1. 1. The potential outcomes for any unit do not vary with the treatments assigned to other units $Y_{ij}(d_1, \ldots, d_J, w_{11}, \ldots, w_{N_j,J}) = Y_{ij}(d_i, w_{ij})$, and, for each unit, there are no different forms or versions of each treatment level, which lead to different potential outcomes $(d_i, w_i) \in \{0,1\}^2 \quad \forall i$. SUTVA is not necessarily satisfied since spillover effects might ocure within villages (individuals share cash transfers) or across villages (fiscal transfers change general equilibrium environment).
   2. By randomization, the ATEs can be written in terms of momentsof the (Y, D, W) distribution

$$\begin{aligned}
\tau_d &= \mathbb{E}[Y_{ij}(1,0) - Y_{ij}(0,0)] \\
&= \mathbb{E}[Y_{ij}|(D_j, W_{ij}) = (1,0)] - \mathbb{E}[Y_{ij}|(D_j, W_{ij}) = (0,0)] \\
\tau_w &= \mathbb{E}[Y_{ij}(0,1) - Y_{ij}(0,0)] \\
&= \mathbb{E}[Y_{ij}|(D_j, W_{ij}) = (0,1)] - \mathbb{E}[Y_{ij}|(D_j, W_{ij}) = (0,0)] \\
\tau_{dw} &= \mathbb{E}[Y_{ij}(1,1) - Y_{ij}(0,0)] \\
&= \mathbb{E}[Y_{ij}|(D_j, W_{ij}) = (1,1)] - \mathbb{E}[Y_{ij}|(D_j, W_{ij}) = (0,0)]
\end{aligned}$$

   i.e. the differences in CEFs identify them since potentional outcomes are balanced across groups.
   3. The population estimator consisting of differences in CEFs corresponding to the identified ATEs is constructed above. Since $\mathbb{E}[Y_{ij}|(D_j, W_{ij})] = 0$ in the population regression equation

$$Y_{ij} = \mu + \tau_d D_j(1 - W_{ij}) + \tau_w W_{ij}(1 - D_j) + \tau_{dw} D_j W_{ij} + \epsilon_{ij},$$

   also subtracting $\mathbb{E}[Y_{ij}|(D_j, W_{ij}) = (0,0)] = \mu$ from each of the $\mathbb{E}[Y_{ij}|(D_j, W_{ij}) = (1,0)] = \mu + \tau_d$, $\mathbb{E}[Y_{ij}|(D_j, W_{ij}) = (0,1)] = \mu + \tau_w$, $\mathbb{E}[Y_{ij}|(D_j, W_{ij}) = (1,1)] = \mu + \tau_{dw}$ corresponds to the ATEs.
   4.

$$Y_{ij} = \mu + \alpha_j + \beta W_{ij} + \epsilon_{ij}$$

   Based on the perspective, treatment effects are either 1) random samples from a large population of treatments or 2) fixed, unknown quantities to be estimated.

   1) $\theta = (\mu, \sigma^2_\alpha, \sigma^2_\epsilon)$ estimate σ^2_α in a variance components model (e.g. by restricted ML or leave-one-out MM), test H_0: $\sigma^2_\alpha = 0$ against H_1: $\sigma^2_\alpha > 0$ using the ANOVA F statistic.

   2) $\theta = (\alpha_1, \ldots, \alpha_J, \sigma^2_\epsilon)$: estimate α_j in the equation above (omit grand mean/constant or include $J-1$ indicators) via OLS and test H_0: $\alpha_j = \alpha_k \quad \forall j,k$ by comparing the between/within group sum of squares ratio to the $F_{J-1,N-J}$ quantile critical value at the desired level of significance
   5.

```
set obs 100 // gen village = _n // gen fiscal = runiform() < .5
gen casheffect = rnormal(2,1) // expand 10 //
bys village: gen individual = _n // gen cash = runiform() < .5
gen po = rnormal(0, 1) // gen pofiscal = rnormal(1, 1)
gen pocash = rnormal(casheffect, 1) // cap program drop createoutcomes
program define createoutcomes, // syntax[, meaninteraction(real 0)]
  gen interaction = rnormal(`meaninteraction', 0.2)
  gen poboth = pofiscal + pocash + interaction
  gen fiscalcash = fiscal*cash
  gen outcome = (1-fiscal)*(1-cash)*po + fiscal*(1-cash)*pofiscal ///
  + (1-fiscal)*cash*pocash + fiscalcash*poboth // end
```

