

Evaluating SME Programs in Mexico Using Panel Firm Data

Hong Tan and Gladys Lopez-Acevedo
The World Bank

SME Evaluation Workshop
Mexico City, September 23-24, 2004

Goals of Presentation

- ☛ Describe two impact evaluation exercises
 - Re-evaluation of CIMO / PAC program of STPS
 - Using 2001 ENESTYC to compare program impacts
- ☛ Highlight evaluation approaches and the issues they try and address
- ☛ Report some tentative findings
- ☛ Suggest ways of improving program evaluations

I. Re-evaluation of CIMO/PAC

- Overview of CIMO/PAC program
- 1995 & 1997 Impact Evaluation Studies
- Re-evaluating the impacts of CIMO/PAC
 - Positive intermediate outcomes, negative productivity impacts
 - Selectivity bias from weak firms participating in CIMO/PAC, non-comparable control group
 - Using difference-in-differences approach
 - Results and suggestions

Overview of CIMO/PAC

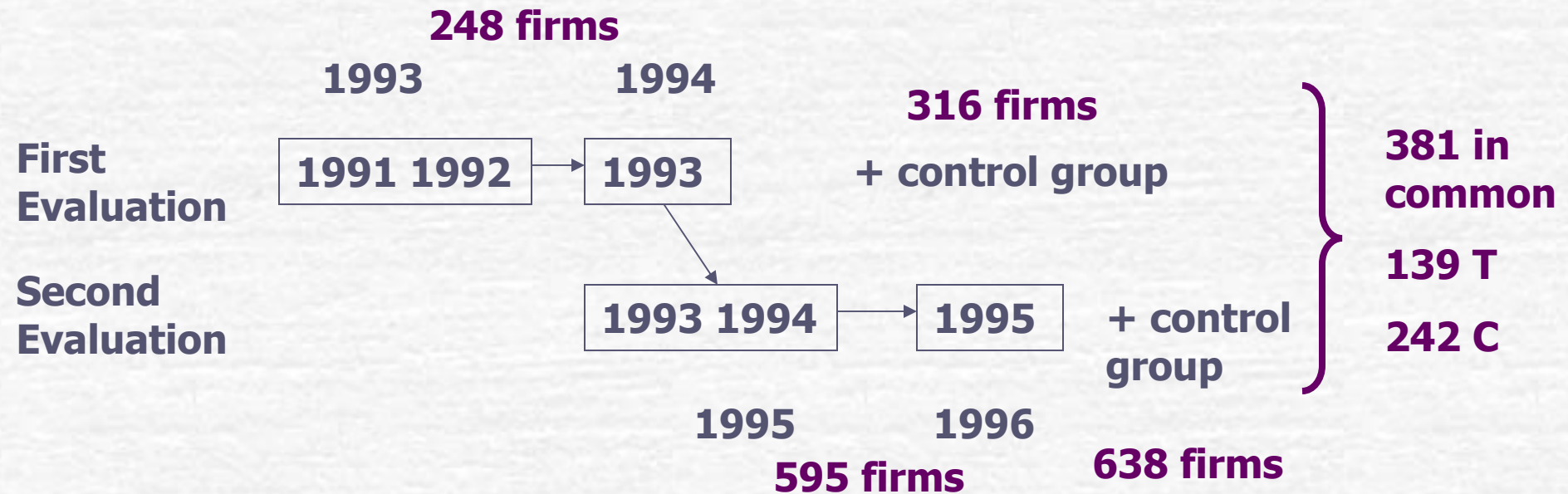
- Subsidizes training provision and other support services to MSMEs by public or private providers
- Promoters do diagnostic to identify production and skills constraints – training & technical assistance offered on a cost-sharing basis
- In 2001, CIMO/PAC provided support to 94 thousand firms (3% of Mexican firms), benefiting 333,500 workers

Previous Evaluations

STPS 1995 and 1997

- ☛ Quasi-experimental design
 - treatment group from CIMO/PAC program and a matched control group that did not participate, but otherwise similar in firm size, sector, and geographic location
- ☛ Surveys applied by CIMO/PAC to treatment group, and by INEGI to the control group

