treatment arm is 50% (and independent of respondent characteristics). The probability of being in control condition 1 or 2 is 25% each. Full video scripts (in English (translation) and Bahasa Indonesia) are found in Appendix 9.1. The videos themselves will be made available online.

5   Conjoint analysis

In the conjoint experiment, 15 pairs of candidates are presented to respondents. They differ on the following:

- gender

- age

- religion

- ethnicity

- education

- marriage status

- occupation

- position on a contentious issue

- valence item

The characteristics are presented in the above order to all respondents. The detailed list of characteristics and the distribution of each level of the attribute is presented in appendix 9.2. This distribution is benchmarked against the real distribution of candidates running in the General Election in the previous cycle (2019).

Basic characteristics   Religion can take the following values, with probabilities in parenthesis "Islam" (.7) ,"Kristen Protestan" (.1), "Katolik" (.1), "Hindu" (.05), and "Buddha" (.05). These reflect the actual distribution of religious affiliation in Indonesia, but higher probability is assigned to the minority ones.

Ethnicity   The distribution of the ethnicity of the candidate is province-specific. The respondent is asked in which province they vote, and the respective province-specific distribution is used to randomize the ethnicity. The detailed list of provinces and candidates ethnicity distribution is presented in appendix 9.3. To increase the realism of candidate distribution in each province, and to have a consistency in all provinces in coding ethnic affiliation, the ethnicity distributions of the candidates are designed so that half of the candidates is from the native titular ethnicity in the province (defined as the largest ethnicity), 40 percent of the candidates will come from other native ethnicities (2nd - 4th largest native ethnicities in the province), meanwhile, the rest (10%) is assigned to a minority ethnicity (Chinese Indonesian–Tionghoa).

Contentious issue   This characteristics maps one-to-one to the items on divisive issues in the baseline survey. Below we present the wording used in the conjoint profile of candidates:

1. The role of the Indonesian National Armed Forces (TNI) vs. the government

   - Strengthening the role of the Indonesian National Armed Forces (TNI) within the state institutions for the sake of stability and the people's security.
   - Prioritizing a democratic system by ensuring the neutrality of the Indonesian National Armed Forces (TNI).

2. Traditional morality

   (a) Abortion:
      - Strengthening law enforcement for the protection of children, even starting from during pregnancy.
      - Ensuring the protection of women's rights to abortion in cases of rape and indications of medical emergencies.

   (b) Prostitution:
      - Upholding societal norms, morals, and religious values, such as criminalizing activities related to prostitution.
      - Ensuring the protection of the rights of female workers in the prostitution industry.

   (c) Homosexuality
      - Implementing local regulations in accordance with Sharia, such as regulating issues related to adultery and sexual deviations.
      - Ensuring human rights and rejecting laws that can discriminate against specific groups.

3. Freedom of association

   - Rejecting radicalism, extremism, and intolerance that can threaten regional security.
   - Ensuring the freedom of organization and association.

4. Freedom of speech

   - Strengthening law enforcement against the spread of hate speech targeting public officials within the digital realm.
   - Ensuring the right to freedom of speech and expression in the digital domain, including the ability to criticize public officials.

5. Economic nationalism:

- Promoting a foreign investment climate to expand job opportunities and support regional economic growth.
- Protecting the rights and obligations of local workers in the region.

For each profile, one of the issues is selected randomly (with equal probability .2) and then the polarity of the statement is selected (with 50% probability). For the traditional morality (selected with probability .2) we select one sub-issue (with 1/3 probability) and then the polarity (with .5 probability).

Valence The candidates differ on the following valence dimensions. This item maps to questions included in the baseline survey, where respondents are asked to list two most important problem to tackle in the country. Yet, we treat these as positive/negative traits of the candidate regardless of whether the respondent indicated, in the baseline, that a given dimension was important.

1. Corruption:

   - High
     - No track record of corruption.
     - Successfully reduced the level of corruption in local government.
   - Low
     - Previously involved in corruption cases.
     - Involved in bribery cases for business permits.
     - Former convict.

2. Education and health:

   - High
     - Improved the quality of schools in the region.
     - Improved the quality of service and infrastructure for health facilities in the region.
     - Established free schools.
     - Increased school participation rates in the region.
   - Low
     - Failed to reduce maternal and infant mortality rates and cases of stunting in the region.
     - Unsuccessful in curbing the dropout rate in the region.

3. Economy:

   - High
     - Successfully reduced poverty, inequality, and unemployment rates in the region.

- Improved guidance and support for SMEs to boost the local economy.

- Low
  - Failed to control the stability of basic commodity prices in the region.
  - Increased poverty, inequality, and unemployment rates in the region during the term.

4. Environment:

- High
  - Successfully emphasized carbon emissions in the region.
  - Addressed the emergency waste issue in the region.

- Low
  - Failed to address air pollution and river pollution from factory waste.
  - Failed to address the waste emergency in the region.
  - Granted permits for the opening of mining companies.

5. Rights of women and children:

- High
  - Proposed various new programs for the empowerment of women.
  - Improved the quality of reporting services for domestic violence.

- Low
  - Unsuccessful in reducing the number of domestic violence cases in the region.
  - Failed to control the number of child laborers in the region.

For each profile, first an issue is selected (with equal probability 1/5) then the polarity is selected with equal probability (50%), then a wording within the chosen issue-polarity combination is selected (with equal probability). Please note that the number of alternative items (specific wording) within a given issue-polarity combination is not balanced; yet the probability of selecting any one issue is constant at 1/5 and the probability of having a positive (negative) valence trait is constant at .5. We use several alternative wordings for a given issue to increase the natural appearance of different issues across candidate profile choice, as the appearance of two identical positions across profiles could be potentially confusing and detract from the realism of the choice exercise.

Exclusion of highly unlikely combinations of ethnicity and religion.
We exclude from the distribution of traits a handful of combinations of ethnicity and religion that are highly unlikely to appear in the real world. Notice that these would themselves be unlikely (albeit possible) in the joint randomization. Yet, to avoid possible distraction caused by these among respondents, we eliminate them completely. The excluded combinations are those for religions that have less than .0005 probability (equivalent to 500 people per million) conditional on ethnicity according to the 2010 census.