   By linearity of expectation, subtracting $\mathbb{E}[Y_{ij}|(0,0)] = 0$ from $\mathbb{E}[Y_{ij}|(1,1)]$ and using the notation from above: $\tau_{dw} = \tau_d + \tau_w + \mathbb{E}[X_{ij}]$. ϕ can therefore be interpreted as the average difference between the individual transfer effects and their total synergy effect / their mean interaction effect.

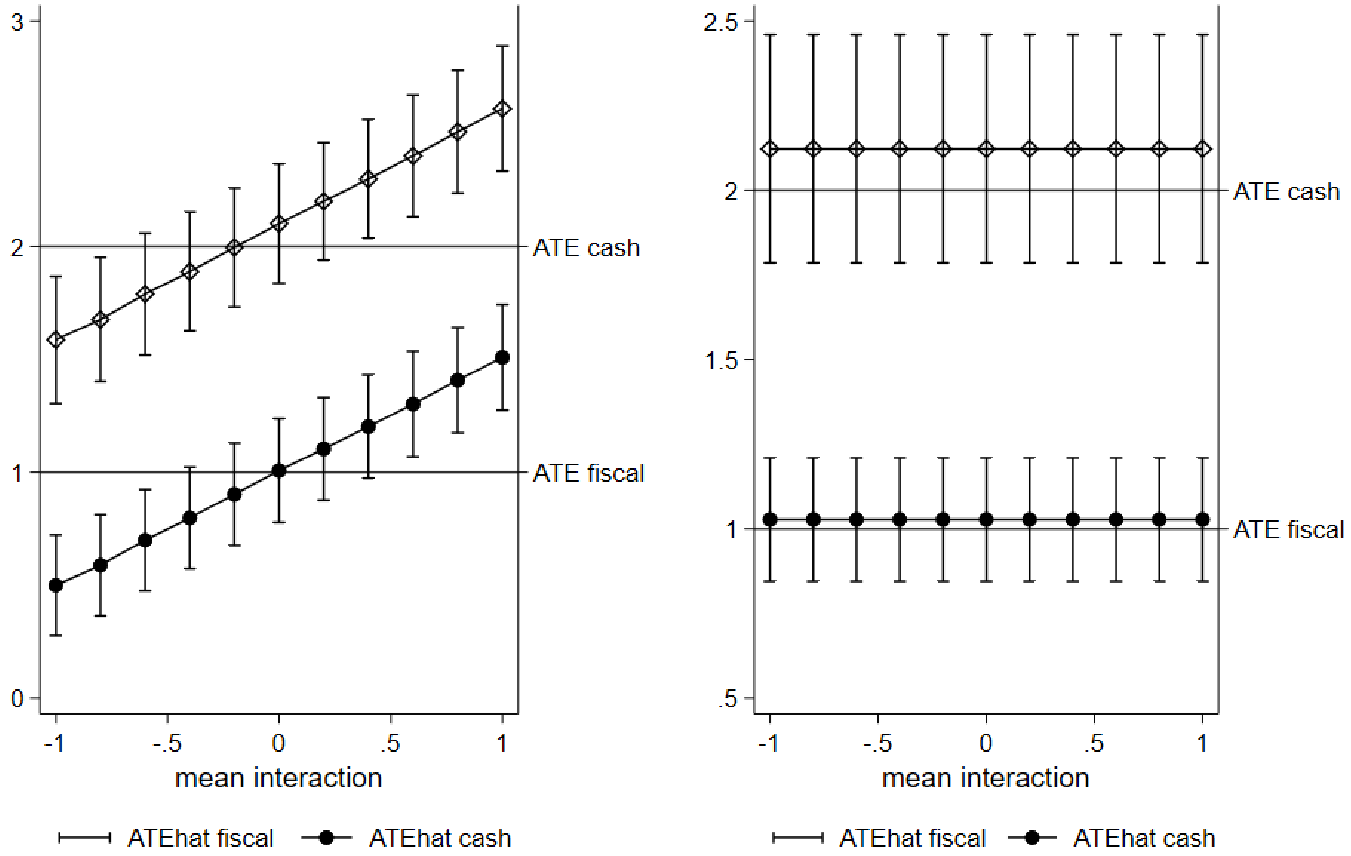
[1] mach5689@student.su.se

6.

```
forval meaninteraction = -1(.2)1 {createoutcomes, meaninteraction
(`meaninteraction') // reg outcome fiscal cash, cl(village)
reg outcome fiscal cash fiscalcash, cl(village) }
```

Figure 1 shows coefficient and onfidence intervals from regression specification including indicators (left), and additionally including interaction (right). Standard errors are clustered at the village level to adjust for within-village correlation of the fiscal transfer.