Data Collection Strategy 1995 & 1997 CIMO Studies



Summary of Previous Studies

- ☛ In comparison to control group, both studies found positive impacts on intermediate outcomes:
 - treatment group more likely to provide training, higher training spending per worker, introduced organizational changes, and implemented quality control systems
- ☛ But negative impacts on productivity levels in the treatment group as compared to the control group
 - Both studies found lower productivity levels in the treatment group than in the control group
 - Production functions estimated on post-program data do not take this into account

Re-evaluation of CIMO/PAC

- Challenge: resolving apparent contradiction between positive intermediate outcomes but negative final program impacts on firm performance
- Re-examination of CIMO panel data
 - cleaning and making data comparable over time
 - addressing selectivity bias - CIMO attracting weaker firms into program than other SMEs
 - using a difference-in-differences approach that fully uses the panel data

A First Look at the Data

- Program improved intermediate outcomes of participating firms relative to the control group

Variable	Period	Change Over Time
Mean private investment (per worker) in training	1991-1993	192.7
Mean private investment (per worker) in training	1993-1995	170.0
Share of firms with quality control mechanisms	1994-1995	23.1

Note: All estimated effects are statistically significant.

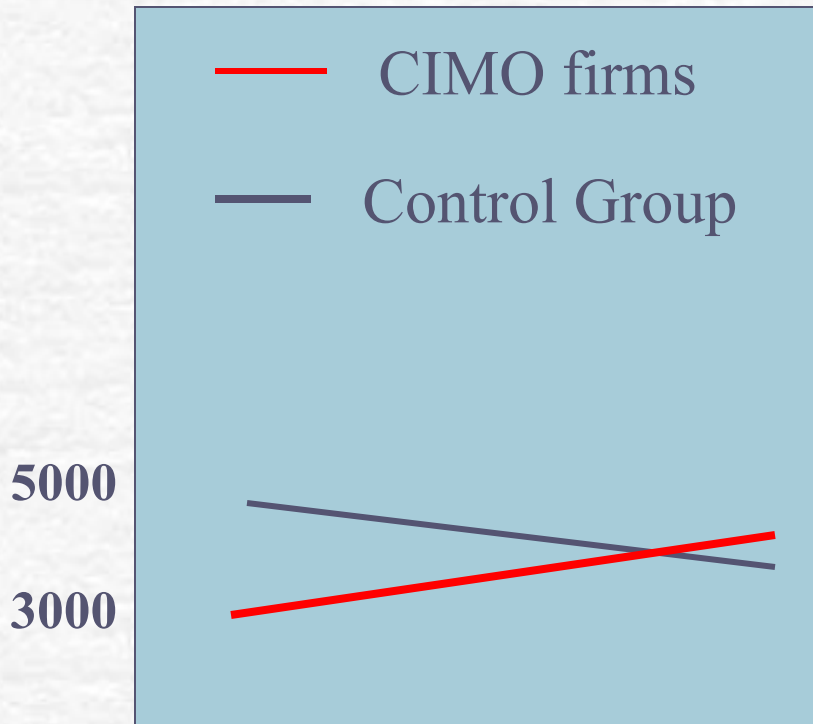
Group Means in Labor Productivity

	Value-Added per Worker			Production minus Materials per Worker		
	1991	1993	1995	1991	1993	1995
<i>Firms only in the first study</i>						
Control Group	47,391	34,964		63,898	49,140	
CIMO/PAC	31,741	35,146		35,078	37,024	
<i>Firms only in the second study</i>						
Control Group		178,734	207,989		186,299	225,120
CIMO/PAC		79,704	84,114		71,735	78,964
<i>Firms in both studies</i>						
Control Group	46,738	65,459	128,253	57,752	55,658	113,618
CIMO/PAC	30,874	51,815	51,372	38,174	38,884	51,203