5.1 Data definitions

As per standard voting models, we differentiate between positional issues, identity, and candidate valence.

For the positional issues, we can classify each profile as close or distant to the respondent based on the answers the respondent provided in the baseline survey. The answers are given on a 4-point scale. To calculate the distances that enter the random utility model described below, we take the square of difference between the position of the candidate (respectively -1, 1 or 0 if not mentioned) and the position of the respondent (where strongly agree is coded -1, agree -.5, disagree .5, and strongly disagree 1.). Hence the distances range from 0 to 4. In the AMCE estimation, we simply consider candidates that are congruent with, or divergent from, the position stated by the respondent in the baseline (recoded so that "strongly agree" and "agree", and 'strongly disagree" and "disagree", are recoded into a binary agree/disagree feature).

We consider that ethnicity and religion can be pure identity items, but also "valence" characteristics (so that, for instance, being ethnic Javanese makes a candidate more appealing also to other ethnic groups). Hence these features enter the random utility models in two different ways:

- as the difference between respondent ethnicity (religion) and candidate ethnicity (religion), where the identity congruence item is equal to one if respondent and candidate have the same ethnicity (religion) and equal to zero if respondent and candidate do not share the same ethnicity (religion)

- as valence items, where being member of one ethnicity or religion confers some advantage/disadvantage to the candidate independent of the traits of the respondent.

Similarly, we are agnostic about whether age, gender and education should be treated as identity or valence issues. Therefore in the random utility estimation we estimate: a full model that includes terms both for distance (e.g., absolute difference between age of respondent and age of candidate; university education of both candidate and respondent vs different level of education; etc.) –hence treating these as identity markers– and for values of the candidate characteristic (e.g., dummies for age group of candidate and for university education of the candidate) –hence treating them as valence terms; a model that only treats them as identity items (hence only including the differences between candidate

10

and respondent); and a model that only treats them as valence items (hence only including the dummies for candidate characteristics.

5.2   AMCE (Average Marginal Component Effect) estimation

The data from the conjoint experiment can be analyzed following the approach in Hainmueller, Hopkins, and Yamamoto 2014. Under the assumptions spelled out there, we can pool all the data from the 15 choices, for a total of 30 profiles.

Given that assumption 1 in Hainmueller, Hopkins, and Yamamoto 2014 might be violated, as the sequence in which profiles are presented might affect how respondents evaluate them, we also present the results of a restrictive analysis only on the first pair of choices presented to each respondent.

For substantive purposes, we group the components into three subsets:

- descriptive/identity proximity

- ideological/policy proximity

- valence

In the first group we include ethnicity, religion, (plus gender, education, and age). Hence the proximity variables for identities are defined to be equal to one if the respondent shares ethnicity, religion, (and gender, education, or age group) with the candidate, and zero otherwise. We estimate the AMCE for each of these re-defined binary predictors.

In the second subset there is the proximity on the divisive (ideological) issues. This is defined to be equal to one if the position of the respondent (as stated in the baseline) is on the same side as the position of the candidate (respondent strongly agrees or agree in a given item in the baseline, and the candidate is on the same side.) We estimate the AMCE for each of these re-defined binary predictors. In supplementary analysis, we use the ordinal scale based on the difference between how respondents place themselves and the position of the candidate.

We analyze the AMCE for these binary attributes. These are redefined (compared to the randomization) as they incorporate congruence between respondent's characteristics and candidate features.

Finally, the valence issue can take the value of 1 if the candidate has a positive attribute, and -1 if the candidate has a negative attribute. We estimate the AMCE, dimension-wise,[2] for having a positive characteristics vs not having anything; having a negative characteristic vs not having anything; and positive vs. negative.

We also create an alternative set of profile-level attributes, where a candidate is defined: as identity-congruent if it shares *both* ethnicity and religion with

_____

2. The valence dimensions, as described above, are: Corruption; Education and Health; Economy; Environment; Rights of Women and Children. In addition, we treat religion and ethnicity –and in additional analysis also demographics– also as valence items.

the respondent; as ideologically congruent if it stands on the same side as the respondent on the divisive issue; and as high (low) valence if it has a positive (negative) valence attribute. We then use only these three components and estimate the AMCE for each of these. In practice, for ideology and valence in this piece of analysis we ignore the specific issue dimension that was presented in the profile.

All inferences for AMCE and ACIE (Average Component Interaction Effect) are based on the standard error clustered at the respondent level as per Hainmueller, Hopkins, and Yamamoto 2014

5.2.1 Interaction between candidate characteristics

We estimate the ACIE for

- the binary identity-congruence variable (equal to one if *both* ethnicity and religion match between respondent and candidate, zero otherwise) and respectively the binary valence and the binary ideological congruence (i.e., ignoring the specific issue dimension included in the profile)

- ethnic (religious) congruence and respectively the binary valence and the binary ideological congruence (i.e., ignoring the specific issue dimension included in the profile)

- ethnic (religious) congruence and respectively issue-wise valence and issue-wise ideological congruence.

5.2.2 Subgroup analysis

We estimate heterogeneity according to the following variables:

- education: binary, defined as college and above vs everyone else

- gender

- age: binary variable for above/below ($\leq$) median age (approximate median age of the voting age population, calculated as 35) and categorical variable for age group (17-19, 20-24,25-29, 30-34,35-39, 40-44, above 45)

- factual political knowledge: above/below median, and continuous interaction, with the political knowledge score as defined above (estimated from IRT)

- conception of democracy: above/below median, and continuous interaction with the "conception of democracy" score as defined above

- Pancasila support: median split and discrete interaction (the answer to the Islam vs Pancasila stance of the respondent is three-levels)

- religiosity (based on the scale defined above): binary indicator for median split (above/below median religiosity) and continuous interaction with the religiosity score

- religion: Islam, Christian Protestant, Catholic, Hindu, Buddhist, other; alone and crossed with the median split of religiosity

- binary agree/disagree with the statement regarding non-Muslims allowed to become president, and categorical agreement disagreement (with five categories, including "unsure"), for the whole sample and restricting to Muslim respondents

- binary agree/disagree with the statement regarding whether equal opportunites should be provided to the non-Javanese to become president, and categorical agreement disagreement (with five categories, including "unsure"), for the whole sample and restricting to Javanese respondents

Heterogeneity is assessed via the ACIE using linear regression with interactions (and standard errors clustered at the respondent level) as suggested in Hainmueller, Hopkins, and Yamamoto 2014.