Figure 1: ATE estimates over ϕ



7.

```
forval meaninteraction = 0/1 {createoutcomes, meaninteraction
(`meaninteraction') // reg outcome fiscal cash, cl(village)
reghdfe outcome cash, absorb(village) cluster(village)}
```

The village fixed effects estimates (even columns) of the cash transfer coefficient do not change significantly from the indicator specification estimates (odd columns) presented in the table below.

	$\phi = 0$	$\phi = 0$	$\phi = 1$	$\phi = 1$
fiscal	1.013		1.516	
	(8.70)		(12.63)	
cash	2.107	2.103	2.618	2.616
	(15.70)	(15.18)	(18.40)	(18.27)

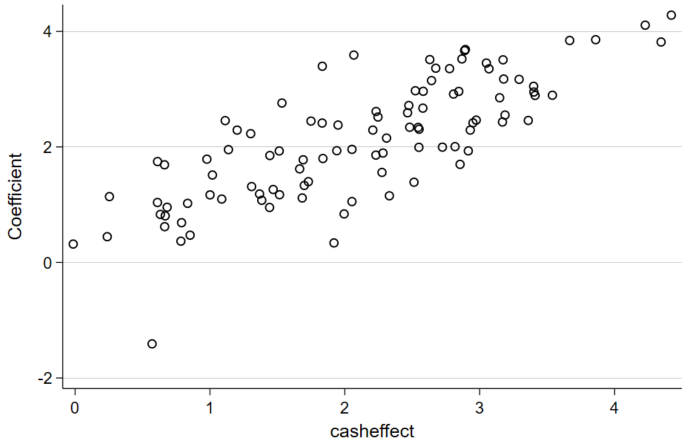
t statistics in parentheses

8.

```
qui createoutcomes, meaninteraction(0) // gen cashvillage = cash*village
reghdfe outcome i.cashvillage, absorb(village) cluster(village) nocons
```

Figure 2 compares the estimated village cash transfer effects (y-axis) to their true values (x-axis).

