The Australian National University

Centre for Economic Policy Research

DISCUSSION PAPER

The Social and Economic Determinants of Voting 'Yes' in South Australia's Federation Referenda

William Coleman

DISCUSSION PAPER NO. DP700

December 2017

ISSN: 1442-8636
ISBN: 978-1-921693-84-7

The Social and Economic Determinants of Voting 'Yes' in South Australia's Federation Referenda.

William Coleman,

Research School of Economics, ANU

December 2017

The paper uses data from South Australia's census of 1901 to throw light of the attributes of electors and electorates that encouraged or discouraged voting Yes in the 1898 and 1899 South Australian federation referenda. It concludes that British-birth and an industrial occupation contributed powerfully to voting No. It additionally concludes that in the 1899 referendum industrial occupation disappeared as a discouragement to voting Yes.

- The author gratefully acknowledges the stimulus of the conversation and kindred research of Tim Hatton, Andrew Sudol and Rohan Alexander.

I Introductory

The Question

Between 1898 and 1900 ten referenda were held across Australia on the draft Constitution Bill that proposed to federate New South Wales, Victoria, Queensland, South Australia, Western Australia and Tasmania. All resulted in Yes majorities.

This paper investigates what factors inclined South Australian voters to vote Yes or No. The question is historical, but the method is social scientific. The paper uses statistical methods to infer the dependence of South Australian voting results on socioeconomic variables reported by a census.

Some of the results are striking: British birth strongly encouraged voting No in both referenda, while being in an industrial occupation even more strongly encouraged No in 1898, but not 1899.

Previous literature

The first quantitative explorations of the Federation referendums were by Robin Parker (1949), and, in critique, Geoffrey Blainey (1950). Norris (1969) belongs to the same literature, and is specifically directed to South Australia. These papers tend to use either the simplest sample statistics, or a 'case study' method, where particular districts are selected as corroborating a thesis, or serving as a counter example to another. The strength of this method is that great detail may be brought to bear on a given 'case'. But it shares the weakness of any method that selects the evidence it makes inferences from, and it cannot cope with chance concurrence of votes and variables.

The first use of sophisticated statistical method on the federation referenda was undertaken by Rhodes (1988). This study spans the whole of Australia, and does several

impressive things: it investigates, for example, the impact in each electorate of the local newspaper's stance on Federation. But it neglects the bulk of the census data available in South Australia.[1]

The data setting

This paper is concerned with South Australia's vote.

On 4 June 1898 a referendum was held in South Australia on the draft federal constitution prepared by the Constitutional Convention of 1897-98. There were 136,387 registered voters, out of a population of around 360,000, who cast 53,120 valid votes. The Yes votes amounted to 67.4 percent of total votes, and No votes 32.6 percent. Data on voting is available by 27 electoral districts (see Rhodes 2002). Excluding the Northern Territory, the highest Yes vote was in the electorate of Light (80 percent); the lowest in West Torrens (35 percent). Of 27 electorates only 2 recorded a majority No vote, West Adelaide, and West Torrens.

On 29 April 1899 a second referendum was held on a slightly revised version of federal constitution. On this occasion the Yes votes amounted to 79.5 percent of total votes, and No votes just 20.5 percent. Of 27 electorates none recorded a majority No vote. Indeed, the lowest Yes vote was 67 percent, in Torrens West. Evidently, there was a substantial shift to Yes in 1899, and in pooling the observations of the two referenda into one data set it will be necessary to allow for that shift.

On March 31 1901 a census was taken in South Australia. A considerable part of the results are reported in terms of South Australia's 27 electoral districts. (*Historical Census and Colonial Data Archive)*

[1] A qualitative precis of this study is reported in Rhodes (2002), without any detail.

Thus data is available on voting by electorate; and data is also available for several demographic characteristics by electorate. These characteristics are,

Sex
Conjugal Condition
Country of Birth
Age
Educational Attainment (ie literacy)
Occupation
Religion

Regrettably, the census provides no data on income or wealth.

The census data is not cross tabulated, save (sometimes) by Sex and Age. Thus we know the number of adults of the electorate of Encounter Bay who worked in primary industry; *and* we know but number of adults of the electorate of Encounter Bay who were Lutherans; but we do not now the number of Lutherans in Encounter Bay who worked in primary industry. Given the secrecy of the ballot, there is, of course, no cross tabulation of the Census data with voting results.

One of the 27 electorates deserves special consideration: the Northern Territory. It is abnormal; huge in area, in 1898 it recorded just 167 votes, out of a South Australian total of 53,000. The enrolled electors were overwhelmingly single, male, young, miners, with an unusually high illiteracy rate. Further, the issue of Federation was understood solely in terms of a much hoped-for separation from South Australia. And the electorate voted 95 percent Yes. It is well known that outlier observations can exert an impact on statistical inferences out of all proportion to their frequency. The paper drops the Northern Territory for its analysis.

