the same residuals in regression (2.33) and a regression like (2.39), we need to purge the dependent variable of the second segment, which can be seen from the figure to be equal to $\hat{\alpha}_1\boldsymbol{\iota}$.

This suggests replacing $\boldsymbol{y}$ by what we get by projecting $\boldsymbol{y}$ off $\boldsymbol{\iota}$. This projection would be the line segment perpendicular to the page, translated in the horizontal direction so that it intersected the page at the point $\hat{\alpha}_2\boldsymbol{z}$ rather than $\hat{\boldsymbol{y}}$. In the general context, the analogous operation replaces $\boldsymbol{y}$ by $\boldsymbol{M}_1\boldsymbol{y}$, the projection off $\boldsymbol{X}_1$ rather than off $\boldsymbol{\iota}$. When we perform this projection, (2.39) is replaced by the regression

$$\boldsymbol{M}_1\boldsymbol{y} = \boldsymbol{M}_1\boldsymbol{X}_2\boldsymbol{\beta}_2 + \text{residuals}, \tag{2.40}$$

which yields the same vector of OLS estimates $\hat{\boldsymbol{\beta}}_2$ as regression (2.33), and also the same vector of residuals. This regression is sometimes called the **FWL regression**. We used the notation "+ residuals" instead of "+ $\boldsymbol{u}$" in (2.40) because, in general, the difference between $\boldsymbol{M}_1\boldsymbol{y}$ and $\boldsymbol{M}_1\boldsymbol{X}_2\boldsymbol{\beta}_2$ is not the same thing as the vector $\boldsymbol{u}$ in (2.33). If $\boldsymbol{u}$ is interpreted as an error vector, then (2.40) would not be true if "residuals" were replaced by $\boldsymbol{u}$.

We can now formally state the FWL Theorem. Although the conclusions of the theorem have been established gradually in this section, we also provide a short formal proof.

Theorem 2.1. (Frisch-Waugh-Lovell Theorem)

1. The OLS estimates of $\boldsymbol{\beta}_2$ from regressions (2.33) and (2.40) are numerically identical.
2. The OLS residuals from regressions (2.33) and (2.40) are numerically identical.

Proof: By the standard formula (1.46), the estimate of $\boldsymbol{\beta}_2$ from (2.40) is

$$(\boldsymbol{X}_2^\top\boldsymbol{M}_1\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2^\top\boldsymbol{M}_1\boldsymbol{y}. \tag{2.41}$$

Let $\hat{\boldsymbol{\beta}}_1$ and $\hat{\boldsymbol{\beta}}_2$ denote the two vectors of OLS estimates from (2.33). Then

$$\boldsymbol{y} = \boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{y} + \boldsymbol{M}_{\boldsymbol{X}}\boldsymbol{y} = \boldsymbol{X}_1\hat{\boldsymbol{\beta}}_1 + \boldsymbol{X}_2\hat{\boldsymbol{\beta}}_2 + \boldsymbol{M}_{\boldsymbol{X}}\boldsymbol{y}. \tag{2.42}$$

Premultiplying the leftmost and rightmost expressions in (2.42) by $\boldsymbol{X}_2^\top\boldsymbol{M}_1$, we obtain

$$\boldsymbol{X}_2^\top\boldsymbol{M}_1\boldsymbol{y} = \boldsymbol{X}_2^\top\boldsymbol{M}_1\boldsymbol{X}_2\hat{\boldsymbol{\beta}}_2. \tag{2.43}$$

The first term on the right-hand side of (2.42) has dropped out because $\boldsymbol{M}_1$ annihilates $\boldsymbol{X}_1$. To see that the last term also drops out, observe that

$$\boldsymbol{M}_{\boldsymbol{X}}\boldsymbol{M}_1\boldsymbol{X}_2 = \boldsymbol{M}_{\boldsymbol{X}}\boldsymbol{X}_2 = \mathbf{O}. \tag{2.44}$$