In 1994 prices

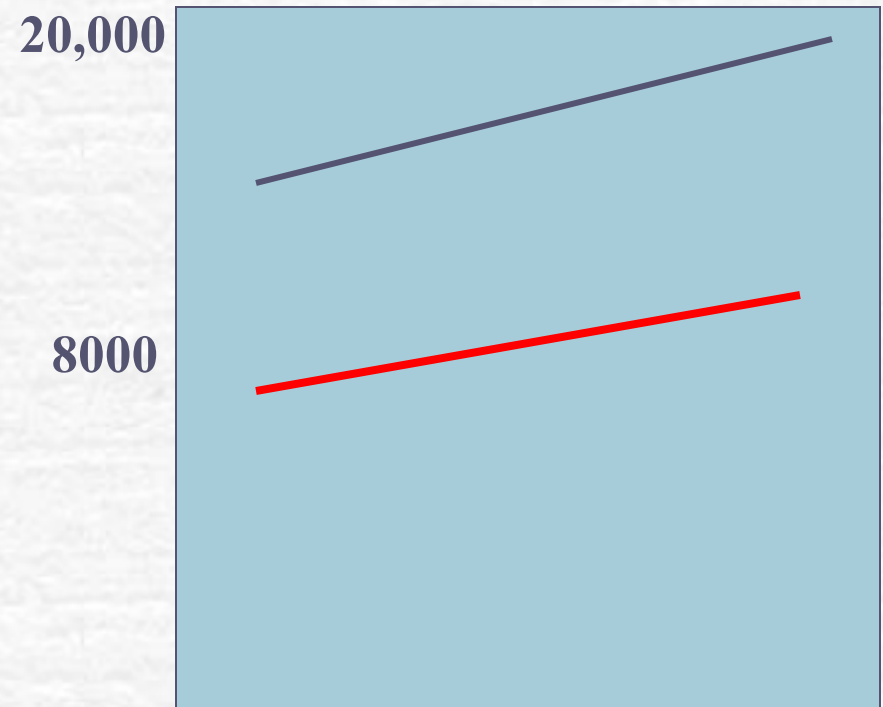
Productivity Growth CIMO and Treatment Groups

Value-added per worker



First study 1991-1993

Value-added per worker



Second study 1993-1995

Selectivity Bias

CIMO and Non-CIMO firms not directly comparable

- Control groups have **higher productivity levels** than CIMO firms with similar observable attributes
- Due to selection of weak firms into CIMO, or poor choice of control group, or both.
- A productivity regression will generally yield a **negative coefficient** (impact) on a CIMO indicator variable
- **The solution**: estimate a “fixed effects” or “first difference” model, remove level differences between the two groups, study changes over time in outcomes

Addressing Selectivity Bias

Estimating CIMO effects θ in a “levels” and “first differenced” production function

Levels Model

$$\text{Log}(VAt) = \alpha \text{Log}(Kt) + \beta \text{Log}(Lt) + \theta \text{CIMO}$$

VA=value-added, K=capital, L=labor, t=year

First Differenced Model

$$\Delta \text{Log}(VAt) = \alpha \Delta \text{Log}(Kt) + \beta \Delta \text{Log}(Lt) + \theta \text{CIMO}$$

$$\Delta = X_t - X_{t-1}$$

Levels versus Differences

Impacts on productivity & productivity growth

	Value Added		Production minus Materials	
	Levels	Differences	Levels	Differences
1991-1993	-0.098	0.058 *	-0.184 *	0.108 *
1993-1995	-0.527 *	-.051	-0.607 *	-0.004
1991-1995	-0.485 *	-.006	-0.574 *	0.019

* Denotes significance at the 5% level.

Summary of Results and Conclusions

- CIMO/PAC has positive effects on intermediate outcomes – training, organization change, QC
- Negative impacts on productivity attributable to selectivity bias and choice of control group
- Positive impact on productivity growth in the 1991-1993 period, but not in 1993-1995.
- LESSONS: The critical importance of
 - Selecting an appropriate control group
 - Addressing selectivity bias in program participation

Suggestions

- CIMO/PAC quasi-experimental design a good model to use to evaluate impacts of specific SME programs
- Time-line of 2 years to collect pre- & post-program data on treatment and control groups, 6 months to 1 year for analysis – plan and budget accordingly
- Design surveys to collect information specific to programs and common outcome or performance indicator variables – for comparability
- Reports to include details on data collection and analytic methods – for transparency