Specification

The present study's object of explanation is the proportion of votes in electorate i that are Yes votes, S_i, a continuous variable ranging between zero and one.[2]

The determinants of S_i are a range of variables which are also proportions in the given electorate, X_j^i, where j indexes the type of determinant.[3] Because no single observation of any variable –either dependent or independent - can be construed as any individual's 'decision' there is no grounds for modelling the data as produced by a single individual's decision making process. We simply take as our starting point the proposition that an increase in the proportion of persons of a certain attribute may impact on the proportion of Yes votes in that electorate, and seek to map that relationship.

Dropping the i subscript/superscripts for the sake of concision, the relationship may be represented,

$$S = f(X_j)$$

dS/dXj is the impact on the percentage Yes vote of increasing the share of the adult population that are 'j' by one percent. Thus if j = over 65s and dS/dX = 0.30, then an increase in the over 65s from (say) 15 to 16 percent of the adult population would increase the Yes vote by 0.3 of one percent. We denote dS/dX as 'impact'.

[2] An alternative object of explanation would be the proportion of *registered* voters that cast a Yes vote.

[3] In most cases the present study uses proportions of the number of persons aged 21 or over ('adult') in the given electorate. This is sometimes varied in a minor way according to the dictates of convenience.

The functional form of f(X) is, of course, unknown. The simplest functional form is also the most intuitive and natural one: a linear one,

$$S = \beta X_j + u$$

This linear specification implies 'impact' is parametrical.

$$\frac{\partial S}{\partial X_j} = \beta$$

This specification does have the merit of lending itself to a simple rationalisation of measured 'impacts' in term of probabilities. The rationalisation is that persons of a given attribute (eg being married) have a certain, exogenous probability of voting Yes, which is independent of the frequency of any other attribute in the electorate. Thus if Xj is the proportion married persons, then α is the probability of an *un*married person voting Yes, and $\alpha+\beta$ is the probability of married person voting Yes. We have a simple and comprehensible 'additive vision' of the proportion voting Yes in an electorate, that is entirely congruent with the methodological individualism of economics: just as the temperature of a bucket of water is the average of the temperatures of the various cups of water that are poured into it, so the proportion voting Yes an electorate is an average of the propensities to vote Yes of the various 'demographics' composing it. Given its intelligibility and simplicity it is not surprising that the linear/additive specification remains in use in academic studies of referendum results (Urbatsch 2013).

But it is well-known the linear/additive modelling of proportions has two deficiencies.

Firstly, the linear/additive vision may falsely characterise the causal process. The key presumption of the linear/additive model is that the probability of (say) a young person voting Yes is parametrical. That is, the probability is unaffected by else; including the proportion of the electorate being young. But what if, for example, the propensity of a

young person to vote Yes tends to change as the number of elderly persons in their electorate changes? Perhaps the young are influenced by the views of the old, or perhaps they are allergic to them. The family is one obvious context whereby the attributes of one person may colour, or 'infect', the positions of another. In the same vein a marital partner who is some 'z' (a Lutheran, an immigrant etc) might increase the tendency of their marital partner to vote in the way a Lutheran etc tends to. Other contexts for 'attitudinal contagion' are the workplace, the church, the reading room, the public house. If the positions of the X group on Federation are 'contagious', then we will not have an 'additive model'.

Secondly, the linear/additive modelling of proportions does not supply valid statistical inferences. Standard methods of estimation of the linear/additive model neglect the necessity that

$$0 \leq S \leq 1$$

This restriction implies that, whatever the coefficients on the included variables, the disturbance term, u, cannot be normally distributed, as a normal distribution would (inadmissibly) make it possible for actual S to exceed 1, or be less than zero. Thus any assumption of the normality of u in small samples is invalidated, and so the validity of the standard errors of OLS estimates in small samples is compromised.

A 'logistic' specification of the dependence of S on X_j simultaneously avoids additive causation and the statistical invalidity of the linear/additive model,

$$S = \frac{1}{1 + e^{-\alpha - \beta X_j - u}}$$

Evidently, this specification always ensures actual S is always between 1 and zero,

Secondly, the 'impact' can be shown to be,

$$\frac{\partial S}{\partial X_j} = \beta[1 - S]S$$

The impact is, evidently, not parametrical. Thus the logistic specification does not impute unconditional, parametrical probabilities of voting Yes to individual members of demographic groups: it does not allow unconditional inferences such as 'A Young person will vote Yes with a probability of such and such percent'.

The impact of any given demographic, instead of being parametrical, is an inverted U function of S; zero when S = 0, and zero when S =1, but non-zero in between. This can be rationalised by imagining that the vehicle of infection is not so much the individual but 'the atmosphere' that each individual breathes in. So we might imagine a largely native-born electorate creates a nativist atmosphere. It is plausible that a single diluting Briton can barely alter this atmosphere: what is one British-born among 10,000? Thus a few more Britons will do little to change the existing (nativist) atmosphere. But, going the other direction, suppose an electorate has become overwhelmingly British-born; a new Britannia, no less, and a 'Britannic' atmosphere has been established. Again, the arrival a few more Britons will do little to change the now existing (Britannic) atmosphere; it is already hegemonic. It seems that increase in British-born becomes least important when an some atmosphere – nativist or neo-Brittanic – is established; and most important when no single atmosphere has established itself, but instead Federation-hostile and Federation-friendly atmospheres are in contention. Thus we arrive at the inverted U shaped between the impact of an expansion in the presence of some demographic and the size of the demographic presence; little impact when the demographic is weak (and the prevailing atmosphere is hostile), the strongest impact when it is middling (and it is a matter of contention which atmosphere shall prevail), and little impact again when the demographic is strong (and the atmosphere is favourable).