5.2.3 Effects of the randomized video treatment

We expect the informational (treatment) video to increase the weight assigned to valence characteristics, and to ideological/policy proximity, and reduce the weight assigned to identity characteristics. This general hypothesis is tested in two ways. We estimate ACIE where the features (both issue-wise and combining all issues) are interacted with an indicator for video treatment. As a robustness check, we also estimate AMCE as described above on the sample split by video treatment status (treatment vs control conditions 1 and 2, treated as one single control condition).

We also assess treatment effect heterogeneity based on the dimensions of heterogeneity listed above, one at a time.

5.3 Random utility estimation

We can model the utility respondent i receives from candidate j based on positional features on dimension k and valence issues m as

$$U_{ij} = -\left(\sum_k \gamma_{ik}(x_i^k - \tilde{x}_j^k)^2\right) + \delta_1 \mathbb{1}(E_i = \tilde{E}_j) + \delta_2 \mathbb{1}(R_i = \tilde{R}_j) + \sum_m \beta_{im} V_{jm} + \epsilon_{ij} \tag{1}$$

where γ_{ik} and β_{im} are respectively the weight that respondent i assigns to positional issue k or valence characteristic m; individual observed position x_i^k on issue k is compared to (randomly assigned) candidate position $\tilde{x}_j^k$; individual has ethnicity E_i and religion R_i, while the candidate j has ethnicity $\tilde{E}_j$ and religion $\tilde{R}_j$ and $\mathbb{1}$ is the indicator function; and ϵ_{ij} is a random shock to the utility of candidate j for individual i.

We observe, for a pair of candidates $j = 1, 2$, whether $U_{i1} \geq U_{i2}$ based on the stated choice across the pair of profiles. We assume that $\epsilon_{ij'}$ and $\epsilon_{ij''}$ are

independently distributed for a pair (j', j''). This yields an error term ϵ_i for the difference in utilities.

Weights are recovered by estimating a probit model where the outcome is an indicator variable for choosing candidate 1, equal to zero if the respondent choose candidate 2. Each observation is a respondent-pair combination (hence every respondent contributes 15 observations). We cluster standard errors at the respondent level.

The predictors are based on the difference between the characteristics of the candidates in the pair, where the various elements of the utility of candidate 2 for respondent i are subtracted from their counterparts for candidate 1.

We can take the difference between U_{i1} and U_{i2}, as in equation 1, for a pair of candidates (1,2). Define $\sigma_{ik} = -(x_i^k - \tilde{x}_1^k)^2 + (x_i^k - \tilde{x}_2^k)^2$, the difference in the distances between the respondent and the two candidates, where dependence on the candidates is suppressed to simplify notation. Similarly define $\xi_i^E = \mathbb{1}(E_i = \tilde{E}_1) - \mathbb{1}(E_i = \tilde{E}_2)$ and $\xi_i^R = \mathbb{1}(R_i = \tilde{R}_1) - \mathbb{1}(R_i = \tilde{R}_2)$

This yields

$$\Delta U_i = \sum_k \gamma_k \sigma_{ik} + \delta_1 \xi_i^E + \delta_2 \xi_i^R + \sum_m \beta_{im}(V_{1m} - V_{2m}) + \epsilon_i \qquad (2)$$

The outcome variable of the probit model is the sign of ΔU_i.

Given that in a given profile, only one of the positional and one of the valence dimensions are displayed, we impute zeros to the position of candidates on the issues (for valence and for ideology) that are not displayed in the profile. Due to randomization (with equal probability) of "high" and "low" positions, zero is the expected value. In other words, if for instance a profile does not mention the candidate position on military intervention (which, if displayed, can take two values, -1 and 1) we treat the candidate as having an intermediate position 0 on that issue. This corresponds to the expected value (taken over the randomization of profiles) of the position of candidates on military intervention.

In addition, we estimate also a simpler specification where a candidate is either

- ideologically close or distant

- high or low-valence

- congruent both in terms of ethnicity and religion, or different in at least one

regardless of the specific dimension that was displayed. Ultimately this is equivalent to estimating a probit model with three predictors: difference in ideological congruence for the two candidates; difference in valence of the two candidates; and difference in ethno-religious congruence between the two candidates.

Informational treatment effect on weights   We expect the treated (those who receive the informational video about accountability) to display

- lower weight assigned to identity characteristics

- higher weight assigned to valence characteristics

- higher weight assigned to ideological proximity

We test these hypothesis by split sample, estimating the model defined in equation 2 separately for the respondents who received the informational video and those who received the neutral video or no video (treated as one single condition). In robustness checks, we also estimate probit regressions with the interaction between the video treatment indicator and all the predictors.

Treatment effect heterogeneity   We assess treatment effect heterogeneity according to the same dimensions specified in the subsection 5.2.2 above.

5.4   Response time

We record the response time for each pair of profiles. Arguably, longer response time indicates a harder choice to make. We use this as an additional outcome variable. We model log response time as a function of the absolute differences in candidate traits, with dummies that take the value of one if the profiles differ on that feature, and zero otherwise. We estimate a linear regression with standard errors clustered by respondent, and including individual fixed effects. A positive coefficient on the difference in ethnicity, or a negative coefficient on the difference in valence, would respectively indicate that choosing between profiles that differ in ethnicity is harder and that choosing between profiles that differ in terms of valence is easier. Given the full randomization of the candidate traits, all these have a causal interpretation. Individual fixed effects allow for removal of variation across individuals in reading speed, cognitive ability, or impulsiveness.

