# The employment gap for female parents in London

Women with children in London have an average employment rate 6 percentage points (pp) 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 (see Figure 1).

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, people aged 16-64.

London’s demographic composition differs from the UK overall

London’s population is more ethnically and religiously diverse than the UK average; more Londoners are single than the average UK resident but fewer are disabled. These differences are reflected in the average employment rate of London relative to the UK overall, as shown in Table 1 below, which breaks down the population by demographic characteristics

For instance, the average employment rate of London residents in the BAME ethnicity group is 67%, somewhat lower than the average of 68% in the UK overall. However, the BAME group made up 40% of London’s population but only 16% of the UK total. In the opposite direction, London has relatively fewer residents with disabilities, who have lower average employment rates, at a share of 18% of the population compared to 21% in the UK overall.

It is not clear which factors affect female parental employment to such an extent that the gap between London and the UK is as large as observed. We therefore employ a simple regression to disentangle these effects on a more narrowly defined population.

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 person 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.

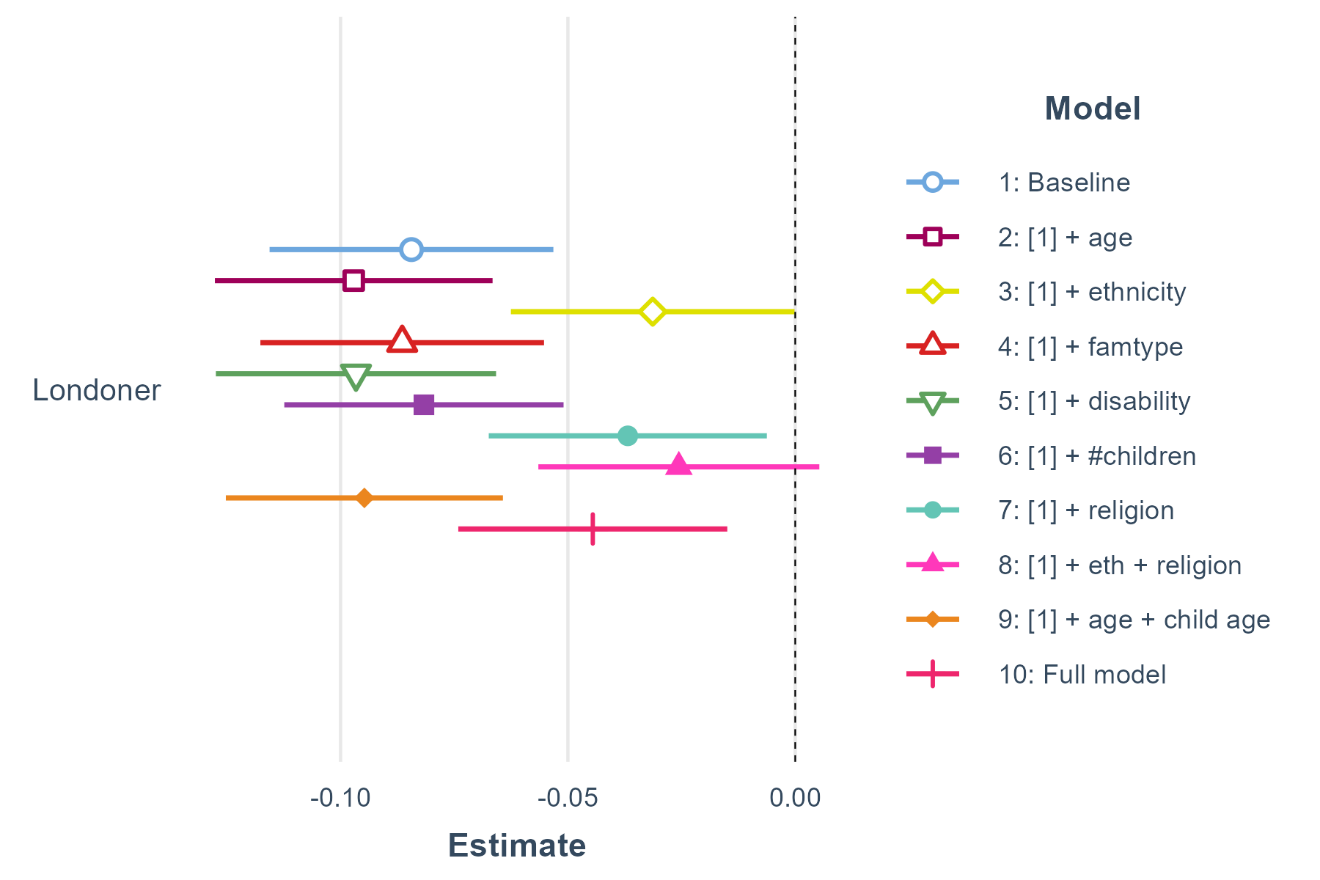
Results: the gap persists despite controlling for characteristics