The logistic model, too, has problems.

Firstly, there is no reason why the fields of 'attitudinal contagion' should conterminous with electoral boundaries. West Torrens with 30.5 per cent of the adult population British-born (compared to a South Australian average of just 23.5 percent) may have exercised some attitudinal contagion on adjacent electorates. Or visa versa. Or perhaps the entire state was a single field of 'attitudinal contagion'.

This 'atmospheric' rationalisation of the logistic specification also beg questions about what can be rationally inferred regarding the probability of members of a given demographic voting Yes. One sweeping answer to this question would be to suppose that the probability voting Yes in electorate i is the same for of *all* demographics, S_i. This amounts to an extreme case of 'atmospheric' explanation: each of the various groups do make a certain contribution to that atmosphere, but all the demographics are equally and totally susceptible to that atmosphere. This is an acutely Durkhemian vision of social causation, and in blunt contrast with the methodological individualism of economics. And this particular variant of the Durkheimian vision is, surely, factually false; different demographics do exhibit a different frequency of voting within a given community. There is, however, a less extreme, but still Durkheimian vision, of inferring a demographic's probability of voting Yes from the coefficients of a logistic specification. It is to suppose that, with atmosphere having been established and producing a certain S_i, the demographic groups assume a probability of voting Yes that is consistent with (i) S_i, and (ii) the marginal 'impact' of the presence of a demographic implied by the logistic function.

$$\frac{\partial S}{\partial X_j} = \beta[1-S]S$$

Thus if S_i = 0.75 and β equals 0.2 and 60 percent of the population are j then the probability of a member of demographic j voting Yes is 0.765, and the probability of a non-member of demographic j voting Yes is 0.7275. [4] This procedure at least allows for

[4] Check: $0.75 = S = p_J Xj + p_{\sim J} (1-Xj) = 0.765*0.6 + 0.7275*0.4$

More generally, the logistic model implies,

the observed fact that different demographic groups have different probabilities of voting Yes. But the hypothesis is contrived solely to accommodate this observed fact with an extreme Durkheimian assumption of an all-deciding 'atmosphere'. There is nothing else to recommend the assumption.

In reality, atmosphere doubtless has a role in determining an individual's vote, but so does their individuality. Thus the logistic vision solves an econometric problem by invoking a lop-sided causal cosmology. Therefore despite the 'best practise' status commonly awarded the logistic specification, its superiority over the linear/additive model is, in this paper's view, not complete. We will proceed by first using logistic specification, and in a later section using a linear/additive model.

Estimation Issues

The logistic specification can be rewritten,

$$\ln \frac{S}{1-S} = \alpha + \beta X + u$$

Because the RHS is linear this may be estimated by OLS (Manning 1996).

$dS/dXj = \beta\,[1-S]S$

and an additive representation supposes

$dS/dX_J = p_J - p_{\sim J}$

implying

$S = (p_J - p_{\sim J})X_J + p_{\sim J} = \beta\,[1-S]S\,X_J + p_{\sim J}.$

or

$(S - \beta\,[1-S]S\,X_j) = p_{\sim J}$

From this equality $p_{\sim J}$ - the probability of a non-J voting Yes - can be inferred, as S is predicted, β is estimated, and X_j is measured,

The assumptions of OLS are relatively undemanding, at least in the present context, but three pervasive difficulties merit consideration.

1. The RHS variables need to be exogenous, rather than endogenous. This need not be strictly true here. Consider the association of Educational Attainment and voting Yes. It is plausible that a higher Educational Attainment will encourage a voter to support Federation (ie Educational Attainment is exogenous). But it is conceivable that supporting Federation will induce a voter to increase their Educational Attainment (i.e Educational Attainment endogenous). The possibility is intelligible, but this paper will take it that it is safe to assume that all regressors are exogenous.

2. Excluding relevant variables will bias the estimates of any included variables they are correlated with. Thus we have no data on income; but income is surely correlated with many variables we do have data on. So if a variable is estimated to be significant, is it truly relevant; or is it just correlated with income and income is relevant?

3. The abundance of variables makes regressor choice problematic. We have 52 observations of the dependent variable; but the census provides dozens of religions, places of birth, and age groups. We only wish to include only the relevant regressor, but don't know which is relevant.

