VANUS INVESTMENTS LIMITED

THE DESIGN OF CARNIVAL POLICY AND THE CHALLENGE OF CAUSAL ANALYSIS

Vanus James, Lead Consultant
Vanus Investments Ltd — 15 Henry Rd, Tunapuna, Trinidad
T: (868)663-0500 **E:** vanus.james@gmail.com

0. Introduction

Efforts to design industrial policy in social systems require three distinct types of model-based information. The first is a set of equations (or some other conceptual device) describing the time-evolution of the system. The second is an adequate estimate of the values of the causal parameters of the system. The causal parameters of concern link the rates of change of characteristic state (effect) variables to themselves and to other causal forcing functions, at the level of society as a whole, in its composite sectors, and their constituent operators (household, individuals and establishments/firms). The sector of greatest interest here is carnival, conceived as a satellite sector. The third type of required information is the set of data, including initial/boundary conditions, emanating from adequately randomized experiments, quasi-experiments and instruments. Even if such information were known with certainty, which is typically not the case, the development trajectory is usually not similarly known with certainty. Uncertain knowledge usually reflects the influences of uncontrollable external or endogenous unobserved forces that are changing in uncontrollable and random ways and that also cause the behavior of the state variables of the system to be random. Uncertain knowledge also reflects the fact that development is a complex dynamic social process, set in the global environment, engineered by many interacting sectors, run by free-thinking individuals, establishments and organizations whose behaviors cannot be fully controlled, even though their average behavior can be. In such systems, there are many free variables – degrees of freedom – whose dynamic activity again leads to random fluctuations in the behavior of the system.

Even beyond lack of knowledge and complexity, development involves a Schumpeterian tendency to creative destruction, in which small adjustments in the underlying technological and organizational parameters of society tend to destroy old sectors and create new free variables, which is to say new free state variables and related new degrees of freedom. These then cause large and cumulative changes in the quantitative *and* qualitative behavior of its characteristic social and economic state variables whenever they are stimulated.[1] Consequently, in a developing society, cause is not tied to effect by constants of proportionality but rather by variable parameters of disproportionality that generate irregular nonlinearity of the sort that does not permit linearization. The irregular nonlinearity usually occurs even when the parameters change in reasonably deterministic ways. One normally wants the change to be favorable but must admit that it might often be unfavorable. Measurement of such changes must take place in accordance with the principles of causal analysis if the measurement is to support sound policy.

Science and culture are the twin dynamic information foundations of the Schumpeterian dynamics. Science refers to a method of knowing – the scientific method, featuring proof of claims of cause and effect based on evidence and logic that emanate from randomizing experiments, quasi-experiments, and credible instruments. The claims may come from culture. Culture refers to beliefs, commonsense-based social information systems, practices, technological traits and skills of any definable social group. These normally translate into activities and products that provision society. Science and culture jointly forge a natural process of creating technology through these dynamic interfaces. Here technology refers to the use of science and culture in production

[1] In linear systems, when the parameters change the quantitative behaviour also changes but the qualitative behaviour, i.e., the type of trajectory, does not change.

engineering, including in creative expression. Culture, science and technology, and hence also economy, co-evolve, and the dynamic is influenced by environment and geography.

We may not yet have documented human society long enough to determine which is more important. However, we know the comparative roles. Culture provides much of the stylized facts and the common sense that defines the premises that can be tested by scientific method, alongside a basket of activities and products designed to provision society. The scientific method, especially in its expression as basic research, yields the most important advances in both activities and products to provision society. Most interesting perhaps, dynamic interface is the norm. Culture provides a vast array of activities and products that can be the target of certain kinds of fusion: (i) Innovation with science; (ii) innovation through cultural fusion; and (iii) the dynamic interfaces of both. Dynamic interface is the norm. In terms of relative importance, the GOTT has settled the issue for the moment by agreeing in 2012, at the Rio+20 negotiations on social progress that the role of science and democratic policy design should be sharply increased in efforts to develop our society. Policy can shape both process and outcome. Culture shapes science and technology; and science and technology shape culture. There is also dynamic feedback and there are chaotic or random outcomes. So, the interacting cultures are all co-evolving. A key to sound policy is to facilitate the networking. There is a Schumpeterian dynamic at work, as the resulting scientific and technological progress make some cultural forms obsolete while creating new ones – creative destruction. Use of culture in economic activity is a major way to solidify a culture's local and global position, but the thrust in this direction will have to be backed by a sound process of financial and real investment in indigenous creativity, as well as a sound exporting process. Trade facilitation, branding and marketing, and investment promotion are all important in the general scheme of things. In the case of Trinidad and Tobago, because of small size, high emphasis must be placed on export to the centres of high demand – beyond the boundaries of one's immediate culture. In this respect, emerging economies are an important target group.

Perhaps the most interesting stimulus to causal variables is socially engineered – i.e., policy to adjust causal factors. Policy-makers must invest in knowing the effects of such interventions, so we discusses this challenge in some detail as an aspect of causal analysis. The overarching question of policy in this case is, what transformations in the causal parameters of the carnival sector would follow from such interventions that would also grow the wider economy and extract economic and social benefits on a sustainable basis; and how are these ordered in terms of impact? From the proactive social perspective, for example, what are the new investment initiatives, supporting policies, and related monitoring indicators on which progress must be promoted in carnival to address Trinidad and Tobago's most urgent social need, poverty reduction and equity? Often, to answer such a question, we also need to investigate the same question for the economy as a whole since that provides the mechanism and much of the environment of propagation of the effects of changes in the causal parameters of a sector. So, we use a multi-sector model of poverty-reduction and economic growth, with equity and inflation. The model of economic growth introduces the sector specific effects, and the sectors specified include the traditional manufacturing sector and all sectors in which carnival plays a definitive role.

It is with this kind of information that analysts attempt causal analysis to support the policy process. In terms of structure, this report first clarifies the interplay of culture, economic progress and social progress in the dynamics of a society and provides some background economic data in the context of which the issues of causal policy analysis arise. Next, the report clarifies the concept of causal analysis and the main method of investigation of policy issues, some of which are identified throughout the presentation for the purpose of illustration. Then, the report addresses the central issue of selection bias and endogeneity, and the related need for random sampling on the causal variables as the primary solution of selection bias. The role of control instruments to be held fixed is explained and also tied to random sampling. The note also explains the centrality of conditional expectation functions in causal analysis and the usefulness of regression analysis as a scientific approximation of such functions. The rest of the paper addresses the manifestation of selection bias in misspecification in regression, including problems such as the omitted variable bias. It motivates the exhaustive search for control instruments to address selection bias. It also motivates the use of fixed effects models, differences in differences models, with and without time trends, and Granger causality as alternative ways of addressing selection bias. Throughout the report, we illustrate how the problems arise with specific policy issues that can be addressed in the process of data collection and analysis. The NCC is well-placed to consider and require adequate scientific design on many of these issues. It is also well-placed to require adequate policy democratisation as a condition for arriving at a more adequate determination of cause.

1. Background Conditions

Sector-specific industrial policy rests on the confidence that data are available about the fundamental initial or boundary conditions of the economy, indicating which sectors possess the technology, productivity and growth trajectories that can address those challenges. This background provides context for the main estimates of the forces affecting growth of output **per capita** in the carnival industries, and hence the contribution of carnival industries to social and economic well-being in Trinidad and Tobago. Crucial among these are the efficiency of foreign exchange use by the sectors and the general growth performance of the economy as a whole as measured by the growth of the GDP or growth of the GNP. The analysis provides data on the size, structure and growth of the GDP from the perspective of the categories of standard national accounting. It also includes data from the supply and use tables, indicating the comparative output per dollar of foreign exchange used by the sectors of the economy, including some sectors (activities) that are considered aspects of the carnival industries.