Figure 2 presents the estimate of (the gap relative to the rest of Great Britain) [[3]](#footnote-3) in eight different models: the basic model without additional covariates, then adding a single variable (as in Equation 1) to the basic model, then all covariates at once.

The models with the lowest point estimates for the gap introduce ethnicity, here a simple BAME dummy, or religion. Using a 90% confidence interval, we cannot reject the hypothesis that the gap is zero when including both variables in model 8, while the point estimate suggests the gap is around 3pp.

Yet, when including all other covariates in the full model, the confidence interval narrows, and the point estimate increases somewhat to 5pp. Goodness-of-fit indicators also suggest that the full model is far better at predicting the employment rate.[[4]](#footnote-4)

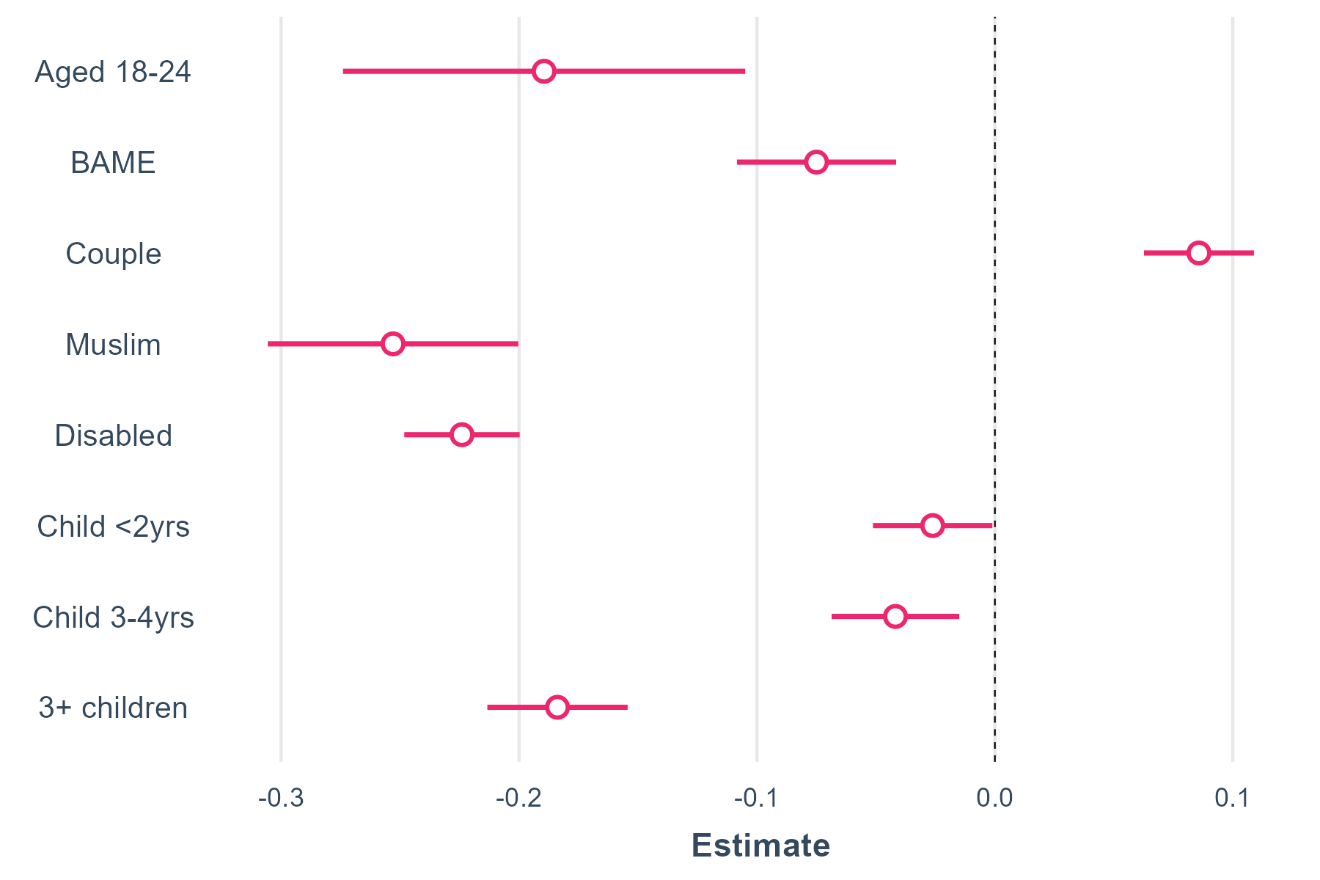
**Figure 2: estimate of female parental gap in London**