In the face of this abundance of potential regressors, the paper investigates, in turn, the explanatory potential of Place of Birth, then Occupation, and finally Religion. We commence each investigation by including as many of the variables of a given explanatory category that the 'degrees of freedom' permit. We also always include a shift dummy for 1899, and three demographic conditioners, Sex (the proportion of over 21 adults that are male), Marital Status (the proportion of adults married) and Age (the proportion between 21 and 35). If not all candidate variables are significant at the 10 per cent level, then an F test is performed on the joint significance of those not significant; if the hypothesis of irrelevance is not rejected then the variables at issue are omitted, and the regression is

repeated with just those that were significant. If the hypothesis of joint irrelevance of individually insignificant regressors is rejected, then the least statistically significant regressor is dropped, and the exercise is repeated until we have reached a suite of variables each of which is significant at the 10 percent level or better. Those regressions are reported below.

Having examined Place of Birth, Occupation and Religion separately, the paper then explores the significance of these types of variables in the presence of each other. The paper runs a horse race that includes all those variables that have been previously identified as statistically significant within each of the three categories. The results of this regression yields the paper's best assessment of the socioeconomic factors that govern Yes voting.

The paper concludes by exploring some extensions and counterfactuals.

II Results

Place of Birth

The great number of places of birth provided by the census necessitates some preselection of place of birth variables. Those preselected where Born in South Australia, Born in Australia but outside South Australia, born in Britain (= England, Wales and Scotland), and Born in Ireland. Winnowing of regressors left three significant;

Table 1: Native-born and British-born

	slope coefficient; point estimate	*slope coefficient; 90 percent confidence interval*	*marginal impact; point estimate*	*marginal impact; upper and lower bounds*
native-born: South Australia	4.65*** (2.86)	1.93, 7.39	0.82	0.34, 1.31
native born: rest of Australia	8.32*** (2.86)	3.45, 13.21	1.47	0.64, 2.31
British born	-2.57* (1.94)	-4.79, -0.35	-0.46	-0.85,-0.06
1899 dummy	0.51*** (4.99)	0.34, 0.69	0.09	0.06,0.12
square of correlation between actual S and predicted S	0.57			

T statistics in absolute values in brackets; *, ** and *** denote a level of significance of 10, 5 and 1 percent respectively. All regressions reported in this paper use a constant term, but they are omitted from the presentation of results. With the exception of one borderline case, all regressions reported in this paper survive a Breusch-Pagan test for heteroscedasticity of the disturbance term.

The second column indicates being British-born makes for voting No. By contrast being born in South Australia contributes significantly to voting Yes, as does being born outside South Australia, but within Australia. The point estimate suggests that the South Australian born were not as Yes-inclined as those who were born in Australia but outside of South Australia. Is this difference in pint estimates a matter of chance? A test of the restriction of the coefficients being the same on those born in South Australia and those born in Australian born (but outside South Australia) is just rejected at the 10 percent level.

How quantitatively significant were these effects? The fourth column in Table 1 reports the 'marginal impact' implied by the point estimate of the slope coefficient, and supposing S = 0.77.[5] These impacts seem large. Thus another 1 percent of the population being born in South Australia increases the Yes vote by 0.82 percent. But what precision can be attributed to these point estimate of impact? The final column reports the upper and lower estimates of the marginal impact implied by the 90 percent confidence interval of the slope coefficient that is reported in the third column. We see that the upper and lower bounds for the impact of another 1 percent of the population being born in South Australia varies from a moderate 0.34 per cent to a somewhat unbelievable 1.31 percent. This conjunction of the 'best guess' of an impact being strong, but both weaker and unbelievably large impacts being accommodated by the data, recurs in this analysis.

Occupation

The census reports these categories of occupation as a percentage of the population of an electorate: Professional, Commercial, Transport and Communications, Manufacturing, Agricultural, Pastoral, Mining, and Domestic. It proves useful to use the following aggregations:

'Business' = Professional + Commercial,
'Industrial' = Transport and Communications + Manufacturing,
'Rural' = Agricultural + Pastoral,

Investigating these aggregations, along with Mining and Domestic, revealed only one statistically significant category: Industrial.

[55] The median Yes vote proportion of the electorates over 1898 and 1899, excluding the Northern Territory, was 77 percent.

Table 2: Economic occupation

	slope coefficient; point estimate	*slope coefficient; 90 percent confidence interval*	*marginal impact; point estimate*	*marginal impact; upper and lower bounds*
industrial occupation	-5.18 *** (5.02)	-6.91,-3.45	-0.92	-1.58.-0.61
1899 dummy	0.51*** (4.84)	0.34,0.69	0.09	0.06,0.12
square of correlation between actual S and predicted S	0.34			

Table 2 reports confidently reports a remarkably strong preference for No of those in an Industrial Occupation. It is, indeed, unbelievably strong; the point estimate has a 1 percent increase in the proportion of the electorate in an industrial occupation reducing the Yes share by 0.82 percent. The negative result does correlate with the of the labour movement's hostility towards Federation. The South Australia Labor leader of the period, E.L. Batchelor, argued 'consistently' against the federation Bill (Bannon 1999). This raises a possible explanation of the negative impact industrial occupation: it is plausible that Labor voting persons – persons who took their cue from the Labor party – tended to be in an industrial occupation. An upshot is the possibility that 'industrial occupation' is simply proxying for 'Labor-voter'. To test if the results of Table 2 are simply a manifestation of the opposition to Federation of the Labor party, the regression was repeated including as a variable the average percentage Labor received in each electorate of the SA Legislative Assembly in the elections of 1896 and 1899.