Population and Employment

Table 1.1 documents the population trends of Trinidad and Tobago since 2000, and Table 1.2 describes corresponding growth patterns. The general tendency is that the rate of population growth is very slow, while the rate of growth of the labour force and employment declined simultaneously. Overall, this has meant falling unemployment, but it is well-known that these employment patterns are highly sensitive to emigration and exogenous shocks, especially shocks to oil and gas prices and hence shocks to government's net revenues and budget balance.

Table 1-1: Population, labour force and employment

	Trinidad & Tobago			Tobago		
Year	POPULATION (000)	Labour Force (000)	Employed (000)	POPULATION (000)	Labour Force (000)	Employed (000)
2000	1,262.40	572.8	502	54.1	22.8	21
2001	1,266.80	576.4	514	54.9	24.8	21.7
2002	1,275.70	586.2	525.1	55.3	25.2	23
2003	1,282.40	596.5	534.1	55.6	26.3	24.1
2004	1,290.60	613.4	562.2	56	27	25.9
2005	1,294.50	623.7	574	56.4	29.8	28.4
2006	1,297.90	625.2	586.2	56.8	28.7	27.2
2007	1,302.20	622.4	587.8	57.3	28.7	27.6
2008	1,308.60	626.6	597.7	57.8	30.1	28.8
2009	1,310.10	620.9	588.4	58.4	29.1	27.8
2010	1317.71	618.8	582.1	58.5	29.2	27.7
2011	1328.02	611.6	581.9	60.9	28.5	27.3
2012	1335.19	628.0	589.6	61.2	30.6	29.5

Source: UNSD; CSO

Table 1-2: Growth Patterns of population, labour force, and employment						
	Trinidad & Tobago			Tobago		
Year	POPULATION	Labour Force	Employed	POPULATION	Labour Force	Employed
2000						
2001	0.35%	0.63%	2.39%	1.48%	8.77%	3.33%
2002	0.70%	1.70%	2.16%	0.73%	1.61%	5.99%
2003	0.53%	1.76%	1.71%	0.54%	4.37%	4.78%
2004	0.64%	2.83%	5.26%	0.72%	2.66%	7.47%
2005	0.30%	1.68%	2.10%	0.71%	10.37%	9.65%
2006	0.26%	0.24%	2.13%	0.71%	-3.69%	-4.23%
2007	0.33%	-0.45%	0.27%	0.88%	0.00%	1.47%
2008	0.49%	0.67%	1.68%	0.87%	4.88%	4.35%
2009	0.11%	-0.91%	-1.56%	1.04%	-3.32%	-3.47%
2010	1.38%	-0.34%	-1.07%	0.17%	0.34%	-0.36%
2011	0.37%	-1.16%	-0.03%	4.10%	-2.40%	-1.44%
2012	0.33%	1.97%	1.32%	0.49%	7.37%	8.06%
Source: UNSD, CSO						

The Government and Trade Balances

Table 1.3 shows the trend in government expenditures, receipts and debt since 2000, while Table 1.4 shows their growth patterns. The general trend is to an overall deficit over the period, and thus to an upward drift of debt and debt service requirements, especially since 2003. Table 1.5 shows the state of the balance of trade over the decade since 2000. Here, the evidence is best summarized as indicating that there was a rising positive balance before the Great Recession of 2007/8 and a sharp negative shock thereafter in 2009. In the leading economies, such as the USA, output is only now recovering. It is well-known that the shock to the local economy is linked to the sharp decline in international energy prices, and there has been significant recovery since 2008. The positive trends have allowed build-up of import cover of up to 14 months. Despite the cushion, the immediate response to the 2007/2008 shock indicates that national development policies must address existing imbalances in the economy, especially the high level of dependence on the energy sector.

Year	GDP at constant 2000 prices ($MNTT)	GOV'T RECEIPTS ($MNTT)	GOV'T EXP. ($MNTT)	GOV'T RECEIPTS ($MNTT)	GOV'T EXP. ($MNTT)	TOTAL DEBT ($MNTT)	DEBT SERVICE ($MNTT)
2000	49,343.3	12,199.0	12,499.0	93.5	351.9	20,749.0	4,893.0
2001	53,512.1	14,381.0	13,991.0	97.1	539.5	20,044.0	4,706.0
2002	57,750.7	14,122.0	14,227.0	106.9	695.9	20,637.0	3,009.0
2003	66,100.2	17,366.0	16,592.0	114.4	680.7	21,461.0	1,742.0
2004	83,652.5	20,885.0	20,674.0	128.6	924.3	22,043.0	3,115.0
2005	75,785.6	29,648.0	27,234.0	168.6	1,177.6	22,287.0	4,449.0
2006	85,795.4	38,911.0	37,085.0	125.4	1,097.9	19,510.0	2,914.0
2007	89,874.3	40,064.0	39,796.0	122.7	1,455.8	22,238.0	3,744.0
2008	92,922.6	56,848.0	53,873.0	150.6	1,885.4	23,621.0	3,341.0
2009	88,841.7	39,045.0	45,731.0	138.2	1,874.4	25,278.0	5,063.0
2010	89,027.0	44,343.0	53,910.9	128.7	2,209.7	31,148.0	5,362.0
2011	87,600.7	49,236.0	60,170.9	152.4	2,466.3	33,719.0	3,697.0
2012	88,934.2	46,671.0	56,704.8	158.6	2,324.2	34,310.0	3,608.0

Source: UNSD,CSO

Table 1-4 Growth Trends in government budget indicators- receipts, expenditure and debt

	Trinidad & Tobago			Tobago		Trinidad & Tobago	
Year	GDP at constant 2000 prices	GOV'T RECEIPTS	GOV'T EXP.	GOV'T RECEIPTS	GOV'T EXP.	TOTAL DEBT	DEBT SERVICE
2000							
2001	8.45%	17.89%	11.94%	3.85%	53.31%	-3.40%	-3.82%
2002	7.92%	-1.80%	1.69%	10.09%	28.99%	2.96%	-36.06%
2003	14.46%	22.97%	16.62%	7.02%	-2.18%	3.99%	-42.11%
2004	26.55%	20.26%	24.60%	12.41%	35.79%	2.71%	78.82%
2005	-9.40%	41.96%	31.73%	31.10%	27.40%	1.11%	42.83%
2006	13.21%	31.24%	36.17%	-25.62%	-6.77%	-12.46%	-34.50%
2007	4.75%	2.96%	7.31%	-2.15%	32.60%	13.98%	28.48%
2008	3.39%	41.89%	35.37%	22.74%	29.51%	6.22%	-10.76%
2009	-4.39%	-31.32%	-15.11%	-8.23%	-0.58%	7.01%	51.54%
2010	0.21%	13.57%	17.89%	-6.84%	17.89%	23.22%	5.91%
2011	-1.60%	11.03%	11.61%	18.34%	11.61%	8.25%	-31.05%
2012	1.52%	-5.21%	-5.76%	4.12%	-5.76%	1.75%	-2.41%

Source: CSO

Table 1-5 The Balance of Trade and FDI				
Year	GDP at constant 2000 prices ($MNTT)	Trade Balance at constant 2005 prices ($MN US)	Foreign Direct Investment Inflows ($MNUS)	Investment Income Paid Abroad ($MNUS)
2000	49,343.3	2,016.1	679.5	628.5
2001	53,512.1	1,662.1	834.9	539.3
2002	57,750.7	835.7	790.7	479.8
2003	66,100.2	2,735.2	808.3	680.9
2004	83,652.5	2,458.9	998.1	397.3
2005	75,785.6	4,283.8	939.7	760.3
2006	85,795.4	6,425.2	882.7	935.8
2007	89,874.3	5,347.4	830.0	963.7
2008	92,922.6	7,770.0	2,800.8	1,202.2
2009	88,841.7	2,224.8	709.1	996.7
2010	89,027.0	4,285.6	549.4	1,074.5
2011	87,600.7	5,020.6	770.6	3,073.9
2012	88,934.2	4,992.4	839.5	3,424.0
Source: UNSD,CSO, CBTT				