II. Evaluations Using ENESTYC

- ☛ Objectives:
 - Investigate potential of ENESTYC for impact evaluations and comparisons of different SME programs
 - Testing different impact evaluation approaches
- ☛ Overview of 2001 ENESTYC and SME module, and links of 1995 and 1999 ENESTYC
- ☛ Some tentative results on program impacts and implications for evaluation studies
- ☛ Suggestions for improving usefulness of future ENESTYC surveys

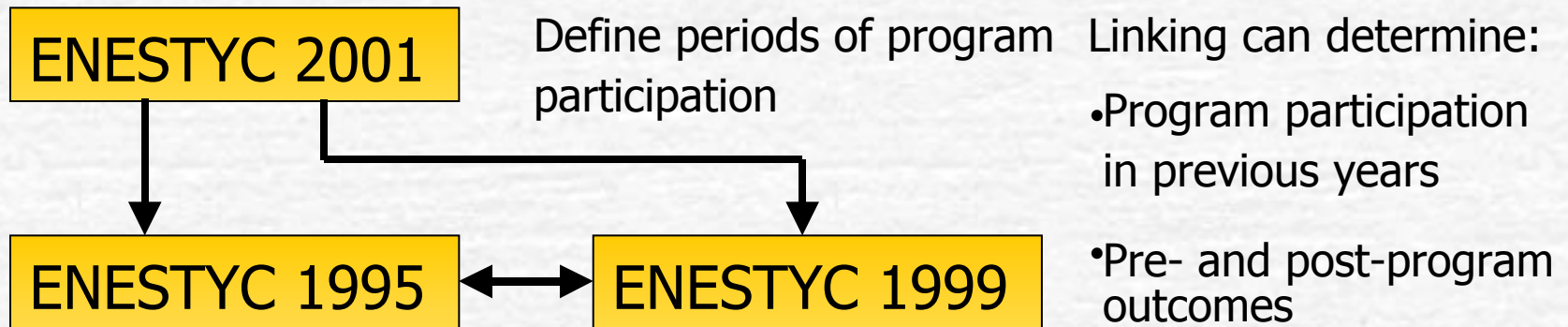
2001 ENESTYC Survey

- Fielded by INEGI with over 8,000 firms
- Firm-level information on ownership, employment, location, workforce attributes, wages, production, technology, workplace practices, and training

2001 ENESTYC Survey

- Fielded for STPS by INEGI with over 8,000 firms
- Firm-level information on ownership, employment, location, workforce attributes, wages, production, technology, workplace practices, and training
- SME module: retrospective questions on
 - **10 major SME programs**
 - **CONOCER, CIMO, COMPITE, CRECE, FIDECAP, FAMPYME, MEX-EX, PATCI, PMT, PCI, PAIDEC (on average 400 firms in the largest programs)**
 - **familiarity, participation, date started, form of participation**

Links with 1995 and 1999 ENESTYC



Linking to create panel data allows:

- Identification of pre- and post-program periods
- Selection of control group from large pool of non-participants
- Estimation of impacts on performance over time
- Control for unobserved heterogeneity and selectivity bias

Data and SME Programs Studied

- Limitations of ENESTYC: Random sampling produces small samples of program beneficiaries when linked to earlier ENESTYC surveys
- Focus on the three largest programs – CIMO, COMPITE and CRECE – with the largest sample sizes of program beneficiaries
- CIPI administrative data base were used to augment self-reported participation information from the 2001 ENESTYC to increase sample sizes

Program Participants

Panel Years	CIMO	CRECE	COMPITE	CIMO, CRECE, or COMPITE
1995 and 2001	160	33	59	215
1999 and 2001	299	60	128	419
1995, 1999 and 2001	141	26	47	186
2001	551	165	278	831

Source: Estimates from ENESTYC databases and CIPI administrative records

Methodology

- Begin with two simple approaches:
 - mean values of key outcome measures of the treatment and control groups
 - production functions to measure impacts on productivity, controlling for firm-specific effects
- Naïve approaches subject to several limitations, but they provide useful initial insights into the program impacts