Table 3: Occupation and Labor Party vote

	slope coefficient; point estimate	*slope coefficient; 90 percent confidence interval*	*marginal impact; point estimate*	*marginal impact; upper and lower bounds*
industrial occupation	-3.88*** (2.81)	-6.19, -1.56	-0.69	-1.10, -0.28
Labor vote	-0.53 (1.41)	-1.15, 0.10	-0.09	-0.20,0.02
1899 dummy	0.51*** (4.88)	0.34,0.69	0.09	0.07,0.12

The Labor vote variable is statistically insignificant. But industrial occupation remains statistically significant, and quantifiably powerful. Thus industrial occupation is not simply proxying for a political allegiance to Labor.

If industrial occupation is not simply a surrogate for Labor, could it still be just the proxy of some other factor? In a careful case study of the two electorates with the smallest Yes vote in 1898 - West Torrens and West Adelaide - Norris (1969) argues that the low Yes was essentially a protectionist vote against the prospect of intra-Australian free trade. Norris' study raises the possibility that the results of Tables 2 and 3 are dominated, indeed traceable, to just these two extreme electorates. To assess the possibility the regression of Table 2 was repeated, omitting West Torrens and West Adelaide.

Table 4: Occupation: West Torrens and West Adelaide excluded

	slope coefficient; point estimate	*slope coefficient; 90 percent confidence interval*	*marginal impact; point estimate*	*marginal impact; upper and lower bounds*
industrial occupation	-3.45 *** (2.77)	-5.5, -1.35	-0.61	-0.98, -0.24
1899 dummy	0.46*** (4.42)	0.28, 0.63	0.08	0.05,0.11

The result of Table 2 persists in Table 4. This does not refute Norris' thesis that protectionism was the underlying operative force, but does suggest that if it was protectionism that was at work in these West Torrens and West Adelaide, then it was also at work in all electorates with a large number of industrial workers.

Religion

Religious affiliation has sometimes been mooted as a factor in explaining support for Federation. Thus Congregationalism – almost at the peak of its strength at the close of the 19th c – was enthusiastically Federationist.

The 1901 census reports for each electorate the number of adherents to these religious affiliations: Church of England, Roman Catholic, Presbyterian, Methodist, Congregationalist, Baptist, Lutheran, and many others.[6] The winnowing of potential regressors by statistical significance left two significant at the 5 per cent level; the proportion of the aggregate of major denominations adhering to the Church of England, and to Baptism. Two such proportions were significant at the 10 percent level: the Presbyterianism and Congregationalism.

[6] Aggregating these into "Protestant' or "Dissenter' did not prove useful.

Table 5: Religion

	slope coefficient; point estimate	*slope coefficient; 90 percent confidence interval*	*marginal impact; point estimate*	*marginal impact; upper and lower bounds*
Church of England	-1.63** (2.51)	-2.73, -0.54	-0.29	-0.48,-0.10
Baptist	-3.92** (1.98)	-7.25, -0.60	-0.70	-0.11, -1.28
Presbyterian	2.04* (1.80)	0.13, 3.93	0.36	0.02,0.70
Congregationalist	4.72* (1.88)	0.50, 8.93	0.84	0.09, 1.58
Male	1.99* (1.68)	0.004, 3.97	0.35	0, 0.70
1899 dummy	0.51	0.34,0.70	0.09	0.06, 0.12
square of correlation between actual S and predicted S	0.51			

The positive sign on Congregationalism and Presbyterianism will not surprise historians of Federation. But what of the negative coefficients on some other Protestant religious affiliations? There may be a temptation to trace the negative sign on Baptists to the attitudinal heritage of this strain of Christianity, that was near the ultimate (English) expression of religious alienation from the state. This wariness might be seen in Baptists' total opposition to State Aid to religion in 19th c Australia (Walker 1962, p173). Of course, it is true that some Baptist ministers were enthusiasts for Federation (Ashley 2001). But the claim is not that every Baptist opposed Federation. The claim is that Baptism strongly encouraged No vote, and the Table above supports the claim .

But what of the negative impact of membership of the Church of England? What doctrinal tenet of that church could tend against Federation? Does the negative coefficient, perhaps, just reflect the opposition of the British-born to Federation? In other words, did members of the Church of England tend to vote No simply because they tended to be British?

A Multi-variable analysis

To put the point more generally, a defect with the single category regressions is that it does not allow for the possibility that a variable in one category is only succeeding on account of it proxying for variable of another category. This possibility is of particular concern given the significant degree of correlation between variables that have been deemed in the previous section to be statistically significant.