Underlying Industry Structure

Table 1.6 shows that under these patterns there is a certain structure of the GDP by sector. It indicates that the share of exploration and production of energy supplies has been growing relative to all others, indicative a rising cost in terms of the rate of depletion of natural resources. Mining is also highly vulnerable to exogenous shocks. The share of manufacturing has also grown, primarily as a result of the activities of Atlantic LNG and food and beverage manufacturing with imported inputs. Manufacturing has a very low import productivity signature, with its mean varying from 0.7 up to 6.7 depending on the subsector (Table 1.7). In the manufacturing sector production of petrochemical derivatives ranks the highest. Import productivity in the mining sector is also moderate, at about 9. These estimates compare with 33.5 for the best performing subclasses of the copyright sector. The fact that the sectors with the highest import efficiency signature are not currently the fastest growing also points to underinvestment in them, when judged against the experience of negative shocks and the need for endogenous responses to an appreciating real exchange rate through the domestic production system. Put differently, by increasingly dominating the economy, the mining and manufacturing sectors are systematically suppressing the productivity of imports, moderating the productivity growth effects of the rest of the economy, and thereby are constraining growth of the capacity of the economy to save foreign exchange through the production system. One way this happens is to suppress investment in capacity-building targeting the high productivity sectors at a sufficient pace to maximize growth of their productivity. This could happen even if the high productivity sectors are growing as a result of private investment.

Table 1-6 : Value Added by Economic Activity, Percentage Distribution (Shares of GDP Constant 2000 Prices)													
	2000	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**
Agriculture	**1.3**	**1.1**	**1.1**	**0.8**	**0.6**	**0.6**	**0.4**	**0.4**	**0.4**	**0.4**	**0.4**	**0.4**	**0.4**
Mining and Utilities	**20.0**	**19.7**	**20.8**	**23.4**	**22.8**	**24.3**	**24.7**	**24.0**	**23.0**	**24.0**	**25.1**	**24.6**	**23.7**
Exploration and Production	18.2	17.9	19.1	21.7	21.7	22.7	23.3	22.6	21.6	22.6	23.6	23.0	22.0
Asphalt Production	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1
Electricity and Water	1.8	1.7	1.7	1.6	1.1	1.5	1.4	1.4	1.4	1.4	1.4	1.5	1.6
Manufacturing	**17.6**	**17.3**	**18.0**	**20.3**	**20.7**	**20.9**	**23.0**	**23.4**	**23.4**	**24.3**	**24.6**	**24.0**	**23.1**
Food, Beverages and Tobacco	3.4	3.4	3.2	2.9	2.3	3.3	3.0	3.5	3.7	4.2	4.4	4.5	4.7
Textiles, Garments and Footwear	0.2	0.2	0.2	0.1	0.1	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
Printing, Publishing etc.	0.8	0.8	0.8	0.9	0.7	0.8	0.8	0.9	0.9	0.8	0.8	0.7	0.8
Wood and Related Products	0.3	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
Chemicals and Non-Metallic Minerals	1.4	1.2	1.1	1.4	1.6	1.4	1.5	1.5	1.3	1.4	1.4	1.5	1.4
Assembly Type and Related Industries	0.8	1.1	1.4	1.3	2.4	1.4	1.5	1.6	1.5	1.5	1.5	1.3	1.3
Miscellaneous Manufacturing	0.4	0.4	0.4	0.3	0.2	0.3	0.3	0.4	0.4	0.5	0.5	0.4	0.4
Sugar Refineries	0.1	0.2	0.1	0.1	0.1	0.2	0.1	0.1	0.1	0.1			
Distilleries	0.3	0.4	0.4	0.3	0.1	0.1	0.2	0.3	0.3	0.2	0.3	0.3	0.2
Refining (Incl. Atlantic LNG)	5.7	5.2	6.0	9.0	6.9	8.6	10.7	10.6	10.8	10.6	10.7	10.2	9.7
Petrochemicals	4.4	4.4	4.5	4.1	6.2	4.8	4.9	4.6	4.3	4.9	4.8	4.7	4.3
Construction	**7.8**	**7.9**	**7.0**	**7.5**	**7.1**	**8.2**	**7.7**	**7.9**	**8.0**	**7.7**	**5.5**	**5.1**	**4.9**
Trade	**20.3**	**18.4**	**17.3**	**15.4**	**14.9**	**14.6**	**14.6**	**14.4**	**15.9**	**13.8**	**12.7**	**13.3**	**13.3**
Distribution and Restaurants	17.0	15.3	14.3	12.8	12.7	12.0	12.2	12.1	13.7	11.4	10.3	10.9	10.9
Hotels and Guest Houses	0.4	0.4	0.4	0.4	0.4	0.4	0.3	0.3	0.2	0.3	0.3	0.3	0.3
Distribution	2.8	2.6	2.5	2.2	1.9	2.1	2.1	2.1	2.0	2.1	2.1	2.1	2.2
Transport	**8.9**	**8.9**	**9.0**	**8.3**	**6.5**	**7.7**	**7.2**	**7.2**	**7.6**	**7.9**	**8.1**	**8.3**	**9.2**
Other Services	**28.7**	**26.7**	**26.4**	**24.4**	**27.2**	**23.5**	**21.3**	**21.8**	**21.2**	**22.1**	**23.4**	**24.7**	**25.3**
Finance, Insurance, Real Estate	14.8	13.8	14.2	13.3	14.0	13.8	12.3	13.0	13.0	12.9	14.0	14.9	15.4
Government	7.9	7.2	6.9	6.0	7.1	5.3	4.6	4.6	4.5	5.6	5.6	5.7	5.7
Education and Cultural Services	2.9	2.6	2.6	2.3	2.7	2.2	2.1	2.0	1.9	2.0	2.0	2.1	2.0
Personal Services	1.6	1.6	1.5	1.6	1.3	1.3	1.3	1.3	1.2	1.3	1.4	1.3	1.4
Service Contractors	1.5	1.6	1.2	1.2	2.1	0.9	1.0	1.0	0.7	0.3	0.4	0.7	0.8
Adjustments	**4.5**	**0.1**	**0.4**	**0.1**	**0.2**	**0.3**	**1.2**	**0.9**	**0.5**	**0.2**	**0.3**	**0.4**	**0.1**
FISM	4.5	3.7	3.5	3.3	3.6	3.1	3.1	3.5	3.3	3.5	3.2	3.2	3.3
VAT	NA	3.8	3.9	3.3	3.8	3.4	4.3	4.4	3.9	3.3	3.5	2.8	3.4
Source: CSO													

Table 1-7 Import Productivity by Sector, ranked from highest to lowest	
Sector	Import Productivity (2000)[!]
Personal Services (9219,5020,9301,9302,9303)	33.5
Business Services(7010,7430,7111-7129,7492)	29.8
Insurance(6601,6603)	27.1
Oil & Gas Distribution (5050,4020)	20.9
Restaurants (5520)	20.8
Hotels & Guest Houses (5510)	17.9
Finance (6511,6519,6592,6599,7530)	15.3
Bakeries (1541)	13.6
W/sale & Retail Distribution (52/51)	13.3
Quarries & Asphalt (1410.1429)	12.4
Wood (2010,3610,2029)	11.9
Fruit & Vegetable processors (1513)	11.5
Oil & Gas production (1110)	9.0
Gas Processing (2411)	7.9
Fish processors (1512)	7.4
Petroleum & Gas Refineries (2320,2320)	6.7
Construction Materials(2422,2693,2695,2694,2520,2811)	6.7
Electricity (4010)	6.6
Transport(6021,6022,6304,6309,6301,6304,6302,6411,6412)	6.5
Communication (6420)	5.8
Construction (4510,4520,4530,4540)	4.3
Service Contractors (1120)	3.1
Water (4100)	3.0
Petrochemicals (2411)	3.0
Feed & flour mills (1531,1532,1533)	2.9
H/hold Chemicals (2423,2424)	2.6
Other Mfng (1911,3691,2519,3699)	2.3
Alcohol/Soft Drinks / Tobacco (1551,1553,1554,1600)	2.2
H/hold Appliances(3420,3140,3430,2511,3220,3512,3610,2899)	2.1
Textiles (1711,1810,1920,1729)	2.0
Poultry processors (1511)	1.9
Printing (2212,2211)	1.7
Meat Processors (1511)	1.5
Plastic Products (2520)	0.8
Dairy factories (1520)	0.3
Misc. Food Mfgrs.(1543,1544,1549)	0.2
Paper Converters (2109)	0.1
[!]Measured as the ratio of value-added to the value of imports used in production	