Figure 2: village effects: estimated over true



9. Estimating $\{\mu, \sigma^2_{Y|\alpha_j}, \sigma^2_{Y|\alpha_j}, p\}$ by the grand mean, "within" variance, "overdispersion" in averages and the sample propensity to receive cash transfer allows for shrinking the mean village cash effects towards a prior by imposing a parametric restriction

$$\bar{Y}_j|\alpha_j \sim N\left(\alpha_j, \frac{\sigma^2_{Y|\alpha_j}}{Np(1-p)}\right), \quad \alpha_j \sim N(\mu, \sigma^2_\alpha)$$

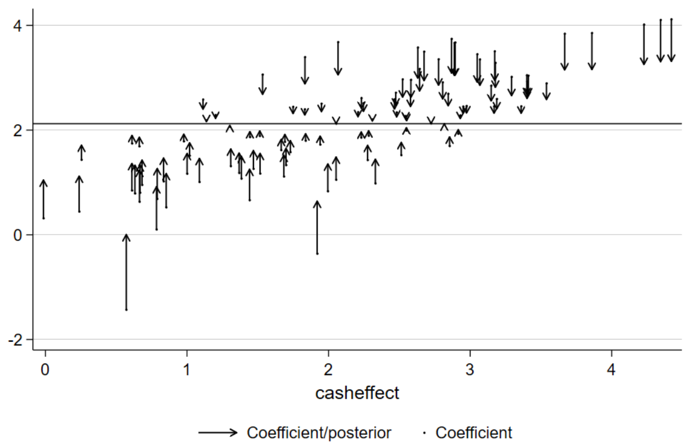
The posterior estimates

$$\hat{\alpha}_j = \hat{\mu} + \frac{\hat{\sigma}^2_\alpha}{\hat{\sigma}^2_\alpha + \frac{\hat{\sigma}^2_{Y|\alpha_j}}{10\hat{p}(1-\hat{p})}}(\bar{Y}_j - \hat{\mu})$$

are shown in Figure 3.

```
sum cash // scalar cashpropensity = `r(mean)'scalar cashvariance =
`=cashpropensity' * (1-`=cashpropensity') // scalar villagers = 10
qui reghdfe outcome, absorb(village cashvillage) // scalar withinvar
= `e(rmse)'^2 sum coef, d // scalar prior = `r(mean)' // reg coef
scalar overdispersion = `e(rmse)'^2 -
`=withinvar'/(`=villagers'*`=cashvariance')
scalar signaltonoise = `=overdispersion'/(`=overdispersion' +
`=withinvar'/(`=villagers'*`=cashvariance'))
gen posterior = `=prior' + `=signaltonoise' * (coef-`=prior')
```

Figure 3: Empirical Bayes



2. 1. $E[Y_{11}|\theta] = \alpha_1 + \gamma_1 + \psi_1, E[Y_{12}|\theta] = \alpha_1 + \gamma_2 + \psi_2, E[Y_{21}|\theta] = \alpha_2 + \gamma_1 + \psi_2, E[Y_{22}|\theta] = \alpha_2 + \gamma_2 + \psi_1$
   Under conditional mean independence of ϵ_{it}, θ is point-identified if each of the 6 parameters can be expressed from the 4 equation CEF system: at least 2 of the parameters need to be normalized. Else the rank condition is violated by multicolinearity, the individual effects cannot be separated.

   2. θ is not point-identified. For suppose it is: if time effects are normalized and each worker effect is lower than the machine effects, then the latter cannot be recovered (and vice versa), contradiction.

   3. Adding two CEFs $E[Y_{31}|\theta] = \alpha_3 + \gamma_1 + \psi_1, E[Y_{32}|\theta] = \alpha_3 + \gamma_2 + \psi_1$ with an additional parameter to the system in part 1, solving 6 equations with 7 unknowns requires one normalization. Answer to part 2 does not change, the fundamental problem remains and analogous argument still holds.