Table 6: Correlation of Regressors

(all variables measured as decimal proportions)

	Yes	*SA born*	*native born*	*British*	*C of E*	*Baptist*	*Presb*	*Cong*	*industrial*	*men*	*married*
SA born	0.43										
native born	0.51	0.91									
British	-0.46	-0.65	-0.67								
C of E	-0.28	-0.60	-0.44	0.51							
Baptist	-0.36	-0.19	-0.29	0.31	0.14						
Presb	0.16	-0.18	0.16	-0.04	-0.09	-0.29					
Cong	-0.08	-0.07	-0.11	0.38	0.31	0.33	-0.27				
industrial	-0.53	-0.80	-0.79	0.52	0.51	0.42	0.06	0.19			
men	0.24	0.45	0.54	-0.55	-0.09	-0.43	0.05	-0.48	-0.31		
married	-0.08	0.35	0.17	0.03	0.04	-0.54	0.25	-0.20	0.15	-0.24	
under 35	0.02	-0.27	-0.10	-0.01	-0.01	0.28	-0.10	0.14	-0.14	0.39	0.44

Notice, in particular, the substantial positive correlation of membership of the Church of England with British birth, and with Industrial Occupation; and the substantial negative correlation of membership of the Church of England with being native-born. Notice also the significant negative correlation of Industrial Occupation with being native-born.

To sort out what variables perform only on account of their proxying for others, a ‘horse race’ regression was conducted that included all variables that have thus far been found statistically significant - South Australian Born, British born, Church of England, Baptist, Presbyterian, Congregationalist and Industrial Occupation. These were then winnowed by statistical significance. The procedure left just two variables significant at 10 per cent or better.

Table 7: The Horse Race

	slope coefficient; point estimate	*slope coefficient; 90 percent confidence interval*	*marginal impact; point estimate*	*marginal impact; upper and lower bounds*
British born	-2.69** (3.02)	-4.56,-0.82	-0.48	-0.81,-0.15
industrial occupation	-3.74*** (3.25)	-5.67, -1.81,	-0.66	-1.02,-0.32
1899 dummy	0.51*** (5.07)	0.34,0.68	0.09	0.06,0.11
square of correlation between actual S and predicted S	0.62			

Table 7 is the culmination of the paper's method, and is the 'preferred regression'. Its first message is that being British-born contributed powerfully to a No vote: an extra one percent of the electorate being British-born reduced the Yes vote by 0.48 of one percent, according to the point estimate. The second message is that being of industrial occupation was also powerfully disposing to vote No: an extra one percent of the electorate being in an industrial occupation reduced the Yes vote by 0.66 of one percent, according to the point estimate.

III Some Extensions

Accounting for the shift to Yes at the 1899 Referendum

The data set deployed in section II has pooled the results of the 1899 referendum with those of 1898, and the analysis has sought to cope with the palpable shift to Yes in 1899 by dummy shift term. This amounts to assuming the slope coefficients were unchanged

between 1898 and 1899, and this in turn comes down to assuming all the shift to Yes was equally shared by all groups, so that the differentials of attraction/repulsion to Federation between the various demographics was preserved. An alternative hypothesis would allow those differentials to have changed. This hypothesis would call for including not just a shift dummy, but also slope dummies, and so explore possibility of a reduction the propensities of different groups (British-born, Industrial occupation etc) to vote No. Adding a slope dummy for British born makes no difference; the impact of being British born in voting No did not change between 1898 and 1899. However, adding a slope dummy for industrial occupation does make a significant contribution. In fact, the shift dummy term loses all statistical significance in the face of a slope dummy for industrial occupation. Thus the entirety of the shift towards Yes in 1899 seems to be locatable in the decline in the propensity of industrial voters to vote No. Further, this decline is so substantial that an F test cannot reject the hypothesis that the sum of the slope coefficient and the dummy slope coefficient is zero; meaning that the data cannot reject the *complete absence* of a No vote effect for industrial occupation in 1899. Thus whatever No effect the presence industrial workers might have induced in 1898, that effect wholly disappeared in 1899.

Table 8: Accounting for 1899 by a slope dummy

	slope coefficient; point estimate	*slope coefficient; 90 percent confidence interval*	*marginal impact; point estimate*	*marginal impact; upper and lower bounds*
British born	-2.69* (2.74)	-4.34 -1.04	-0.48	-0.77, -0.18
industrial occupation	-5.96*** (5.57)	-7.76 -4.17	-1.06	-1.37, -0.74
industrial occupation: 1899 slope dummy	4.44*** (6.78)	3.34, 5.54	0.79	0.59, 0.98

The message of Table 8 is worth amplifying. It implies that the 1899 referendum saw no increase in the Yes propensity across each demographic. Rather, the propensity to vote No of one 'resisting' group – industrial workers – collapsed. This collapse is made even more spectacular when we consider Table 8's estimate of the strength of the No vote among industrial workers in 1898: according to the point estimate, an increase in 1 per cent of the share of the population in industrial occupation reduced the Yes vote by 1.06 percent.