Trinidad and Tobago has been investing considerable resources into the development of carnival and into other industries that rely mainly on copyright. Not much data are available for carnival itself, but many of the components of carnival are also the main activities in the copyright sector, so some basic impressions can be formed by considering that sector. The WIPO estimates for the copyright-based sector as a whole indicate the following:

1. In 2000, copyright contributed approximately TT$1,998 million to the economy –
    a. more than 3 times as much as agriculture (TT$626 million),

b. 2.8 times as much as hotels and guest houses (TT$710 million),
c. Nearly 23% of manufacturing (TT$8,700 million).

2. In real terms, the copyright sector contributed TT$3,630 million in 2007 and TT$4,102 in 2011.
   a. The real size of the copyright sector grew at an average rate of 11.7% per year between 2000 and 2007
   b. At a lower but still positive average rate of 3.2% per year between 2007 and 2011.

The estimates picked up the trend relative and absolute decline in press and the trend relative and absolute growth in radio and new media. Radio was somewhat surprising but it makes sense when streaming online and online ads are considered. In 2011, within the core copyright sector, there were four dominant sectors, ordered in terms of size: (i) *radio and television* (**TT$666.1 million**); (ii) *press and literature* (inclusive of academic publishing) **TT$200.2 million**); (iii) *advertising services* (**TT$147.6 million**), and (iv) *software and databases* (**TT$90.3 million**). Other significant sectors are *motion picture*, *video and sound* (**TT$69 million**), and the organisational and advocacy services provided by the professional organisations that are dedicated to promoting and protecting the interests of the copyright sector (**TT$25.2 million**).

The supply and use tables that provided the data for Table 1.7 have not been updated in Trinidad and Tobago since 2000. However, in the absence of up to date information about the efficiency of foreign exchange use by sector, some continuous policy guidance can be obtained from trends in the gross national product per dollar of imports used. In that regard, let p the domestic price level, $\frac{\varepsilon p_j}{p}$ the relative average price of imports defined in terms of the exchange rate (domestic price of foreign currency), C the level of consumption, I the rate of investment, $(G-T)$ the primary budget gap, $\frac{\varepsilon p_x}{p}X$ the real value of exports, $\frac{\varepsilon p_j}{p}J$ the real value of imports, $\frac{\varepsilon}{p}NFFI$ the real value of net factor incomes flowing into the economy, and $\frac{NRD}{p}$ the real level of depletion of domestic natural resources. The $NFFI$ is the difference between the inflow of factor incomes from foreign economies and the outflow to foreign economies, i.e., the net factor incomes flowing into the country. Then, the real values of the gross national product (GNP) is defined as:

1-1 $$GNP = C + I + (G-T) + \left(\frac{\varepsilon p_x}{p}X - \frac{\varepsilon p_j}{p}J\right) + \frac{\varepsilon}{p}NFFI - \frac{NRD}{p}$$

A useful way to use this equation is to normalize with the volume of income, so as to indicate the relative importance of initiatives to create it or dispose of it. For $y_j = \frac{GNP}{J}$ the reciprocal of the Keynesian import share and Y the GNP, it can be rewritten as

1-2 $$1 = \frac{C}{Y} + \frac{I}{Y} + \frac{(G-T)}{Y} + \left(\frac{\varepsilon p_x}{p}\frac{X}{Y} - \frac{\varepsilon p_j}{p y_j}\right) + \frac{\varepsilon}{p}\frac{NFFI}{Y} - \frac{NRD}{pY}$$

The term $\left(\frac{\varepsilon p_x}{p}\frac{X}{Y}-\frac{\varepsilon p_j}{p y_j}\right)+\frac{\varepsilon}{p}\frac{NFFI}{Y}-\frac{NRD}{pY}$ is the major indicator of the (relative) capacity of the economy to shape foreign tastes and preferences, whether of producers or consumers – the capacity to compete in the global systems of high demand. It declines with depletion of natural resources, if the export capacity depends on natural resources. The capacity grows if exports per dollar of imports grows relative to the rate of depletion of natural resources. In the absence of compensating growth of the natural resource pool, it would be necessary to use culture, science and industry to grow $\frac{I}{Y}$ and thus to grow export capacity. Thus, for the purpose of policy design in Trinidad and Tobago, a central question is whether, as GDP or GNP grows, sectors other than energy can use science and culture to offer a long-run promise of growing $\left(\frac{\varepsilon p_x}{p}X-\frac{\varepsilon p_j}{p}J\right)+\frac{\varepsilon}{p}NFFI$ relative to $\frac{NRD}{p}$. Faced with large energy exports, some of these sectors must also be able to moderate or resist altogether the effects of the dominance of energy exports and the attendant appreciating exchange rate, since the latter discourages exports of non-energy output. This is one of the great discoveries of Arthur Lewis[2] and Dudley Seers[3].

We can assume that, because of the role of foreign investment, the energy exports shape the exchange rate rather than is shaped by it. Non-energy exports are not so fortunate. If the real exchange rate is appreciating ($\frac{\varepsilon p_x}{p}$ is falling for example through inflation), so that the value (price) of the domestic currency (in terms of the foreign currency) is rising, discouraging non-energy exports and encouraging imports, then the economy must adjust by a combination of two strategies. It must first recognize that the reduction in $\frac{\varepsilon}{p}$ does not affect $\frac{\varepsilon p_x}{p}\frac{X}{Y}$ and $\frac{\varepsilon p_j}{p y_j}$ in the same way. The greater risk might be that $\frac{\varepsilon p_x}{p}\frac{X}{Y}$ falls relatively faster than $\frac{\varepsilon p_j}{p y_j}$. Compensating strategies are needed to bring down $\frac{\varepsilon p_x}{p}\frac{X}{Y}$ more slowly than $\frac{\varepsilon p_j}{p y_j}$ and this can be achieved through the production system by raising y_j. The economy might use price-based cost-saving measures to make import-competing domestic industries more cost efficient and quality-competitive so that their prices fall and they can compete more effectively with the imports. Further, since both exports and import competing industries use necessary imports as inputs, the economy must also develop sectors that use imports with growing technical efficiency thereby ensuring that there are compensating increases in y_j. The inability to do this is the essence of the problem known as Dutch Disease, first discovered by Lewis and Seers. Success in making such adjustments is the key to transformation of the non-energy economy and it also builds capacity to adjust to random shocks in any direction to the energy sector and the fiscal space. It is the capacity to undertake such adjustments that makes the growth of output per dollar of foreign exchange a central feature of the carnival sector's performance. One can view this from another perspective. High import productivity is known to be linked to high output of the domestic capital asset called goodwill, especially intellectual property, and is needed to compensate for the negative effects of Dutch Disease. Carnival thrives even in recessions by building up domestic capital making this adjustment.

