# The employment gap for female parents in London

Women with children in London have an average employment rate 6 percentage points lower than their peers in the rest of the UK. This very large gap is unique to this particular group: the employment rates of men, whether parents or not, and women without children are the same for those respective groups within London and in the UK overall.

Why does this gap exist? Differences in the demographic composition of London relative to the rest of the UK does not appear to explain the gap entirely, though it is reduced somewhat. Instead, it is possible that high cost-of-living pressures in large cities may be at fault.[[1]](#footnote-1)

**Figure 1: employment rates in London and the UK overall for selected groups, 2022**

Source: GLAE calculations on LFS Households April-June 2022 data.

Exploring possible explanations: a regression approach

We want to find characteristics which are correlated with the observed gap, but we do not necessarily need to understand what is causing it directly. We can therefore construct a simple regression equation designed to “predict” an employment rate based on personal characteristics, without being too concerned about the direction of causality or collinearity.[[2]](#footnote-2)

This equation provides an intuitive approach to investigating the employment gap using data on female parents only:

|  |  |  |
| --- | --- | --- |
|  |  | (1) |

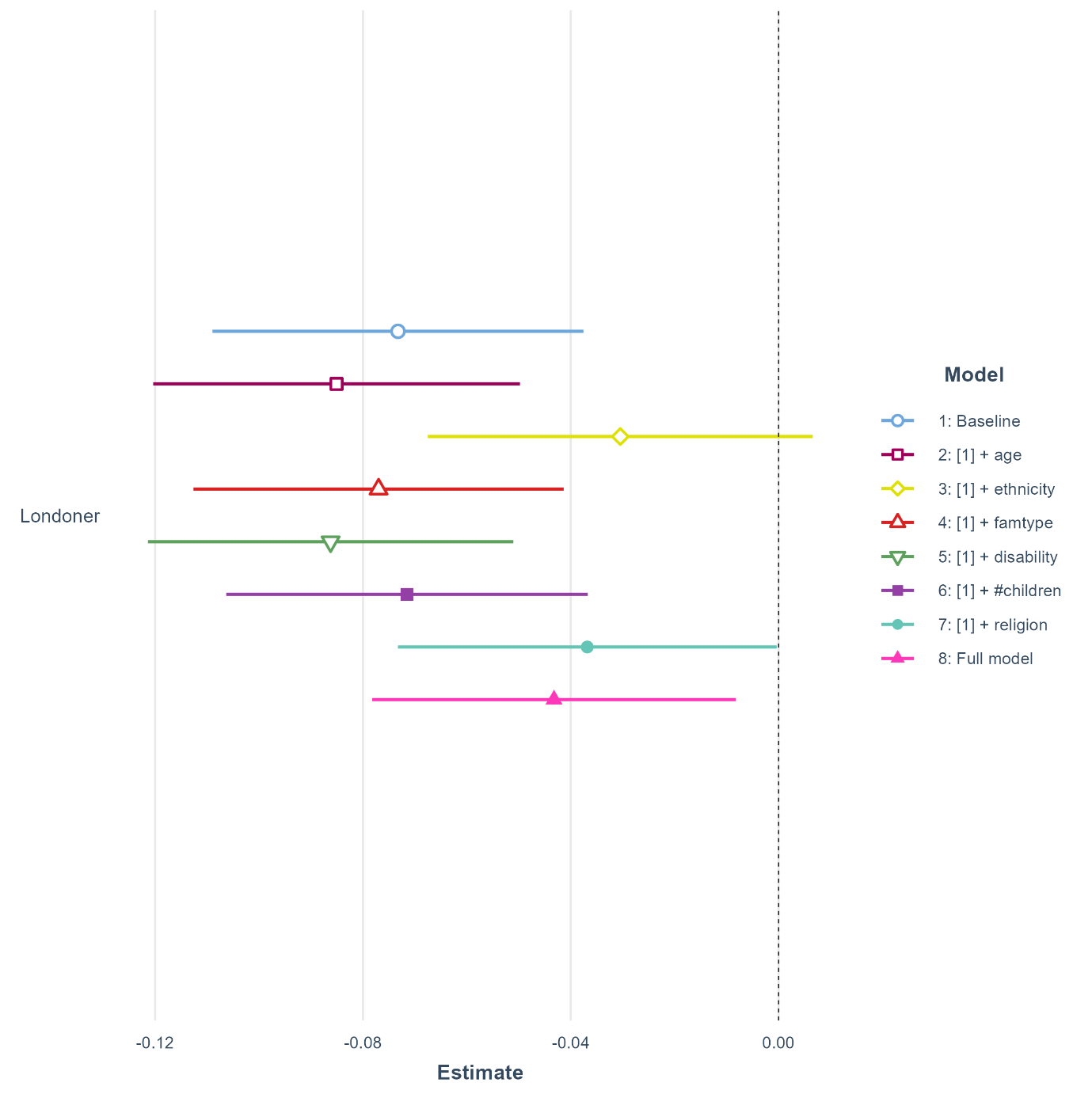
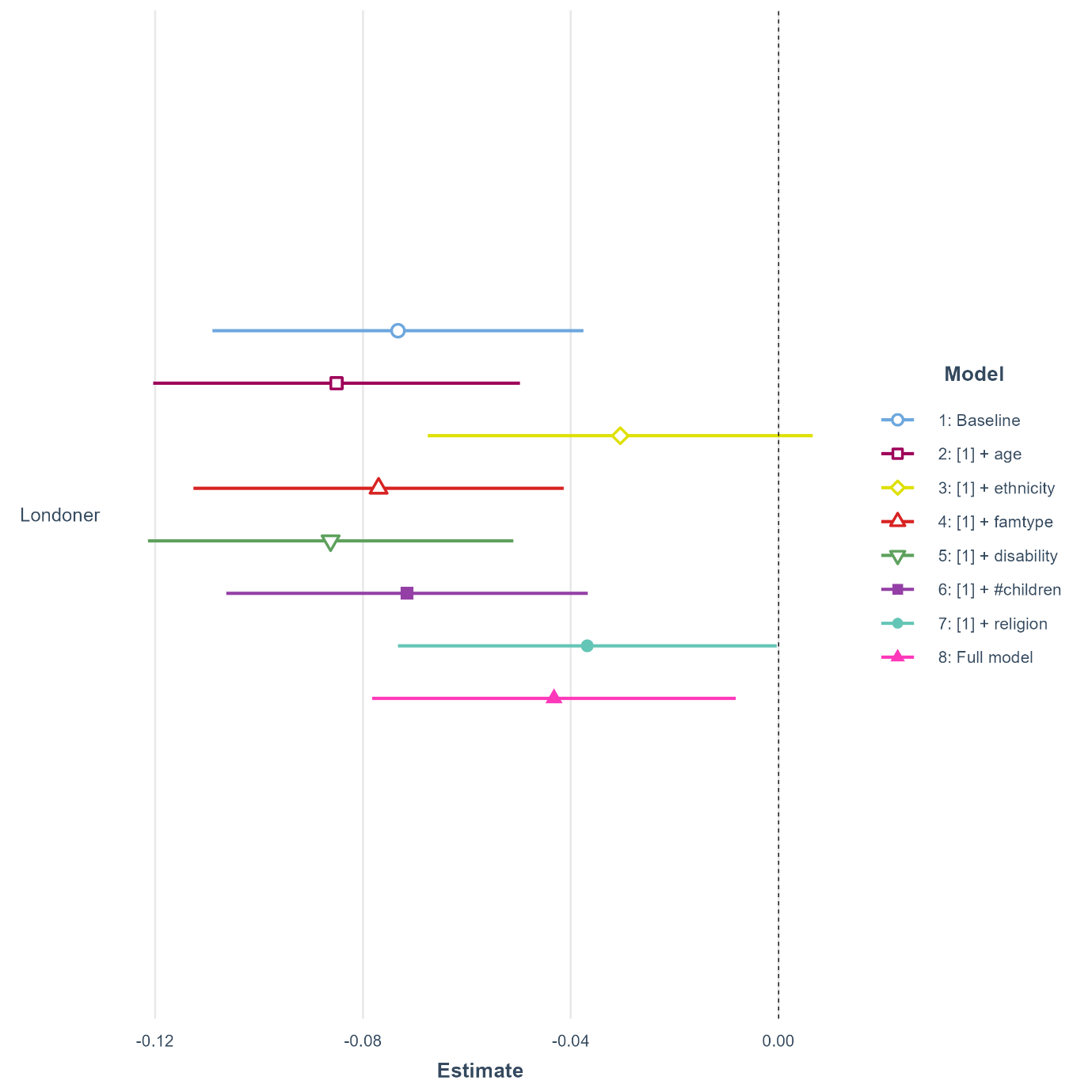
The dependent variable on the left-hand side is the employment rate, which we attempt to explain through independent variables characterising each woman in the dataset. The variable is a dummy. The coefficient of interest , highlighted red, is therefore the gap shown between the London and UK employment rates for female parents shown in Figure 1 above.