Interaction Effects

Is it possible that what counted against voting Yes was not so much British birth as such, and not so much being of Industrial Occupation as such either, but being a British-born person of an Industrial Occupation? The census does not provide cross-tabulations by Place of Birth and Occupation. But if Place of Birth and Occupation are independent events in the statistical sense – a strong assumption, certainly – then the proportion of the population that are British born in an Industrial Occupation will equal the product of the two proportions. Table 9 reports the effect of introducing such an 'interaction' variable.

Table 9: Interaction of British birth and Industrial Occupation

	slope coefficient; point estimate	*slope coefficient; 90 percent confidence interval*	*marginal impact; point estimate*	*marginal impact; upper and lower bounds*
British born	-0.78 (0.26)	-4.31,5.88	0.14	-0.76, 1.04
industrial occupation	-2.39 (3.25)	-6.21, 11.00,	0.42	-1.10, 1.95
interaction term: British born and industrial occupation	-25.969 (1.23)	-61.42, 9.49	-0.46	-10.88, 1.72
1899 dummy	0.51*** (5.10)	0.34, 0.68	0.09	0.06, 0.12

No variables apart from the shift dummy are significant. The introduction of an interaction term is completely without success.

The Issue of Participation: Choosing to Not Vote

Since 1924 compulsory voting has been a barely questioned pillar of the Australian electoral mechanisms. But voting in the Federation referenda was not, of course, compulsory. The question arises, could the fact that voting was not compulsory, and that a choice could be made to not vote at all, bias the estimates?[7] Insight into that question requires a model of voting decision. In one simple model each voter compares their valuation on voting with the cost of voting, and votes if the valuation exceeds its cost. Given a range of valuations within an electorate, such a value:cost model of voting decision implies a 'demand schedule' for voting Yes for each electorate, where

[7] Most registered voters in 1898 chose not to vote.

'demand' (ie the number choosing to vote Yes) is a negative function of voting cost. The value:cost model also implies a 'demand schedule' for voting No for each electorate, where again 'demand' (the number choosing to vote No) is a negative function of cost. Each demand curve will have a certain 'elasticity to cost', and that elasticity may differ between Yes and No. If we suppose the cost of voting varies across electorates then our neglect of participation *may* have biased the results. Consider following scenario. Suppose the demand to vote Yes of those in industrial occupations is, in truth, identical with those not of industrial occupation; *and* that the demand to vote No is also identical across industrial and non-industrial occupations; *but* that the (common) No demand schedule is *more elastic* to cost of voting than the Yes is to the cost of voting; *and* (finally) that voters of industrial occupation reside disproportionately in low voting cost electorates. Then No voters in an industrial occupation will come out to vote more than other voters. The upshot is that industrial occupation will be negatively correlated with Yes. But this is not a reflection of some greater propensity of industrial occupations to vote No – by assumption all occupations have the same Yes and No demand schedules – but simply of the scenario's assumptions that the No vote is more elastic to cost *and* industrial occupations are concentrated in low voting cost districts.

In the scenario above, voting cost is obviously a relevant explanatory variable, and by excluding it we may gave biased results; and wrongly inferred that those of industrial occupation have a different propensity to vote No when, by the assumption of the scenario, they do not.[8] To eliminate that bias we would need to include a measure of voting cost. But what would that be? Participation might be one; the lower voting cost, higher participation. Thus the regression below includes participation as a proxy for voting cost.

[8] If cost of voting is the same in all electorates, its exclusion would have no impact on the regression since its variance would be zero. This conclusion obviously encompasses the case where the cost of voting is zero, or, in the eyes of voters, uniformly trivial.

Table 10: Allowing for Participation

	slope coefficient; point estimate	*slope coefficient; 90 percent confidence interval*	*marginal impact; point estimate*	*marginal impact; upper and lower bounds*
British born	-2.68** (2.41)	-4.55, -0.81	-0.47	-0.81,0.14
industrial occupation	-3.97*** (3.35)	-5.96,-1.98	-0.70	1.36,-0.35
participation	-0.56 (0.84)	-1,67,0.56	-0.10	-0.30, 0.10
1899 dummy	0.60 (5.81)	0.36, 0.83	0.11	0.06, 0.15

Evidently, participation has no explanatory value.

Linear/Additive Specification

The regressions have so far assumed a logistic specification. To explore whether a linear specification would make any difference, linear regression was conducted of all variables that have been found significant above - South Australian born, born in Australia but outside South Australia, British born, Church of England, Baptist, Presbyterian, Congregationalist and in Industrial Occupation – with regressors then winnowed by statistical significance. The procedure left just two variables a significant at 10 per cent or better.

Table 13: The Linear Horse Race

	slope coefficient; point estimate	*slope coefficient; 90 percent confidence interval*
British born	-0.51** (2.33)	-0.88,-0.14
industrial occupation	-0.80*** (3.51)	-1.18, -0.46,
1899 dummy	0.10*** (5.00)	0.066, 0.13
square of correlation between actual S and predicted S	0.53	

Thus the importance of being British born and of Industrial Occupation in voting No is again found. And the greater strength of No voting among persons of Industrial Occupation than persons of British vote is again found.