[2] Lewis, W.A. (1964). Jamaica's Economic Problems. Kingston, Jamaica: The Gleaner Co.
[3] Seers, D. (1964). The Mechanism of an Open Petroleum Economy. *Social and Economic Studies*, 13(2) (June).

So, there are good analytical reasons for considering the path of import productivity. Some additional insight can therefore be gained from the data in Table 1.8, which show that the overall import productivity of the economy has stagnated since 2000. These data, taken together with the data showing that the shares of most productive sectors are not growing the fastest suggest a significant possibility of underinvestment in those sectors. Indeed, the data seems to suggest a high-risk growth path that leaves the economy vulnerable to potentially significant problems on the balance of payments and government debt, because adequate investment is not initiated to take advantage of the high levels of productivity in carnival and other related industries. In general, these risks have only materialized randomly in the past, but when they do the result is often significant social and political turmoil. On the other hand, to define and initiate such investments in the absence of adequate data it is necessary bring all stakeholders into an adequate joint decision-making process while setting up sound foundations of causal analysis for monitoring the outcomes and for refining the investment package over time.

This Aide Memoire defines a basket of desired initial interventions, identified by the stakeholder community of the National Carnival Commission and the scientists recruited to study various aspects of the carnival industry. The basket of programs will have to be routinely monitored and evaluated to determine the efficacy of each and all of the interventions on the relevant population, and to assist with their continuing prioritization. In this sense, monitoring seeks to ascertain the extent of the causal effect of the program. Like governments around the world at the Rio+20 Conference on Sustainable Development, the Government of Trinidad and Tobago has committed to strengthening the science-policy interface, underpinned by a push for democratized choice from among the policy options generated by science. This amounts to requiring adequate proof of causation as a condition for policy intervention. The scientist guiding, monitoring and evaluating policy must begin by identifying clearly the causal relationship being targeted by the policy-maker (Angrist and Pischke, 2009: 3)[4]. The rest of this report explores various aspects of the challenge of scientific causal policy analysis and the solutions offered by regression modelling and the underlying mathematical statistics, with special reference to the carnival industry.

Table 1-8: Import Productivity Performance and Sector Growth

Year	Growth of Real Exports	Growth of Real Imports	GDP Growth	Crude Import Productivity	Import Productivity Growth	Export Share of GDP
2000				2.62		0.45
2001	0.20%	8.80%	7%	2.51	-9.10%	0.45
2002	-8.90%	6.10%	2%	2.55	-5.50%	0.45
2003	28.80%	-5.90%	23%	3.1	7.10%	0.38
2004	13.40%	27.70%	16%	2.62	-21.60%	0.39
2005	28.30%	8.60%	19%	2.56	-7.90%	0.39
2006	22.20%	3.20%	14%	2.81	-2.90%	0.37
2007	-4.80%	7.20%	17%	2.75	-6.50%	0.37
2008	24.60%	8.60%	25%	2.62	-8.50%	0.38
2009	-48.70%	-25.20%	-37%	3.34	34.50%	0.38
2010	22.30%	-5.70%	8%	3.55	6.90%	0.33

[4] Angrist, J.D and Pischke, J-S. (2009). Mostly Harmless Econometrics: An Empiricist's Companion. Princeton: Princeton University Press.

2011	32.00%	44.10%	14%	2.43	-30.20%	0.42
2012	-3.90%	-6.10%	-1%	2.62	6.90%	0.37
Source: Computed from UNSD Data						

2. The Policy Hypothesis and Causal Analysis

The central hypothesis of the scientific policy model in the NCC intervention regime has to be that deliberate intervention or participation of one sort or another (say employment) in the carnival industries improves individual and household incomes as well as their general well-being. Affirmation of this as a causal relationship allows conditional statements, 'predictions', of the effects of changing policies or adjusting circumstances. It would indicate what would happen to income or well-being in the 'counterfactual' or alternative worlds of representative samples of persons participating in carnival and a representative sample of persons who do not. This is essentially similar to comparing the average level of output per person or of some other indicator of well-being in the carnival industry with that in one or more other industries. From a policy standpoint, we can imagine offering rewards and incentives for participation in carnival and then considering the growth and well-being outcomes over time.

While we do not dwell on the matter here, we recognize the fundamental importance of internally consistent theory in shaping hypotheses, indicating what should be counted and how, and thus indicating what should be important for policymakers to address. However, evidence-based proof of the claims of theory is also needed, and must be supplied by econometrics through the discovery of relevant instruments or other relevant tools of analysis. Further, we also recognize the importance of history, especially in the sense of generating theory that fits encouraged by Best (1968; 1975).[5] History tends to matter in efforts to shape theory about what causes what. History is also significant for its emphasis of the use of detail and local context to achieve external consistency among competing theories – competing theories must be mutually reconcilable at some stage (Morck and Yeung, 2011:1).[6]. In other words, economics (or any other social science) is not econometrics but is aided by it in a dynamic interplay as the former adds causal assumptions.

The essence of the call for evidence-based policy is the same as the call for claims about cause, even when firmly rooted in history and theory, to meet the minimum standards set by Kuhn (1962).[7] Logically valid claims about causation are provisionally true. They must be checked to see if they can be falsified, and can be left standing until they are. However, we need to be even sharper in our thinking by probing how to go beyond the patterns of association of standard statistical analysis, i.e., the computations from data and analysis of

[5] Best, L. A. (1968). A Model of Pure Plantation Economy. *Social and Economic Studies*, 17(3): 283-326. Best, L. (1971a); Best, L. (1975). A Biography of Labour in G. Beckford (ed.), *Caribbean Economy*: Dependence and Backwardness. Mona: ISER, pp.147-158.

[6] Morck, R. and Yeung, B. (2011). Economics, History, and Causation. NBER Working Paper 16678. http://www.nber.org/papers/w16678.

[7] Kuhn, T. (1962). *The Structure of Scientific Revolutions*. University of Chicago Press.

statistical distributions in the manner clarified by Pearl (2009).[8] The key principle to be added is that causal assumptions cannot be verified or falsified even in principle, "unless one resorts to experimental control." (Pearl, 2009: 101). Causal monitoring and evaluation "requires some knowledge of the data-generating process" and use of that knowledge to examine "the dynamics of beliefs under changing conditions" and thus to determine how a joint distribution "would change if external conditions were to change". History and free will are vital in this determination because to know how one property of a distribution changes when another property changes, it is necessary to rely on them to add "causal assumptions which identify relationships that remain invariant when external conditions change". This is the substance of the idea that "behind every causal conclusion there must lie some causal assumption that is not testable in observational studies" (Pearl, 2009: 97; 99).

Method of Investigation

To aid clarity, we discuss causal analysis in the context of constructed examples. Our examples are mostly specific to carnival and for that purpose, we take income or well-being as our main focus. Causal analysis examines the relationship between events that might be described as cause and events that might be described as outcomes, subject to the general principles that (i) the cause must predate or coincide in time with the outcome,[9] and that (ii) the cause makes unique changes in the outcome in the specific sense that it serves as an intermediate flow of information retained in the outcome (Hoover, 1990; 2001).[10] The main method of causal policy investigation for the carnival industries is to make reliable and valid comparisons of the income or well-being of persons participating in carnival treated as if they participate in carnival even if they actually do not, and a sample of persons who actually do not (Angrist and Pischke, 2008). Let D be a categorical random variable $\mathrm{D_i} = (0,1)$, reflecting 0 for not participating and 1 for participating. Also, let $\mathrm{Y_i}$ be the income of the persons of working age. The data setup for such a situation is illustrated in Table 1:

Table 1

Unit ID	Year	Income	Participation
1110	2014	1	0
1111	2014	2.3	1
1112	2014	1.5	0
1113	2014	1.8	1

For any individual of working age, there are two potential income variables:

2-1 $$Potential\ Y_i = \begin{cases} Y_{1i}\ if\ D_i=1 \\ Y_{0i}\ if\ D_i=0 \end{cases}$$

[8] Pearl, J. (2009). Causal inference in statistics: An overview. *Statistical Surveys*, **3**: 96-146.