Test for Differences in Means

Treatment versus Control Groups

	CIMO			CIMO, COMPITE, CRECE		
	1995	1999	2001	1995	1999	2001
Value Added	4,034	7,399	3,560	-633	182	2,743
Employment	59.5	125.3	56.2	20.5	67.4	38.0
Wage of Directors	-2.1	-857.2	-1,057.2	-2.4	-857.2	-512.0
Wage of Employees	-0.4	-533.2	-986.2	-0.3	-534.1	-850.4
Export Share	-3.4	4.5	3.6	-4.0	9.2	5.0
Export orientation (%)	-4.2	1.9	2.1	-5.0	8.6	4.0
Provide training (%)	13.1	38.5	66.4	11.2	38.4	53.6
In-house training (%)	-20.9	1.8	18.2	-18.0	7.9	15.4
External training (%)	21.1	35.7	48.0	18.2	29.6	38.0
Technology Adoption (%)	-9.1	23.4	31.1	-1.7	21.6	31.9
Quality control (%)	0.7	10.4	8.4	-0.7	4.9	8.2

Bold: significant at the 10% level

Production Function Estimates in Levels and First Differences

	Levels		First Differences	
	CIMO	CIMO, CRECE or COMPITE	CIMO	CIMO, CRECE or COMPITE
Log(Fixed Assets)	0.333	0.331	8.16E-08	8.23E-08
Log(Labor)	0.541	0.544	0.304	0.302
Program Participation	0.024	-0.002	0.001	0.003
Treatment	0.198	0.153	0.000	0.019
Observations	2339	2418	1646	1712

Bold: significant at the 5% level

Methodology

- Propensity Score Matching addresses inappropriate choice of control group
- Differencing addresses potential selection bias associated with program participation

Propensity Score Matching

- Duration of the pre- and post-participation period varies across cohorts from 3 to 6 years (outcomes may only appear with a time lag)

Cohort	Participation	Pre-participation	Post-participation
1995-2001	After 1995	1995	2001
1999-2001	After 1999	1999	2001
1995-1999	After 1995 and before 1999	1995	1999

Propensity Score Matching

- ☞ Match each of the cohort treatment groups with a control group using one summary indicator
- ☞ Logit model to predict program participation
- ☞ The indicator is the predicted probability or propensity score the that a firm would participate
- ☞ Variables included in model:
 - economic sector, state, firm size, age of the firm, share of permanent workers, share of unskilled labor, and fixed assets per worker

Propensity Score Matching

- The matching was based on *pre-program* participation characteristics for each cohort
- Matching algorithm was the method of **nearest neighbor with equal weights**
- Impacts estimated using the difference-in-differences (DID) approach

Estimated Program Impacts (DID)

	Cohort 1999-2001		Cohort 1995-2001	
	CIMO, CRECE, or COMPITE	CIMO	CIMO, CRECE, or COMPITE	CIMO
<i>Final Outcomes</i>				
Value added	1,899	16,962	-24,300	2,796
Employment	-17.1	11.6	-41.3	-80.1
Wage of Directors	2	0.2	0.1	2.7
Wage of Employees	16.3	-0.1	0.1	0.8
Export orientation (%)	-1.1	-1.7	-5.3	-6.4
Export share (%)	-1.7	-88.7	-2.2	-2.2
<i>Intermediate Outcomes</i>				
Formal training (%)	8.9	6.8	6.8	10.3
Internal training (%)	-5.9	1.6	-9.3	4.5
External training (%)	14.8	4.5	10.5	3
Technology adoption (%)	1.8	-4.2	4.8	28.2
Quality control (%)	-3.3	-11	9.7	2.6

Bold: significant at the 5% level

Summary and Implications

- Programs appear to have positive impacts on intermediate outcomes, impacts on final outcomes are still elusive
- May be due to small sample size; or may suggest need to improve program design and delivery, and if warranted, even consolidation or termination of some non-performing programs
- Future research
 - may need more sophisticated methods than a time-invariant DID estimator
 - estimate the effects of differential treatment doses

Suggestions

- ENESTYC potentially useful vehicle for impact evaluations of specific programs and cross-program comparisons
- Need to add purposive sample to augment sample sizes of program beneficiaries, from CIPI data base
- Augmenting ENESTYC sample requires additional budget, contributions from different agencies?
- Greater coordination and knowledge-sharing across SME programs, of evaluation methods and lessons learnt
- Cross-program comparisons not a replacement for program specific evaluations and continuous monitoring