5.5   Stated importance of dimensions

At the end of the conjoint exercise (after choosing between the 15 pair of profiles) we ask to each respondent to rank the relative importance they assign to one dimension (chosen at random) compared to education of the candidate (which is used as the reference point). Respondents answer on a three-point scale: the shown dimension is more, less, or equally important than education of the candidate. The dimensions shown are:

- Share the same religion/ethnicity

- Corruption record / Integrity

- Financial capacity / Being Rich

- Programs, policy preferences

From the answers, an average ranking (for any sufficiently large subset of respondents) of all the issues can be backed out. We perform (descriptive) subgroup analysis based on the dimensions of heterogeneity listed in subsection 5.2.2, one at a time, splitting the sample. In addition, we estimate, by split sample, the effect of the informational treatment video on stated importance of the issues. Finally we perform treatment heterogeneity analysis by the dimensions in subsection 5.2.2 by split sample of combinations of the (binary) treatment indicator and binary respondent characteristics, one at a time.

6 Video effects on turnout

In the baseline (hence pre-treatment with respect to the video intervention), we ask a question regarding intention to turn out in the general election. In the post-election wave, we ask respondents whether they turned out in the general election. We analyze the effect of the treatment, considering it, in this case, a three-arm design: accountability video, neutral election procedure video, and no video.

We estimate these models:

- probit regression for turnout in the general election, including demographic characteristics and indicator variables for the two video treatments

- probit regression for turnout in the general election, including demographic characteristics and indicator variables for the two video treatments, and interactions between treatment indicators and: education (binary); indicators for religion; indicators for age groups

7 List Experiment

The list experiment aims at answering several questions:

1. Does vote buying, defined as the exchange of individual voter's vote for a small gift, exist? If so, what is the extent of this activity?

2. Is this strategy effective to gather votes?

    - Do voters accept the gift from the party machine?
    - Do voters uphold their end of the transaction by actually voting for the candidates giving them the gift?
    - Does providing information about the accountability concept of democracy reduce acceptance and efficacy of vote buying?

The list experiment contains two types of lists, and each respondent answers both:

1. Party activities list

2. Individual voter activities list

The party activities list contains four activities that may be performed by party cadres/volunteers during the election campaign, one of which (threats) is very unlikely. There is one sensitive item, hence the longer list contains a total of 5 activities. The items below are included in the party activities list:

- Put up posters or signs in your neighborhood

- Organize a concert in your areas

- Invite people to have a party in a restaurant or other places

- Threatened you to vote for them

- A sensitive item, only deployed in the treatment group list

    - Giving you a gift or money

The voter activities list contains four ordinary activity items and three sensitive items, one randomized in the pre-election (wave 2) survey –called arm 0 below– and two that that are randomized in different groups in the post-election (wave 3 survey) –called arm 1 and arm 2 below. The list of activities includes:

- Attending a campaign rally

- Buying a new car

- Attending a funeral

- Going to work (asked in the pre-election survey) / Going to the polling station to vote (asked in the post-election survey)

- A sensitive item. Respondents selected to receive the long list receive only one out of the three alternative below. In the pre-election survey (second wave) there is one single arm. In the post-election (third wave) survey there are two arms.

    - Accepting money/gift from campaign activists [pre-election – arm 0]
    - Accepting money/gift from campaign activist and vote for candidate that give you the money/gift [post-election – arm 1]
    - Accepting money/gift from campaign activist and do not vote for candidates that give you the money/gift [post-election – arm 2]

17

7.1 Detecting the existence and the magnitude of vote buying

The magnitude of vote buying is estimated as the difference in the average number of items mentioned in the "control" group (respondents with the short list) and the average number of items mentioned in the "treatment" group (respondents with the list including the sensitive item) from the party activity list. We can calculate the prevalence of party offers from the party activities list experiment. From the vote activity list we estimate the prevalence of acceptance.

Party offers are given by the difference in means between the number of items in the control (short list) and treatment (list with sensitive item) group in the question about party activities, pooling wave 2 and wave 3.

Voter acceptance of the "gift" is given by the difference in means between control (short list) and treatment (list with sensitive item, regardless of arm) group in the voter activities question, pooling wave two and wave three.

We can also estimate which fraction of the offers made are accepted, from the double difference between prevalence of party offers and prevalence of acceptance. The estimate is based on the difference between two counts, ie. the estimate of the prevalence of party offers (itself from the difference, in the party activities questions, between "treated" and "controls") and the estimate of voter acceptance obtained in a similar fashion.

7.1.1 Heterogeneity in vote buying

We estimate the maximum likelihood model with covariates of Blair and Imai 2012, including the variables used for heterogeneity analysis in the conjoint exercise and listed in subsection 5.2.2. We estimate separate models for party offers (based on the party activities list) estimated on the full sample; for acceptance (combining arm 0, arm 1, and arm 2 into a single binary indicator), estimated on the full sample; and as a two-arm list experiment (arm 1 vs arm 2, i.e., acceptance with or without compliance) estimated on the post-election (wave 3) sample.

7.2 Estimating the effectiveness and efficacy of vote buying

There are three measures of effectiveness we plan to calculate, i.e. the output of vote buying (proportion of offers that are accepted, discussed above) and the outcome of vote buying (proportion of offers that are accepted and are complied with by the voter, and proportion of offers that are accepted but not complied with). These can be estimated as the difference between the estimated means. The means from arm 1 and arm 2 clearly come only from wave 3 (post-election).

7.3 Informational treatment effect

We want to estimate whether being exposed to the informational (vs. the neutral or no) video decreases: acceptance of the vote buying offer; and fulfillment of the obligation to vote as the machine requested when making the offer.