[9] The requirement that cause predate or coincide in time with outcomes is open to debate. In human society, the future might cause the past via expectations and pre-planning for a desired future.

[10] Hoover, K. D. (1990). The Logic of Causal Inference: Econometrics and the Conditional Analysis of Causality, *Economics and Philosophy* 6(2), 207-234; Hoover, K. D. (2001) *Causality in Macroeconomics*. Cambridge: Cambridge University Press.

That is, Y_{0i} is the income of the individual if he had not participated in carnival, irrespective of whether he or she actually does, whileY_{1i} is the individual's income if he participates in carnival. For the i^{th} individual, we want to compute the difference between Y_{1i} and Y_{0i}, which would be the causal effect on a person of participating in the carnival.

The observed income, Y_i can be written in terms of the potential income as

2-2 $$Y_i = \left\{\begin{matrix} Y_{1i} \; if \; D_i=1 \\ Y_{0i} \; if \; D_i=0 \end{matrix}\right\} = Y_{0i} + (Y_{1i} - Y_{0i})D_i$$

Here, $(Y_{1i} - Y_{0i})$ is the causal effect of participation in carnival for individual i. In the population, there will be a distribution of the incomes Y_{1i} and Y_{0i}. However, we never see both potential incomes for the same person, so to investigate the effects of carnival, we must compute the mean income of persons who participate and persons who do not and extract the causal element in the difference. We cannot just do a naïve comparison of average incomes, although that would be informative. Instead, we need a comparison of means that takes account of selection bias.

An interesting example of the use of counterfactuals is what Holland (1988) describes as an 'encouragement' design. Suppose we want to consider the effects of fete on worker punctuality (discipline) during the carnival season. There is widespread belief that workers are not punctual during the carnival season because of the amount of feting (FET) done. As above, there are two experimental treatments: (i) for $D_i = 1$, throughout the year, before the carnival fete season starts, a treatment that encourages workers and gives them incentives to fete when carnival comes around; and (ii) for $D_i = 0$, a treatment that does not encourage workers to fete. Incentives might include democratisation or courses in personal responsibility, just to name a few options. After some exposure to the treatment, workers then decide on how much to fete during carnival. Subsequently, the NCC's research team can conduct measures of worker punctuality (Y_i) for each treatment group during the carnival season. The only experimental manipulation is the encouragement and incentive to fete, with room to provide different degrees of encouragement as the design team decides. Clearly, the analysts could then measure the effect of the incentive/encouragement D_i on FET and on Y_i. One can also estimate the effects of FET on Y_i.

3. Selection Bias

In statistics, "bias" refers to errors that appear consistently throughout the estimation of the moments of a distribution, using survey or experimental data, as a result of the method of sampling or data collection. It measures systematic differences between comparison groups in response to some factor such as participation in carnival. Selection bias occurs when the selection of cases (hence data points) for treatment and measurement is not sufficiently random to draw a general conclusion. The selection bias must be measured and taken into account if the comparisons of the means are to be accurate and reliable.The required decomposition is obtained by first noting that

3-1 $E(Y_i: D_i = 1) - E(Y_i: D_i = 0) = E(Y_{1i}: D_i = 1) - E(Y_{0i}: D_i = 0)$

For a relevant decomposition that focuses on the effects of participation, one needs only add and subtract $E(Y_{0i}: D_i = 1)$, which monitors how the expected Y_{0i} (income of the individual if he had not participated in carnival) is affected by actual participation. The result is:

3-2 $$E(Y_i: D_i = 1) - E(Y_i: D_i = 0) = E(Y_{1i}: D_i = 1) - E(Y_{0i}: D_i = 1)$$
$$+E(Y_{0i}: D_i = 1) - E(Y_{0i}: D_i = 0)$$

The decomposition has two key components. The first component is $E(Y_{1i}: D_i = 1) - E(Y_{0i}: D_i = 1)$, which is the true *causal effect of participation in carnival*. Here, we can factor out the condition of participation and write:

3-3 $E(Y_{1i}: D_i = 1) - E(Y_{0i}: D_i = 1) = E(Y_{1i} - Y_{0i}: D_i = 1)$

This says that the average causal effect of participation on those who actually participated in carnival is the average difference between (i) the income of the actual participants, $E(Y_{1i}: D_i = 1)$, and (ii) the average income participants would have gotten had they not participated, $E(Y_{0i}: D_i = 1)$.

The second component in the decomposition is the measure of selection bias $E(Y_{0i}: D_i = 1) - E(Y_{0i}: D_i = 0)$, which is the most important problem faced by an empirical researcher. This is the difference of the average Y_{0i} (the income of the individual if he had not participated in carnival) between those who participated and those who did not participate. If this term is either positive or negative, then it must be considered in understanding the difference of the average incomes of actual participants and non-participants. It tends to arise when there is an unbalanced representation of $D_i = 1$ and $D_i = 0$ in the sample of cases involved in forming the expectations. It should be clear that this computation also requires that all the potential factors which might affect income, other than carnival, are held fixed and equal. That is, the *ceteris paribus condition* must hold. We return to this shortly.

Random Sampling

Selection bias is quite difficult to measure, so much of the time statisticians tend to focus on prevention. We know that if there is a flaw in the sampling process, such that a subset of the data is systematically excluded or underrepresented (relative to size) due to a particular attribute, then the excluded subset can cause the moments to be biased as above (because of the effects of the attribute) and can influence the statistical significance of tests about the moments. The result is the 'spotlight fallacy'. Among other factors, sample size, time scale differences and attrition can cause selection bias. In the example on encouragement design above, random assignment of incentives would normally be possible but once done, workers self-select their own extent of exposure to feting, FET and this transforms the design from a full experimental design into a quasi-experimental design (Campbell and Stanley, 1963).[11]

It can also be treated, though not completely, with Heckman-type estimates and measurement of unobservable background forces affecting the outcome of interest or unobservable factors affecting the selection of members in the sample. In regression-type models, instruments are needed for this purpose. However, ultimately, random allocation is the main protection against this sort of bias and in a setting of cluster sampling this will only be achieved if the sample size is sufficiently large in the first place.

Proper random sampling (or random allocation) works because it makes D_i independent of any of the likely outcomes. This can be observed by noting that, in (3.1), independence implies that $E(Y_{0i}\colon D_i = 1) = E(Y_{0i}\colon D_i = 0)$, so the selection bias term disappears in equation (3.2) and in particular,

3-4
$$\begin{aligned} E(Y_i\colon D_i = 1) - E(Y_i\colon D_i = 0) &= E(Y_{1i}\colon D_i = 1) - E(Y_{0i}\colon D_i = 0) \\ &= E(Y_{1i}\colon D_i = 1) - E(Y_{0i}\colon D_i = 1) \\ &= E(Y_{1i} - Y_{0i}) \end{aligned}$$

The effect of randomization of participation on the participants in carnival is the same as the effect of participation in carnival on a randomly chosen person and either way selection bias is eliminated. So, if proper random sampling is done and then it is found that $E(Y_{1i} - Y_{0i}) > 0$, then this would be a good case for using participation in carnival to increase employment and earnings.

[11] Campbell, D. T., and Stanley, J.C. (1963). ***Experimental and Quasi-Experimental Designs for Research***. Chicago: Rand-McNally.