Source: GLAE calculations on LFS Households AJ 2022 data, female parents aged 16-64.

Note: point estimates and 90% confidence intervals shown.

**Figure 3: Coefficients of selected covariates in full model**



Source: GLAE calculations on LFS Households AJ 2022 data, female parents aged 16-64.

Note: point estimates and 90% confidence intervals shown.

Characteristics and associated impact

While we do not expect our regression model to establish causal links between personal characteristics and employment, it is useful to consider the respective average impacts of the characteristics on the employment rate implied by the coefficients.

It is important to remember that the coefficients show the average impact of a certain characteristic *relative to a base case* (Table 1)[[5]](#footnote-5). We have defined the base cases to highlight impacts on particular characteristics. For instance, we use “4-18yrs” as the base case for the variable “Youngest child age” so that we can immediately see the impact of having young children through the coefficients.

**Table 1: Base cases by characteristic**

|  |  |
| --- | --- |
| **Characteristic (covariate)** | **Base case** |
| Age group | 25-34 |
| Ethnicity | White |
| Family type | Single |
| Youngest child age | 5-18yrs |
| Disability | Not disabled |
| Number of children | One child |
| Religion | No religion |

Figure 3 shows the coefficient estimates of selected covariates from the full regression model. The association between employment rates and being Muslim or disabled are particularly large and negative, at -25pp and -22pp, respectively. These estimates are even larger than that of the age group “18-24yrs”, who are much more likely to be in education and therefore generally have lower employment rates on average.

Having young children is also correlated with lower employment rates with estimates of a negative 3-4pp impact from children younger than 5. Having 3 children or more is associated with a much higher negative impact at -18pp. On the other hand, being in a couple is associated with a higher employment rate by about 9pp.

Conclusion: 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.