To estimate the effect of the informational treatment on vote buying and selling, we estimate models where the video treatment indicator (dummy) is a covariate, using in all cases the ML model of Blair and Imai 2012 as implemented in the `ictreg` function in the `list` package in the R environment. We also split the two "control" conditions, hence creating one indicator variable equal to one for the respondents who received informational video treatment and one indicator variable equal to one for the neutral election procedure video.

As a robustness check, we also estimate the equivalent (but less efficient) counterparts in linear regression form (with heteroskedasticity-robust standard errors):

- regression of listed number of voter activities on indicator for short/long list, indicator for informational video, and their interaction (this is estimated on the full sample, coding the indicator equal to 1 for arm 0, 1, and 2)

- regression of listed number of voter activities on indicator for arm 1, indicator for arm 2, indicator for having received the informational video, and interaction between indicator for informational video and respectively indicators for arm 1 and for arm 2 (this is estimated on wave 3 respondents)

- regression of listed number of voter activities on the indicators for arm 0, arm 1, and arm 2, the indicator for having received the informational video, and the interaction between the indicator for informational video and respectively indicators for arm 0, for arm 1, and for arm 2 (this is estimated on the full sample).

Negative statistically significant coefficients on the interaction between the informational video dummy and the short vs long list dummies indicate a causal effect of the informational treatment:

- on acceptance of gifts from the machine, for arm 0

- on acceptance and compliance, for arm 1

- on acceptance with non-compliance, for arm 2.

In addition we estimate the versions of the regression models discussed above but splitting the two types of control condition. Namely

- regression of listed number of voter activities on indicator for short/long list, indicator for informational video, indicator for neutral video, and the interaction between the two video dummies and long/short list(this is estimated on the full sample, coding the indicator equal to 1 for arm 0, 1, and 2)

- regression of listed number of voter activities on indicator for arm 1, indicator for arm 2, indicator for having received the informational video, indicator for neutral video, and interaction between indicators for informational video and neutral video and respectively indicators for arm 1 and for arm 2 (this is estimated on wave 3 respondents)

- regression of listed number of voter activities on the indicators for arm 0, arm 1, and arm 2, the indicator for having received the informational video, indicator for having received the neutral video, and the interaction between the indicators for informational video and neutral video and respectively indicators for arm 0, for arm 1, and for arm 2 (this is estimated on the full sample).

8   Dynastic Politics Analysis

In this piece of analysis we attempt to estimate whether an incumbent president can award an electoral advantage to their progeny. The set-up of candidates running in the Indonesia presidential election 2024 provides an opportunity to explore this question, as the son of the incumbent president is running as a vice-president, in a different party ticket than the incumbent president party, and is paired with the challenger of the incumbent president in the 2019 election. Hence, voting for the incumbent's son is equivalent to voting for the challenger ticket; voting for the incumbent party ticket is equivalent to *not* voting for the incumbent's son.

We exploit a number of questions included in the baseline, pre-election and post-election survey:

- In the baseline, we ask a number of questions about the level of satisfaction about the incumbent administration and who is responsible for it. All these are answered on a 4-point scale (Strongly agree, Agree, Disagree, Strongly disagree). We treat the "unsure" option as missing.

  – "How satisfied are you with the achievements of the current administration".

  – For those who answered: *Satisfied* or *Very satisfied*, we provide the following question: "Whom do you think contributed the most to the achievements of the current administration?"

  – For those answered: *Dissatisfied* or *Very dissatisfied*, we provide the following question: "Whom do you think should be held responsible for the outcome of the current administration?"

- In the pre-election survey, we ask for voting intention in both the presidential and the legislative races (in the latter, we ask the party of the candidate the respondent intends to vote)

- In the post-election survey, we ask for the actual candidate and party the respondent has voted in the general election.

We estimate the effect of dynastic politics by comparing the intention and the actual votes for the presidential ticket that includes the son of the incumbent president (dynastic ticket). We hypothesize that higher satisfaction with the incumbent administration is positively correlated to the higher vote for the dynastic ticket, controlling for other observable characteristics. Moreover, this correlation is expected to be especially higher among those who attribute the success/failure of the current administration to the incumbent president himself, and not to other agents, e.g. the main incumbent party, the cabinet, or the coalition party.

The model we estimate has the general form of a probit regression where the outcome variables are respectively pre-election vote intention (for the ticket with the incumbent's son) in the presidential race and post-election self-reported vote choice (for the ticket with the incumbent's son) in the presidential race. The predictors are related to evaluation and attribution of incumbent performance.

We define *government performance evaluation and attribution* as a categorical variable that takes the following values:

- very positive and attributed to the president

- very positive and attributed to other political actors

- positive and attributed to the president

- positive and attributed to other political actors

- very negative and attributed to the president

- very negative and attributed to other political actors

- negative and attributed to the president

- negative and attributed to other political actors

This can then be dummied out as a categorical predictor.

We estimate four versions of the model: a "short" one with only the "government performance evaluation and attribution" dummies; a "long" one which also includes dummies for legislative voting (intention at baseline and actual in the post-election wave), and demographic controls; and three interactive models where the evaluation and attribution dummies are interacted respectively with education (binary, college vs. non-college), with age group, and with political information (score of factual information about politics as described above, treated as continuous). We also estimate a simpler model (with demographic controls) in which performance evaluation (positive vs negative) and attribution (president vs. others) are included as dummies and interacted with each other.

Evidence of dynastic transmission is provided if attribution of positive (negative) performance to the president is associated with more (less) support for the challenger ticket.

References

Blair, Graeme, and Kosuke Imai. 2012. "Statistical analysis of list experiments."
    Political Analysis 20 (1): 47–77.

Hainmueller, Jens, Daniel J Hopkins, and Teppei Yamamoto. 2014. "Causal
    inference in conjoint analysis: Understanding multidimensional choices via
    stated preference experiments." *Political analysis* 22 (1): 1–30.

9 Appendix

9.1 Video scripts

Duration: 75 seconds

9.1.1 Video 1 - treatment

EN:

Imagine you inherit a convenient store from your grandparents. However, you have your own job and do not have the expertise to manage it.