4. Selection Bias as Endogeneity

With a random sample in hand, we can put the counterfactual equation (2-1) to further use – econometric analysis – and that introduces the well-known interpretation of selection bias as the endogeneity problem. Let us assume that D_i is randomly assigned and that the effect $(Y_{1i} - Y_{0i}) = \beta$ is a constant. This makes sense if the effect of carnival is the same across individuals. Also assume that Y_{0i} has a random component, $e_i = Y_{0i} - E(Y_{0i})$. Then, one can write

4-1 $Y_i = E(Y_{0i}) + (Y_{1i} - Y_{0i})D_i + Y_{0i} - E(Y_{0i}) = \alpha + \beta D_i + e_i$

The second equation in (4-1) is the standard form of a regression of a categorical variable (D_i) on Y_i. The coefficient $(Y_{1i} - Y_{0i}) = \beta$ measures what happens to Y_i if D_i changes by one unit (from 0 to 1). Now taking expectations to the second equation conditional on the value assumed for D_i, we get two distinct equations. For $D_i = 1$, the first is:

4-2 $E(Y_i\colon D_i = 1) = \alpha + \beta + E(e_i\colon D_i = 1)$

The second is

4-3 $E(Y_i\colon D_i = 0) = \alpha + E(e_i\colon D_i = 0)$

Now take the difference and get

4-4 $E(Y_i\colon D = 1) - E(Y_i\colon D = 0) = \beta + E(e_i\colon D = 1) - E(e_i\colon D = 0)$

Using the definition of e_i, consider the second term $E(e_i\colon D = 1) - E(e_i\colon D = 0)$. We get,

4-5
$$\begin{aligned} &E(e_i\colon D_i = 1) - E(e_i\colon D_i = 0) \\ &= (E(Y_{0i}\colon D_i = 1) - E(E(Y_{0i}))\colon D_i = 1) - E(Y_{0i}\colon D_i = 0) + E(E(Y_{0i}))\colon D_i = 0) \\ &= E(Y_{0i}\colon D_i = 1) - E(Y_{0i}\colon D_i = 0) \end{aligned}$$

However, by the principle of iterated expectations, $E(E(Y_{0i}))\colon D_i = 0) = E(Y_{0i}) = E(E(Y_{0i}))\colon D_i = 1)$. So, $E(e_i\colon D_i = 1) - E(e_i\colon D_i = 0)$ is actually the selection bias. It measures the difference in the potential incomes of persons who do not participate in carnival if they did and the potential incomes of persons who do not participate in the carnival. This selection bias is the same as the correlation between the residual and the regressor D_i. This means that selection bias is the same as the famous endogeneity problem of econometrics. For the regression to measure the causal effects of interest, the selection bias (or endogeneity problem) must be eliminated. This condition also requires us to revisit the well-known *ceteris paribus* condition mentioned above.

Significance of Ceteris Paribus – Controlling Confounders

The comparison in equation (4-5) might be contaminated or confounded by other factors affecting the mean incomes. Label these as X_i. The data setup now looks like:

Table 2: Data setup with confounders							
Unit ID	**Year**	**Income**	$\boldsymbol{D_i}$	**X1**	**X2**	**…**	**Xn**
1110	2014	1	0				
1111	2014	2.3	1				
1112	2014	1.5	0				
1113	2014	1.8	1				

The ceteris paribus condition amounts to a requirement that these other factors must be properly balanced across the two comparison groups in order that the means can be comparable. That is to say, they must have no relationship to the regressor D_i. If this condition is satisfied, then they will not affect the estimate of β in equation (4-1). In particular, let γ be a row vector of coefficients for the column vector of factors X_i. Now, consider the regression model

4-6 $Y_i = \alpha + \beta D_i + \gamma X_i + e_i$

If X_i is unrelated to D_i, then the estimate of the coefficient β in (4-6) and (4-1) should be the same or very close and, if anything, should generally be more precise than other estimates of β. In any event, multicollinearity will tend to prevent (accurate) inversion of the matrix formed by concatenating the columns $\mathrm{D_i}$ and $\mathrm{X_i}$ and any related matrices. Therefore, the absence of multicollinearity is also a condition for a causal interpretation of coefficient β.

However, even in the absence of multicollinearity, several other issues affect the predictive power of a model such as (4-6), which must be addressed in the measurement and data collection design. One of these is that a model such as (4-6) relies on inter-unit variation (purely cross-sectional data) to estimate the coefficients. Such a model would tend to be plagued by omitted variable bias that make the predictors endogenous (Cameron and Trivedi, 2010).[12] So, its capacity to predict how income will adjust to changes in the carnival industry will be relatively low unless addressed. We want to detail what the issue really is and see how to treat it. One type of strategies will rely on instrumental variables and another type will look to fixed and random effects models, and related differences in differences analysis. We lay foundations in the analysis of conditional expectation before identifying the solutions.

[12] (Cameron, A. C., and Trivedi, P. K. (2010). Microeconometrics Using Stata. Rev. ed. College Station, TX: Stata Press).

5. Conditional Expectation Explained

Considering equations (4.1) to (4.6), it is clear that, as a general matter, our interests focus on the two conditional expectation functions – both population concepts. The first is the conditional expectation function for the dependent variable, written generally as $E(Y_i|X_i)$ and the second concerns the conditional expectation function for the residual, $\mathrm{E}(\mathrm{e_i}: X_i)$, where we treat the argument X_i as some general $n.k$ column matrix of all the data assembled on the k covariates of Y_i. Clearly, the discussion so far involved the mathematical consideration of (i) the linear independence of the column vectors and (ii) the randomness of the selection of cases documented in these vectors, further assuming the accuracy of the measurements reported in the data. Conditional expectation is a mathematical construct that captures the primacy of randomness in the whole discourse and in any case it is central to a proper understanding of causation and ultimately to the design of policy.

Suppose then that Y_i is a continuous random variable. Moreover, suppose that we can write down a conditional probability density function for Y_i as $f_y(y|X = x)$. Then, the conditional expectation function of the random variable Y_i on the random variable X_i, given that the latter value takes the value x is:

5-1 $E(Y: X = x) = \int_y \; y f_y(y: X = x) dy$

If Y_i and X_i are both discrete random variables, then the conditional expectation function is given by the discrete analogue:

5-2 $E(Y|X = x) = \sum_y \; yP(Y = y|X = x) = \sum_y \; y \frac{P(Y=y,X=x)}{P(X=x)}$

Obviously, we need to impose the requirement that $P(X = x) \neq 0$. It is possible to have a distribution in which one variable is discrete and one is continuous. A problem arises if X is now a continuous random variable while Y remains discrete because, in such a case, $P(X = x) = 0$. We can adjust for this possibility by rearranging on the right hand side to get:

5-3 $E(Y: X = x)P(X = x) = \sum_y \; yP(Y = y, X = x)$

Both (5.1) and (5.3) are trivial for individual values x, since the integral is zero in the first case and both sides are zero in the second. However, if we consider any measurable subset B of the domain of X, then it should hold that

5-4 $\int_B \; E(Y: X = x)P(X = x)dx = \int_B \; \sum_y \; yP(Y = y, X = x) \, dx$

It is now even clearer that in a practical exercise, these probabilities must be computed accurately, which is why biased selection is so much of a dangerous prospect when data are obtained in the form of samples.

Understanding Conditional Expectation with Joint Distributions

Usually, the policymaker is interested in two or more variables at the same time, and these might be jointly distributed. Statistical agencies routinely collect data on many variables that are jointly distributed. The more complex the argument gets, the more we must condense the discussion into mathematical language, and hence functions, that allows us to keep the logic clean. For this, we need some more foundations. We assume two-variables but the claims extend naturally to $n > 2$ variables.

For continuously distributed random variables (X_i, Y_i), the joint cumulative distribution function is defined by

5-5 $F(x, y) = P(X_i \leq x, Y_i \leq y)$

Given this joint cumulative distribution function, the associated joint density function, $f_{xy}(x, y)$ is given by

5-6 $f_{xy}(x, y) = \frac{d^2F(x,y)}{dxdy}$

In practice, $f_{xy}(x, y)$ would be given by an explicit formula (closed-form), but in general it is any integrable function that satisfies the properties $f_{xy}(x, y) = \frac{\mathrm{d}^2F(x,y)}{dxdy} \geq 0$ and $\iint f_{xy}(x_i, y)dxdy = 1$.