**Table 2: Summary table employment rates and populations sizes by characteristic, in London and UK overall[[6]](#footnote-6)**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **London** | | | |  | **UK** | | | |
|  | **Employment rate** | **Unweighted count** | **Weighted count of population** | **Share of population** |  | **Employment rate** | **Unweighted count** | **Weighted count of population** | **Share of population** |
| **Overall** | 74.2% | 4,677 | 6,250,000 | 100% |  | 75.3% | 47,331 | 41,470,000 | 100% |
| **Sex** |  |  |  |  |  |  |  |  |  |
| Male | 78.9% | 2,244 | 3,126,000 | 50% |  | 78.7% | 22,655 | 20,648,000 | 50% |
| Female | 69.5% | 2,433 | 3,124,000 | 50% |  | 71.9% | 24,676 | 20,823,000 | 50% |
| **Age** |  |  |  |  |  |  |  |  |  |
| Aged 16-17 | 15.4% | 177 | 232,000 | 4% |  | 22.7% | 1,693 | 1,465,000 | 4% |
| Aged 18-24 | 53.2% | 480 | 844,000 | 14% |  | 60.4% | 4,550 | 5,356,000 | 13% |
| Aged 25-34 | 81.9% | 975 | 1,605,000 | 26% |  | 83.7% | 8,081 | 8,917,000 | 22% |
| Aged 35-49 | 85.2% | 1,618 | 2,042,000 | 33% |  | 85.6% | 14,514 | 12,680,000 | 31% |
| Aged 50-64 | 72.0% | 1,427 | 1,527,000 | 24% |  | 71.5% | 18,493 | 13,053,000 | 31% |
| **Ethnicity** |  |  |  |  |  |  |  |  |  |
| White | 79.1% | 2,644 | 3,253,000 | 60% |  | 77.1% | 39,894 | 32,823,000 | 84% |
| BAME | 67.2% | 1,530 | 2,196,000 | 40% |  | 68.2% | 5,252 | 6,146,000 | 16% |
| **Disability** |  |  |  |  |  |  |  |  |  |
| Disabled | 55.2% | 828 | 1,114,000 | 18% |  | 53.8% | 10,115 | 8,801,000 | 21% |
| Not disabled | 78.4% | 3,849 | 5,136,000 | 82% |  | 81.0% | 37,216 | 32,669,000 | 79% |
| **Age of youngest dependent child in family** | |  |  |  |  |  |  |  |  |
| 2 yrs or less | 73.0% | 386 | 619,000 | 10% |  | 79.7% | 3,345 | 3,800,000 | 9% |
| 3-4 yrs | 72.9% | 221 | 328,000 | 5% |  | 78.6% | 2,058 | 2,161,000 | 5% |
| 4-18 yrs | 69.2% | 1,370 | 1,664,000 | 27% |  | 74.0% | 13,429 | 11,499,000 | 28% |
| No children | 76.9% | 2,700 | 3,638,000 | 58% |  | 74.9% | 28,499 | 24,010,000 | 58% |
| **Number of dependent children in family under 19** | | |  |  |  |  |  |  |  |
| 0 children | 76.9% | 2,700 | 3,638,000 | 58% |  | 74.9% | 28,499 | 24,010,000 | 58% |
| 1 child | 71.9% | 849 | 1,119,000 | 18% |  | 76.9% | 8,243 | 7,662,000 | 18% |
| 2 children | 74.7% | 845 | 1,042,000 | 17% |  | 79.3% | 7,893 | 7,075,000 | 17% |
| 3+ children | 57.6% | 283 | 451,000 | 7% |  | 63.6% | 2,696 | 2,723,000 | 7% |
| **Family type** |  |  |  |  |  |  |  |  |  |
| 1 person | 67.7% | 1,369 | 1,950,000 | 31% |  | 65.3% | 12,100 | 11,166,000 | 27% |
| Couple | 77.2% | 3,308 | 4,300,000 | 69% |  | 78.9% | 35,231 | 30,304,000 | 73% |
| **Religion** |  |  |  |  |  |  |  |  |  |
| No Religion | 82.0% | 1,528 | 1,978,000 | 36% |  | 78.6% | 19,442 | 18,124,000 | 48% |
| Christian (all denominations) | 72.1% | 1,683 | 2,091,000 | 38% |  | 75.0% | 18,002 | 15,654,000 | 41% |
| Budhist | 71.8% | 53 | 71,000 | 1% |  | 74.3% | 192 | 199,000 | 1% |
| Hindu | 75.7% | 254 | 308,000 | 6% |  | 78.1% | 663 | 706,000 | 2% |
| Jewish | 81.7% | 68 | 77,000 | 1% |  | 74.2% | 164 | 168,000 | 0% |
| Muslim | 59.0% | 441 | 740,000 | 14% |  | 58.4% | 1,440 | 2,015,000 | 5% |
| Sikh | 77.9% | 45 | 57,000 | 1% |  | 74.9% | 247 | 247,000 | 1% |
| Any Other Religion | 69.0% | 94 | 111,000 | 2% |  | 69.9% | 661 | 642,000 | 2% |

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.[[7]](#footnote-7)

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. The data used for the regressions exclude Northern Ireland as respondents are not asked about their religion there. All models therefore use the same number of observations. [↑](#footnote-ref-3)
4. The pseudo-R2 is 0.16 for the full model while not exceeding 0.06 for any of the other models. [↑](#footnote-ref-4)
5. Interpreting single coefficients also imply keeping all other values constant, which in some cases appears unrealistic: religion and ethnicity are likely to be highly correlated, for instance. [↑](#footnote-ref-5)
6. These tables include all people aged 16-64, including men and non-parents. The weighted counts have been rounded to nearest thousand. [↑](#footnote-ref-6)
7. 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-7)