Therefore, you run a search for the store manager who will act on your behalf to manage it, so you need to find someone that you can trust to act for your best interest.

Due to the power you give him, he may be tempted to do things that benefit himself but costly for you. Therefore, you need to check on him from time to time, making sure that he manages your store with competence, dignity, and integrity. If you think he does not fulfill your expectations, or is betraying your trust, you have all the right to fire him and elect someone else to do the job.

This is a metaphor of an election (PEMILU) in democracy. In an election, we choose a political candidate to run the country on our behalf.

Just like the case of the store manager, political leaders will be tempted to abuse their power to benefit themselves at our cost.

Therefore, in an election, we can "punish" them by electing someone else when they were lying, cheating, or under-performing.

So, are you ready to vote in the coming election?

ID

Bayangkan kamu mewarisi sebuah toko serba ada dari kakek kamu. Namun, kamu memiliki pekerjaan sendiri dan tidak memiliki keahlian untuk mengelolanya.

Karena itu, kamu mencari seorang manajer toko yang bertindak atas nama kamu untuk mengelolanya. Orang ini harus bisa kamu percayai akan bertindak sebaik-baiknya demi kepentinganmu.

Karena kepercayaan yang diberikan kepadanya, dia mungkin akan tergoda untuk melakukan hal-hal yang menguntungkan dirinya sendiri tetapi merugikan kamu; sehingga kamu perlu mengawasinya dari waktu ke waktu, memastikan dia mengelola tokomu dengan kompetensi, martabat, dan integritas. Jika dia tidak memenuhi harapanmu, atau mengkhianati kepercayaanmu, kamu punya segala hak untuk memecatnya dan memilih orang lain untuk melakukan pekerjaan itu

Ini adalah perumpamaan PEMILU dalam demokrasi. Dalam pemilu, kita memilih CALEG untuk menjalankan negara atas nama kita. Seperti kasus manajer toko di atas, pemimpin politik sering tergoda untuk menyalahgunakan kekuasaan demi keuntungan sendiri; merugikan kita tanpa kita sadari. Dengan

PEMILU, kita bisa "menghukum" mereka dengan memilih orang lain jika kita dapati mereka berbohong, curang, atau berkinerja buruk.

Jadi, apakah kamu siap untuk memilih pada pemilu mendatang?

9.1.2 Video 2 – control

EN:

Election (PEMILU) is around the corner, but what do we know about election? PEMILU stands for Pemilihan Umum. We held our first democratic election in 1995 and had experience twelfe general election since. An interesting fact for this year's election is that we will held the national as well as the regional level election ALL at once.

Who are eligible to vote? If you are 17 years old or older, or you have had been married before, then you are eligible to vote. Obviously, you need to have a valid national ID (KTP) to certify that you are indeed the citizen of the Republic of Indonesia. Moreover, a person who are eligible to vote has be mentally sound, and they cannot be a member of national police force (POLRI) or a member of national army (TNI).

We should check whether our names are registered as a valid voter (Daftar Pemilih Tetap - DPT) at the KPU website. When you find yourself registered, then you should also check if your information details are accurate. Make sure that your residence is in line with the location where you will vote (Dapil). If you detect any discrepancy, then you should report this to KPU and request for data change. You will then be added to the additional voter list (Daftar Pemilih Tambahan).

So, are you ready to vote in the coming election?

ID:

PEMILU sudah dekat, tapi apa yang kita ketahui tentang pemilu? PEMILU adalah singkatan dari Pemilihan Umum. Kita menyelenggarakan pemilu demokratis pertama kali pada tahun 1995 dan telah menjalani dua belas kali semenjak itu. Fakta menarik pada pemilu tahun ini adalah kita akan menyelenggarakan pemilu tingkat nasional dan tingkat daerah secara serentak.

Siapa yang berhak memilih? Jika kamu sudah berusia 17 tahun, atau pernah menikah sebelumnya, maka kamu berhak memilih. Tentunya kamu harus memiliki KTP yang masih berlaku untuk membuktikan bahwa kamu memang warga negara Republik Indonesia. Selain itu, orang yang berhak memilih harus sehat secara mental, dan anggota POLRI atau TNI tidak boleh ikut PEMILU.

Kita bisa mengecek apakah nama kita terdaftar sebagai pemilih sah dalam Daftar Pemilih Tetap (DPT) di website KPU. Ketika terdaftar, kamu juga sebaiknya memeriksa apakah rincian informasi kamu akurat. Pastikan tempat tinggal kamu sesuai dengan tempat kamu akan memilih (Dapil). Jika kamu

menemukan kesalahan, kamu bisa melaporkannya ke KPU dan meminta pe-
rubahan data. Kemudian, kamu akan ditambahkan ke daftar pemilih tambahan
(DPT).

Jadi, apakah kamu siap untuk memilih pada pemilu mendatang?

9.2 Attributes and Level for conjoint experiment/candidate profiles

Attribute	Level and Distribution	Distribution (%)
Age	23 - 30	10
	30 - 55	70
	56 - 69	15
	70 - 75	5
Gender	Male	60
	Female	40
Ethnicity	Titular-ethnic	50
	Native non-titular	40
	Chinese non-titular	10
Educational Background	Primary school/Junior high school	5
	Senior high school	50
	D1/D2/D3	5
	D4/S1	30
	S2/S3	10
Religion	Islam	70
	Christian	10
	Catholic	10
	Hindu	5
	Buddha	5
Marital Status	Married	75
	Not married	20
	Divorced/Widowed	5
Occupational Background	Worker	50
	Entrepreneur	10
	Domestic	10
	Teacher	5
	Student	5
	Retiree	5
	Politician	5
	Others	10

