b. Proof of the theorem:

Let $y_i = \beta_0 + \beta_1 x_{1i} + ... + \beta_k x_{ki} + ... + \beta_K x_{Ki} + e_i$ be a regression model of independent variables x onto y. For any k, regress x_{ki} onto the remaining independent variables so that we have $x_{ki} = \gamma_0 + \gamma_1 x_{1i} + ... + \gamma_{k-1} x_{k-1i} + \gamma_{k+1i} x_{k+1i} + ... + \gamma_K x_{Ki} + f_i$. Define $\tilde{x_{ki}} = x_{ki} - \hat{x_{ki}}$, where $\hat{x_{ki}}$ are the fitted values of x_{ki} from the auxiliary regression above.

We want to show that $cov(y_i, \tilde{x_{ki}}) = \beta_k \cdot var(\tilde{x_{ki}})$. So we take a look at the left-hand side:

$cov(y_i, \tilde{x_{ki}}) = cov(\beta_0 + \beta_1 x_{1i} + ... + \beta_k x_{ki} + ... + \beta_K x_{Ki} + e_i, \tilde{x_{ki}})$

By definition, $\tilde{x_{ki}} = (\gamma_0 - \hat{\gamma_0}) + (\gamma_1 - \hat{\gamma_1})x_{1i} + ... + (\gamma_{k-1} - \hat{\gamma_{k-1}})x_{k-1i} + (\gamma_{k+1} - \hat{\gamma_{k+1}})x_{k+1i} + ... + (\gamma_K - \hat{\gamma_K})x_{Ki} + f_i$

Then we use the property of covariances: $cov(Y, Z) = E(YZ) - E(Y)E(Z)$:

$cov(y_i, \tilde{x_{ki}}) = E(y_i \tilde{x_{ki}}) - E(y_i)E(\tilde{x_{ki}}) = E(y_i \tilde{x_{ki}})$ because $E(\tilde{x_{ki}}) = 0$ since $E(\hat{\gamma_i}) = \gamma_i$ and $E(f_i) = 0 \ \forall i$

Expanding, $E(y_i \tilde{x_{ki}}) = E([\beta_0 + \beta_1 x_{1i} + ... + \beta_k x_{ki} + ... + \beta_K x_{Ki} + e_i][(\gamma_0 - \hat{\gamma_0}) + (\gamma_1 - \hat{\gamma_1})x_{1i} + ... + (\gamma_{k-1} - \hat{\gamma_{k-1}})x_{k-1i} + (\gamma_{k+1} - \hat{\gamma_{k+1}})x_{k+1i} + ... + (\gamma_K - \hat{\gamma_K})x_{Ki} + f_i])$ (*)

We will multiply and expand the above expression (*) according to the property $E(X+Y) = E(X) + E(Y)$.

This gives us $E(y_i \tilde{x_{ki}}) = E(\beta_0 \cdot \tilde{x_{ki}}) + E(\beta_1 x_{1i} \cdot \tilde{x_{ki}}) + E(\beta_2 x_{2i} \cdot \tilde{x_{ki}}) + ... + E(\beta_k x_{ki} \cdot \tilde{x_{ki}}) + ... + E(\beta_K x_{Ki} \cdot \tilde{x_{ki}})$

Taking out constant terms of the expected value,

$E(y_i \tilde{x_{ki}}) = \beta_0 E(\tilde{x_{ki}}) + \beta_1 E(x_{1i} \cdot \tilde{x_{ki}}) + \beta_2 E(x_{2i} \cdot \tilde{x_{ki}}) + ... + \beta_k E(x_{ki} \cdot \tilde{x_{ki}}) + ... + \beta_K E(x_{Ki} \cdot \tilde{x_{ki}})$

All values will be zero, except for $\beta_k E(x_{ki} \cdot \tilde{x_{ki}})$. This is because the first term disappears from the expectation of $\tilde{x_{ki}}$, and all other terms can be combined in such a way as to write $\tilde{x_{ki}}$ as a linear combination of all other x_i.

Thus, $E(y_i \tilde{x_{ki}}) = \beta_k E(x_{ki} \cdot \tilde{x_{ki}})$

Since $E(x_{ki}|X_{-k}) = \hat{x_{ki}}$, meaning that the expected value of x_{ki} is evaluated using all other independent variables, excluding x_{ki}, we have $x_{ki} = E(x_{ki}|X_{-k}) + \tilde{x_{ki}}$

Now, $\beta_k E(x_{ki} \cdot \tilde{x_{ki}}) = \beta_k E(\tilde{x_{ki}}(E[x_{ki}|X_{-k}] + \tilde{x_{ki}}))$

$= \beta_k (E(\tilde{x_{ki}}^2) + E[(E[x_{ki}|X_{-k}]\tilde{x_{ki}})])$

$= \beta_k var(\tilde{x_{ki}})$, completing the proof. ■

P.S. – Maybe it's because I don't have a strong background in stats, but reading this made me conclude that economists should stick to not proving theorems.