   4. $\theta = (\alpha_1, \alpha_2, 0, \gamma_2, 0, \psi_2) \rightarrow 2 = \alpha_1,\ 4 = \alpha_1 + \gamma_2 + \psi_2,\ 3 = \alpha_2 + \psi_2,\ 3 = \alpha_2 + \gamma_2 \rightarrow \theta = (2, 2, 0, 1, 0, 1)$

   5. ```
      set obs 200 // gen period = _n // gen gamma = rnormal(1, sqrt(.5))
      expand 10 // bys period: gen worker = _n // bys worker: gen alpha
      = rnormal(0,1) // replace alpha = alpha[_n-1] if worker == worker[_n-1]
      gen machine = runiformint(1, 20) // bys machine: gen psi = rnormal(1,1)
      replace psi = psi[_n-1] if machine == machine[_n-1]// gen epsilon
      = rnormal(1, sqrt(.2)) // gen output = alpha + gamma + psi + epsilon
      ```

   6. OLS requires two conditions for identification of fixed effects:
   1) exogeneity holds since all RHS variables are iid and machines are randomly assigned by design
   2) no multicollinearity requires normalization, e.g. setting $\alpha_i = \psi_j = 0$

   7. ```
      reghdfe output i.worker i.machine, absorb(period)
      ```

   Simultaneou estimates are presented in Figure 4. Note lack of normalization resulted in ommiting the first category and estimated coefficients being on a different scale, but with consistent ordering.

   8. $T = 5$ estimates in 5 exhibit reduced precision. Furthermore, only 16 machine estimates were obtained since chance played its role in the assignment process and 3 categories were not observed.

   9. $j(i,t) = f(\alpha_i, \epsilon_{it}) \implies Pr(J(i,t) = j|\epsilon) \neq Pr(J(i,t) = j)$ Violation of exogenous mobility from quality matching implies the parameters in a connected-set fixed effects are not identified: the worker and machine effects are not simply additive anymore: impossible to separetely estimate.

3. 1. Regular DiD $Y_{it} = D_i \cdot Y_{it}(1) - (1 - D_i) \cdot Y_{it}(0), \quad Y_{i,t-1} = Y_{i,t-1}(0)$
   $y_{it} = \alpha + \gamma_i \cdot D_i + \lambda_t \cdot \text{Post}_t + \tau \cdot D_i \times \text{Post}_t + \varepsilon_{it}$

   - Nobody treated at $t - 1$ and some people treated at t.
   - Parallel trends $\mathbb{E}[Y_{it}(0) - Y_{i,t-1}(0)|D_{it} = 1] = \mathbb{E}[Y_{it}(0) - Y_{i,t-1}(0)|D_{it} = 0]$
   - If parallel trends holds: $\tau = ATT = \mathbb{E}[Y_{it} - Y_{i,t-1}|D_{it} = 1] - \mathbb{E}[Y_{it} - Y_{i,t-1}|D_{it} = 0]$

   Writing $Y_{it}(D_{it})$ requires no contamination across units (SUTVA) and no contamination across periods (anticipation/memory). $\tau \equiv \mathbb{E}[Y_{it}(1) - Y_{it}(0)|D_{it} = 1] = \mathbb{E}[Y|D_{it} = 1] - \mathbb{E}[Y_{it}(0)|D_{it} = 1]$ Parallel trends assumption $\mathbb{E}[\Delta Y_{it}(0)|D_{it} = 1] = \mathbb{E}[\Delta Y_{it}(0)|D_{it} = 0]$ allows for identification by imputing the counterfactual trend of the treated group: $\tau = \mathbb{E}[\Delta Y_{it}|D_{it} = 1] - \mathbb{E}[\Delta Y_{it}|D_{it} = 0]$
   TWFE $y_{it} = x_{it}\beta + \tau_{it} \cdot D_{it} + \eta_i + \lambda_t + \varepsilon_{it}$ delivers the $ATT = \mathbb{E}[\tau_{it}|D_{it} = 1]$ if there are only two periods with $D_{i0} = 0$ for all i or constant treatment effects $\tau_{it} = \tau$. Since how τ_{it} varies over time for different cohorts now affects the estimate of τ. Heterogeneity robust estimators are therefore better suited for the staggered rollout considered in this exercise. The causal effect of interest, as proposed in dCdH(2020), is therefore arguably $\tau_S \equiv \mathbb{E}\left[\frac{1}{N_S}\sum_{(g,t):t\geq 2, D_{it}\neq D_{it-1}}(Y_{it}(1) - Y_{it}(0))\right]$, where $N_S = \sum_{(g,t):t\geq 2, D_{it}\neq D_{it-1}}$ are observations in switching cells .