9.3   Ethnicity By Province

Province	Ethnicity	Status	Probability (%)
Aceh	Aceh	Titular ethnic	50
Aceh	Javanese	Native non-titular	10
Aceh	Batak	Native non-titular	10
Aceh	Minang	Native non-titular	10
Aceh	Melayu	Native non-titular	10
Aceh	Chinese	Chinese non-titular	10
North Sumatera	Batak	Titular ethnic	50
North Sumatera	Javanese	Native non-titular	10
North Sumatera	Nias	Native non-titular	10
North Sumatera	Melayu	Native non-titular	10
North Sumatera	Minang	Native non-titular	10
North Sumatera	Chinese	Chinese non-titular	10
West Sumatera	Minang	Titular ethnic	50
West Sumatera	Batak	Native non-titular	10
West Sumatera	Javanese	Native non-titular	10
West Sumatera	Sumatera	Native non-titular	10
West Sumatera	Melayu	Native non-titular	10
West Sumatera	Chinese	Chinese non-titular	10
Riau	Melayu	Titular ethnic	50
Riau	Javanese	Native non-titular	10
Riau	Batak	Native non-titular	10
Riau	Minang	Native non-titular	10
Riau	Banjar	Native non-titular	10
Riau	Chinese	Chinese non-titular	10
Jambi	Jambi	Titular ethnic	50
Jambi	Javanese	Native non-titular	10
Jambi	Melayu	Native non-titular	10
Jambi	Minang	Native non-titular	10
Jambi	Batak	Native non-titular	10
Jambi	Chinese	Chinese non-titular	10
South Sumatera	South Sumatera	Titular ethnic	50
South Sumatera	Javanese	Native non-titular	10
South Sumatera	Melayu	Native non-titular	10
South Sumatera	Sunda	Native non-titular	10
South Sumatera	Sumatera	Native non-titular	10
South Sumatera	Chinese	Chinese non-titular	10

Province	Ethnicity	Status	Probability (%)
Bengkulu	Sumatera	Titular ethnic	50
Bengkulu	Javanese	Native non-titular	10
Bengkulu	South Sumatera	Native non-titular	10
Bengkulu	Minang	Native non-titular	10
Bengkulu	Sunda	Native non-titular	10
Bengkulu	Chinese	Chinese non-titular	10
Lampung	Javanese	Titular ethnic	50
Lampung	Lampung	Native non-titular	10
Lampung	Sunda	Native non-titular	10
Lampung	South Sumatera	Native non-titular	10
Lampung	Banten	Native non-titular	10
Lampung	Chinese	Chinese non-titular	10
Riau Islands	Melayu	Titular ethnic	50
Riau Islands	Javanese	Native non-titular	10
Riau Islands	Batak	Native non-titular	10
Riau Islands	Minang	Native non-titular	10
Riau Islands	Sunda	Native non-titular	10
Riau Islands	Chinese	Chinese non-titular	10
Bangka Belitung	Sumatera	Titular ethnic	50
Bangka Belitung	Javanese	Native non-titular	10
Bangka Belitung	South Sumatera	Native non-titular	10
Bangka Belitung	Bugis	Native non-titular	10
Bangka Belitung	Sunda	Native non-titular	10
Bangka Belitung	Chinese	Chinese non-titular	10
DKI Jakarta	Javanese	Titular ethnic	50
DKI Jakarta	Betawi	Native non-titular	10
DKI Jakarta	Sunda	Native non-titular	10
DKI Jakarta	Batak	Native non-titular	10
DKI Jakarta	Minang	Native non-titular	10
DKI Jakarta	Chinese	Chinese non-titular	10
West Java	Sunda	Titular ethnic	50
West Java	Javanese	Native non-titular	10
West Java	Betawi	Native non-titular	10
West Java	Cirebon	Native non-titular	10
West Java	Batak	Native non-titular	10
West Java	Chinese	Chinese non-titular	10
Central Java	Javanese	Titular ethnic	50
Central Java	Sunda	Native non-titular	10
Central Java	Batak	Native non-titular	10
Central Java	Asing	Native non-titular	10
Central Java	Madura	Native non-titular	10
Central Java	Chinese	Chinese non-titular	10

Province	Ethnicity	Status	Probability (%)
D I Yogyakarta	Javanese	Titular ethnic	50
D I Yogyakarta	Sunda	Native non-titular	10
D I Yogyakarta	Melayu	Native non-titular	10
D I Yogyakarta	Batak	Native non-titular	10
D I Yogyakarta	Madura	Native non-titular	10
D I Yogyakarta	Chinese	Chinese non-titular	10
East Java	Javanese	Titular ethnic	50
East Java	Madura	Native non-titular	10
East Java	Batak	Native non-titular	10
East Java	Sunda	Native non-titular	10
East Java	Sulawesi	Native non-titular	10
East Java	Chinese	Chinese non-titular	10
Banten	Banten	Titular ethnic	50
Banten	Sunda	Native non-titular	10
Banten	Javanese	Native non-titular	10
Banten	Betawi	Native non-titular	10
Banten	Batak	Native non-titular	10
Banten	Chinese	Chinese non-titular	10
Bali	Bali	Titular ethnic	50
Bali	Javanese	Native non-titular	10
Bali	Madura	Native non-titular	10
Bali	Melayu	Native non-titular	10
Bali	Sasak	Native non-titular	10
Bali	Chinese	Chinese non-titular	10
NTB	Sasak	Titular ethnic	50
NTB	NTB	Native non-titular	10
NTB	Bali	Native non-titular	10
NTB	Javanese	Native non-titular	10
NTB	Sulawesi	Native non-titular	10
NTB	Chinese	Chinese non-titular	10
NTT	NTT	Titular ethnic	50
NTT	Kalimantan	Native non-titular	10
NTT	Javanese	Native non-titular	10
NTT	Sulawesi	Native non-titular	10
NTT	Bugis	Native non-titular	10
NTT	Chinese	Chinese non-titular	10
West Kalimantan	Dayak	Titular ethnic	50
West Kalimantan	Melayu	Native non-titular	10
West Kalimantan	Javanese	Native non-titular	10
West Kalimantan	Madura	Native non-titular	10
West Kalimantan	Bugis	Native non-titular	10
West Kalimantan	Chinese	Chinese non-titular	10