V Some Counterfactuals

Compulsory voting, not voluntary voting.

How might the outcome have been different if voting was compulsory? To attempt an answer on the basis of a value:cost model of voting would be a matter of some intricacy. It would require estimates of the elasticity of voting demand schedules for Yes and No, which would in require data on voting cost. But there is *ad hoc* method which might simulate the effect of compulsory voting. It turns on the measurement of Yes as a proportion of all enrolled voters, rather than as a proportion of actual votes: (i) regress the Yes vote, expressed as a proportion of all registered voters, on various variables, including the participation rate; (ii) regress the No vote, as a proportion of registered voters, on variables including the participation rate; and (iii) infer from these two regressions the predicted Yes vote and No vote if (counterfactually) a 90 percent

participation rate is imposed. Since 'compulsory voting' has amounted to about 90 percent of enrolled voters voting this procedure might simulate the outcome a 'compulsory' vote.

1901 not 1881

Schemes of Federation had been mooted since the 1850s. Why did one suddenly come to fruition after four decades? Were there changes in society that might have encouraged this? We can use the preferred estimated equation to predict what would have been the Yes vote in 1881, given data on the proportion of British born adults in 1881, and the proportion in industrial occupations in 1881. There was certainly a substantial shift of the population towards the native-born between the 1881 and 1901. In 1881 39 percent of 15 and over in SA were born in Great Britain. By 1901 24 percent of those over 21 were born in Great Britain. By both logistic and linear estimates, a decline of 15 percentage points in the share of the British born would increase the Yes vote by 7 or 8 percentage points.

Adult Franchise, not Manhood Suffrage

The Constitution Amendment Act of 1894 granted women in South Australia the right to vote on the same terms as men. By contrast, women did not vote in the referenda in New South Wales, Victoria, Queensland and Tasmania. What impact did the extension of the franchise have on the vote in South Australia? The regression analysis of this paper provides zero evidence of any pure 'gender effect' on voting: the proportion of the over-21 population being male has been included in every initial regression, and in every such regression this variable has been dropped as statistically insignificant. But despite the evident absence of a pure 'gender effect', it may be that extending the franchise to women did have an impact on the Yes vote. This is because British-born and Industrial Occupation has been found to be relevant, and fewer women than men were British-born, and fewer women than men were in an Industrial occupation. Both

these differences would have contributed to women voting more in favour of Federation than men.

V Conclusions

The exploration of census data at the electorate level indicates two factors that powerfully contributed to voting No in South Australia's Federation referenda: being British-born, and working in an industrial occupation.

The conclusion that the British-born were No voters is not surprising; the 'nativism' of the federation movement is well-known to historians. Nevertheless, there is a value in confirming this 'prior', and in obtaining a measure of its strength; both logistic and linear models suggest that an extra 1 percent British born share of the population would reduce the Yes vote by about one half of one percent. It is also worth noting the result is restricted to *British*-born; Irish born seem to have no tendency to vote No any more than other non-British.

That industrial occupation distinctly discouraged voting Yes in 1898 would not astonish historians of Federation, either; but again, the strength of the effect is noteworthy: both logistic and linear models suggest that an extra 1 percent the population working in industry would reduce the Yes vote by distinctly more than one half of one percent. And yet the strong indications of the data are that this effect was restricted to 1898, and that 1899 saw the complete annihilation of the No voting propensity of industrial workers. There is a mystery to puzzle over.

References

Ashley, Robert W.P (2002), 'Rev J.A. Soper: Ballarat and federation, the evangel and social concerns' *Victorian Baptist Historical Society*, 10, pp. 100-21.

Bannon, J.C. (1999) 'South Australia' in *The Centenary Companion to Australian Federation*, Helen Irving ed., Cambridge University Press.

Blainey, Geoff (1950) 'The role of economic interests in Australian federation: A reply to professor R. S. Parker', *Historical studies, Australia and New Zealand*, Volume: 4(15), pp. 224-237

Historical Census and Colonial Data Archive http://hccda.anu.edu.au/

Manning, Richard L (1996), 'Logit Regressions with Continuous Dependent Variables Measured with Error' *Applied Economics Letters*, 3 (3),pp.183-184.

Norris, R. (1969), 'Economic influences on the 1898 South Australian Federation Referendum' in *Essays in Australian Federation*, A.W. Martin, ed., Melbourne University Press.

Parker, R. S. (1949) 'Australian federation: The influence of economic interests and political pressures', *Historical studies, Australia and New Zealand*, 4(13), pp 1-24

Rhodes, Glenn (1988) *The Australian Federation referenda 1898-1900: a spatial analysis of voting behaviour*, doctoral dissertation, London School of Economics.

Rhodes, Glenn (2002) *Votes for Australia*, Centre for Australian Public Sector Management

Urbatsch, Robert (2013) 'A Referendum on Trade Theory: Voting on Free Trade in Costa Rica', *International Organisation*

Walker, R.B. (1962) ‘The abolition of State Aid to religion in New South Wales’, *Australian Historical Studies,* 10(38),pp. 165-177

.