   2. FWL $\tilde{D}_{gt} = D_{it} - \bar{D}_g - \bar{D}_t + \bar{D} \implies \tilde{D}_{33} = 1 - 1/2 - 1 + 1/2 = 0, \tilde{D}_{34} = 1 - 1/2 - 1/2 + 1/2 = 1/2, \tilde{D}_{22} = 1 - 1/2 - 1/2 + 1/2 = 1/2, \tilde{D}_{23} = 1 - 1/2 - 1 + 1/2 = 0$ Theorem (dCdH 2020):

$$w_{33} = w_{23} = \frac{0}{1/4} = 0, w_{34} = w_{22} = \frac{1/2}{1/4} = 2$$

3. $\hat{\tau}_{dCdH}$ estimates the ATE of switching cells. In staggered adoption designs, it is the average of the treatment effect at the time when a group starts receiving the treatment, across all groups that become treated at some point. It leverages "joiner-effect": [0, 1] against [0, 0] and "leaver-effect": [1, 0] against [1, 1] and avoids the rest. It requires requires that no group appears or disappears over time. It requires requires that units' treatments do not vary within each (g, t) cell. It requires that the shocks affecting a group's $Y_{g,t}(1)$ be mean independent of that group's treatment sequence. It requires that between each pair of consecutive periods, the expectation of the outcome with treatment follow the same evolution over time in every group: the ATE follows the same evolution over time in every group. It requires that between each pair of consecutive time periods, if there is a "joiner", then there should be another group that is untreated at both dates, and that between each pair of consecutive time periods, if there is a "leaver", then there should be another group that is treated at both dates. It requires Mean Independence between a Group's Outcome and Other Groups Treatments. Then $\mathbb{E}[\hat{\tau}_{dCdH}] = \tau_S$.
4. $\sum_{t=2}^{T}(1/2\text{DID}_{+,t} + 1/2\text{DID}_{+,t}) = (1/2(3-1)+0) + (0+0) + (0+1/2(-2-(-1)) = 1 - 1/2 = 1/2$

Figure 4: $\hat{\alpha}_i, \hat{\psi}_j$ over α_i, ψ_j : $T = 200$

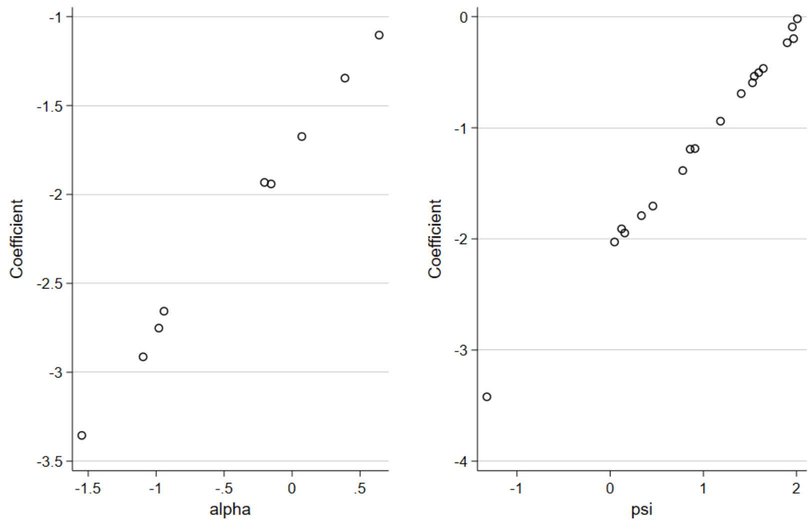


Figure 5: $\hat{\alpha}_i, \hat{\psi}_j$ over α_i, ψ_j : $T = 5$