Province	Ethnicity	Status	Probability (%)
Central Kalimantan	Kalimantan	Titular ethnic	50
Central Kalimantan	Javanese	Native non-titular	10
Central Kalimantan	Banjar	Native non-titular	10
Central Kalimantan	Dayak	Native non-titular	10
Central Kalimantan	Melayu	Native non-titular	10
Central Kalimantan	Chinese	Chinese non-titular	10
South Kalimantan	Banjar	Titular ethnic	50
South Kalimantan	Javanese	Native non-titular	10
South Kalimantan	Bugis	Native non-titular	10
South Kalimantan	Dayak	Native non-titular	10
South Kalimantan	Madura	Native non-titular	10
South Kalimantan	Chinese	Chinese non-titular	10
East Kalimantan	Javanese	Titular ethnic	50
East Kalimantan	Bugis	Native non-titular	10
East Kalimantan	Kalimantan	Native non-titular	10
East Kalimantan	Banjar	Native non-titular	10
East Kalimantan	Sulawesi	Native non-titular	10
East Kalimantan	Chinese	Chinese non-titular	10
North Kalimantan	Javanese	Titular ethnic	50
North Kalimantan	Bugis	Native non-titular	10
North Kalimantan	Kalimantan	Native non-titular	10
North Kalimantan	Banjar	Native non-titular	10
North Kalimantan	Sulawesi	Native non-titular	10
North Kalimantan	Chinese	Chinese non-titular	10
North Sulawesi	Minahasa	Titular ethnic	50
North Sulawesi	Sulawesi	Native non-titular	10
North Sulawesi	Gorontalo	Native non-titular	10
North Sulawesi	Javanese	Native non-titular	10
North Sulawesi	Maluku	Native non-titular	10
North Sulawesi	Chinese	Chinese non-titular	10
Central Sulawesi	Sulawesi	Titular ethnic	50
Central Sulawesi	Bugis	Native non-titular	10
Central Sulawesi	Javanese	Native non-titular	10
Central Sulawesi	Bali	Native non-titular	10
Central Sulawesi	Gorontalo	Native non-titular	10
Central Sulawesi	Chinese	Chinese non-titular	10

Province	Ethnicity	Status	Probability (%)
South Sulawesi	Bugis	Titular ethnic	50
South Sulawesi	Makassar	Native non-titular	10
South Sulawesi	Sulawesi	Native non-titular	10
South Sulawesi	Javanese	Native non-titular	10
South Sulawesi	NTT	Native non-titular	10
South Sulawesi	Chinese	Chinese non-titular	10
Southeast Sulawesi	Sulawesi	Titular ethnic	50
Southeast Sulawesi	Bugis	Native non-titular	10
Southeast Sulawesi	Javanese	Native non-titular	10
Southeast Sulawesi	Makassar	Native non-titular	10
Southeast Sulawesi	Bali	Native non-titular	10
Southeast Sulawesi	Chinese	Chinese non-titular	10
Gorontalo	Gorontalo	Titular ethnic	50
Gorontalo	Sulawesi	Native non-titular	10
Gorontalo	Javanese	Native non-titular	10
Gorontalo	Minahasa	Native non-titular	10
Gorontalo	Bugis	Native non-titular	10
Gorontalo	Chinese	Chinese non-titular	10
West Sulawesi	Sulawesi	Titular ethnic	50
West Sulawesi	Bugis	Native non-titular	10
West Sulawesi	Javanese	Native non-titular	10
West Sulawesi	Makassar	Native non-titular	10
West Sulawesi	Bali	Native non-titular	10
West Sulawesi	Chinese	Chinese non-titular	10
Maluku	Maluku	Titular ethnic	50
Maluku	Sulawesi	Native non-titular	10
Maluku	Javanese	Native non-titular	10
Maluku	Bugis	Native non-titular	10
Maluku	NTT	Native non-titular	10
Maluku	Chinese	Chinese non-titular	10
North Maluku	Maluku	Titular ethnic	50
North Maluku	Sulawesi	Native non-titular	10
North Maluku	Javanese	Native non-titular	10
North Maluku	Bugis	Native non-titular	10
North Maluku	Minahasa	Native non-titular	10
North Maluku	Chinese	Chinese non-titular	10

Province	Ethnicity	Status	Probability (%)
West Papua	Papua	Titular ethnic	50
West Papua	Javanese	Native non-titular	10
West Papua	Maluku	Native non-titular	10
West Papua	Sulawesi	Native non-titular	10
West Papua	Bugis	Native non-titular	10
West Papua	Chinese	Chinese non-titular	10
Papua	Papua	Titular ethnic	50
Papua	Javanese	Native non-titular	10
Papua	Sulawesi	Native non-titular	10
Papua	Bugis	Native non-titular	10
Papua	Maluku	Native non-titular	10
Papua	Chinese	Chinese non-titular	10
Central Papua	Papua	Titular ethnic	50
Central Papua	Javanese	Native non-titular	10
Central Papua	Sulawesi	Native non-titular	10
Central Papua	Bugis	Native non-titular	10
Central Papua	Maluku	Native non-titular	10
Central Papua	Chinese	Chinese non-titular	10
South Papua	Papua	Titular ethnic	50
South Papua	Javanese	Native non-titular	10
South Papua	Sulawesi	Native non-titular	10
South Papua	Bugis	Native non-titular	10
South Papua	Maluku	Native non-titular	10
South Papua	Chinese	Chinese non-titular	10
Highland Papua	Papua	Titular ethnic	50
Highland Papua	Javanese	Native non-titular	10
Highland Papua	Sulawesi	Native non-titular	10
Highland Papua	Bugis	Native non-titular	10
Highland Papua	Maluku	Native non-titular	10
Highland Papua	Chinese	Chinese non-titular	10
Southwest Papua	Papua	Titular ethnic	50
Southwest Papua	Javanese	Native non-titular	10
Southwest Papua	Sulawesi	Native non-titular	10
Southwest Papua	Bugis	Native non-titular	10
Southwest Papua	Maluku	Native non-titular	10
Southwest Papua	Chinese	Chinese non-titular	10