Results: demographics do not explain it all away

Figure 2 presents the estimate of in eight different models: the basic model without additional covariates, then adding a single variable (as ) to the basic model, then all covariates at once.

The only model where the gap becomes statistically insignificant is when including ethnicity, here a simple BAME dummy, though the point estimate remains negative. Yet, when including all other covariates in the full model, the gap is again statistically significant. Goodness-of-fit indicators also suggest that the full model is far better at predicting the employment rate.[[3]](#footnote-3)

**Figure 2: estimate of female parental gap in London**



Source: GLAE calculations on LFS Households April-June 2022 data.

Other unobserved characteristics may play a part

The LFS households dataset does not include information on personal or household income, which may otherwise be a significant factor in determining whether or not a parent decides not to work. London has some of the highest external childcare costs in the UK which may mean families find it more financially viable to have a parent stay at home instead.

APPENDIX: Other models and interpreting the coefficients

Equation (1) above is made as simple as possible by subsetting our original LFS dataset, which contains both men and women who can be parents and non-parents. It is not clear whether subsetting affects our regressions negatively, apart from reducing the sample size.[[4]](#footnote-4)

We could also identify the gap using the full dataset, though the equation would be more complex and involve a number of interaction terms:

|  |  |  |
| --- | --- | --- |
|  |  | (2) |

When using a regression with several dummies, we need to be careful to activate the correct dummies to compare the relevant groups. For the simple model, the base case where all dummies are inactive is a female parent living outside Londoner, as the data is already subsetted. For the more complex equation, it would instead be a non-parent male non-Londoner.

To correctly compare female parents inside and outside London, we need to activate the relevant dummies in Equation (2) and compare the estimated employment rates:

* Female parents not in London:
* Female parents in London:

The difference between these two groups boils down to the coefficients activated when the dummy , or .

When performing the regressions, we can see that the coefficients indeed sum up to the same gap regardless of specification. However, the sheer amount of interaction terms needed make the models cumbersome, so we use the simple model.

An important caveat is that the interpretation of coefficients on the demographic covariates are less broad in Equation (1). The coefficient on the dummy for disability would indicate the effect for the whole population in Equation (2), but when using the female-parent-only dataset, the coefficient only tells us which effect disability has on that specific sub-group.

1. Based on LFS Households data in 2022. Birmingham saw a similar and possibly even larger gap in the data. [↑](#footnote-ref-1)
2. We may be interested in investigating causality *once* we’ve established a likely correlation. We also attempt to avoid perfect collinearity as it may provide spurious results when the coefficients merely lose significance due to a loss of precision in the estimates. [↑](#footnote-ref-2)
3. Introducing religion reduces the sample size due to missing responses, such that the full model uses less observations than the ethnicity-only model. [↑](#footnote-ref-3)
4. It is possible the reduced sample size affects our results. While the gap is not significant when controlling for ethnicity in our main model, it continues to be significant as an interacted term when using a larger dataset. However, it is more difficult to interpret the combined significance of multiple coefficients in an interacted model. [↑](#footnote-ref-4)