Now, once we have an $f_{xy}(x, y)$, we can recover the cumulative distribution function (5.5) using the basic definition of a density function. In particular, by definition,

5-7
$$
\begin{aligned}
P(x \leq b, y \leq b) &= P(-\infty \leq x \leq b, -\infty \leq y \leq d) \\
&= \int_{-\infty}^{d} \int_{-\infty}^{b} f_{xy}(x, y)\, dxdy \\
&= F(b, d)
\end{aligned}
$$

However, the function must apply to all b and all b, d such that $-\infty \leq b, d \leq \infty$. So it must follow that (5.5) is the general definition of the cumulative distribution function evaluated at (x, y).

Marginal Distributions

Further, for any $f_{xy}(x, y)$, the marginal (or one variable) probability density function of $X_i, f_x(x)$, can be obtained by integrating $f_{xy}(x, y)$ with respect to y; and the marginal (or one-variable) probability density function of Y_i, $f_y(y)$, can be obtained by integrating $f_{xy}(x, y)$ with respect to x. That is,

5-8 $f_x(x) = \int \; f_{xy}(x, y)dy$

and,

5-9 $f_y(y) = \int \quad f_{xy}(x, y)dx$

These marginal distributions are needed to find out the probabilities of events in the neighbourhood of x, y. We know that by using the definition in (5-7), we can write

5-10 $P(a \le x \le b, c \le y \le d) = \int_a^b \int_c^d f(x, y)\, dydx$

If we only want to find $P(a \le x \le b)$, we can use

5-11 $P(a \le x \le b, -\infty \le y \le \infty) = \int_a^b \int_{-\infty}^{\infty} f(x, y)\, dydx$

But the inner integral simply generates a function of x. Let us define this as $f_x(x) = \int_{-\infty}^{\infty} f(x, y)\, dy$. Then it must follow from (5-11) that

5-12 $P(a \le x \le b, -\infty \le y \le \infty) = \int_a^b f_x(x)\, dx = P(a \le x \le b)$

The same thing can be done for the $P(c \le y \le d)$. So the concept of the marginal distribution is firmly rooted in the formal definition of the joint density function and is the vital tool needed to finding the probabilities of events in the neighbourhood of x and y.

Conditional Densities

In addition, the conditional densities can be defined. Specifically, the conditional density of $X_i|Y_i = y$ is obtained as

5-13 $f_{X_i|Y_i}(x|Y_i = y) = \frac{f(x,y)}{f_y(y)}, for\ f_y(y) \neq 0$

This can be simply written as $f_{X_i|Y_i}(x|y)$ and is regarded as a function of x with y fixed. It satisfies all the properties of an ordinary single variable probability density function. Similarly, the conditional density of $Y_i|X_i = x$ is

5-14 $f_{Y_i|X_i}(y|X_i = x) = \frac{f(x,y)}{f_x(x)}, for\ f_x(x) \neq 0$

Here again, $f_{Y_i|X_i}(y|x)$ is regarded as a function of y with x fixed. It too satisfies all the properties of an ordinary single variable probability density function. To avoid notational confusion, it should be noted that the mathematical literature would write the conditional probability density function in a variety of ways that mean the same thing: $f_{Y_i|X_i}(y|X_i = x)$ is the same as $f_{Y_i|X_i}(y|x)$ and this might sometimes be condensed to $f(y|x)$, and so on.

One can see how to retrieve probabilities in various neighbourhoods using the joint density functions. Specifically, the probability that (X_i, Y_i) falls in some interesting region B in the $xy - plane$, $P(X_i, Y_i)$, can be obtained as the double integral

5-15 $P(X_i, Y_i) \in B = \iint_B \ f(x, y)dxdy$

With all of this mathematical ammunition, it is a straightforward matter to get expectations via the joint density functions. Given *any* function$g(x, y)$ and the joint density function $f(x, y)$, the expectation of $g(x, y)$ can be obtained as the double integral

5-16 $E[g(x, y)] = \iint g(x, y)f(x, y)dxdy$

This is a rather powerful idea since $g(x, y)$can be any function at all. One of its most interesting application is the law of iterated expectations. Before we get to that, let us make sure that the special single variable cases match the usage in the earlier discussion of conditional expectation.

Suppose that $g(x, y) = y$. In that case, from (5.16),

5-17 $E(y) = \iint yf(x, y)dxdy$

We can pull the y inside and follow the usual rule of integrating first with respect to the x term and then the other. So, we write

5-18 $E(y) = \int y \int f(x, y)dxdy$

Then, on doing the inner integral, we get the result as a function of y, so that we are left with

5-19 $E(y) = \int y\, f_y(y)dy$

Similarly, one can show that

5-20 $E(x) = \int x\, f_x(x)dx$

These are the usual definitions of mean or first moment. It remains only to note that, as a general matter, $f_{Y_i|X_i}(y|x)$ is regarded as a function of y with x fixed, and that it satisfies all the properties of an ordinary single variable (probability density) function of y. So, for the appropriate fixed value of x, the special case is just the same as $f_{Y_i|X_i}(y|x)$ and this might sometimes be condensed to $f(y)$ with no loss of generality in the xy -plane. This is not simply a trick of language. Since $g(x,y)$ can be any function at all, we can set it at $g(x,y) = f(y|x)$ and see what happens with (5.16). We get,

5-21 $E[f(y|x)] = \iint f(y|x)f(x,y)dxdy$

So, here too, we can pull the $f(y|x)$ inside and first integrate $f(x,y)dx$. We write

5-22 $E(f(y|x)) = \int f(y|x) \int f(x,y)dxdy$

Then, on doing the inner integral, we get the result as a function of y. Therefore, we are left with

5-23 $E((y|x)) = \int f(y|x)\, f_y(y)dy$

Thus, in the special case where $f(y|x) = y$, we get exactly the same result as (5-19) – clearly not a linguistic trick.

Independence

To see the power of the law of iterated expectations, we are also going to need the notion of independence, so let us define it here. Independence means that the random variables do not vary together in any way. This is a very strong condition. Random variables X_i and Y_i, with cumulative distribution functions $F_x(x)$ and $F_y(y)$ are said to be independent if, for all (x,y), their joint cumulative distribution function can be factored as $F(x,y) = F_x(x) * F_y(y)$. Further, if X_i and Y_i are independent, then their underlying density functions are such that $f(x,y) = f_x(x) * f_y(y)$. Note here that if the joint probability density functions can be factored, that does not mean that X_i and Y_i are independent. They are usually independent but they are not always independent.

Correlation

We also need to consider the idea of correlation more formally. First, keeping in mind (5.16), we start with the notion of covariance. Observe that if the random variables X_i and Y_i are jointly distributed, with joint probability density functions $f(x,y)$ and with marginal densities $f_x(x)$ and $f_y(y)$, then the covariance between X_i and Y_i is defined by

5-24 $Cov(X_i, Y_i) = E[(X_i - E(X_i) * (Y_i - E(Y_i)]$

Further, for σ_x, σ_y the respective standard deviations, the correlation between the random variables X_i and Y_i is given by:

5-25 $\rho = \frac{Cov(X_i,Y_i)}{\sigma_x \sigma_y}$

An important result is that if the random variables X_i and Y_i are independent, then $Cov(X_i, Y_i) = 0$ and so is the correlation. However, it is easy to see from the algebra of (5.24) that if $Cov(X_i, Y_i) = 0$, hence $\rho = 0$, it is not necessarily true that the variables X_i and Y_i are independent. Another is that the regression coefficient β can be obtained as:

5-26 $\beta = \frac{Cov(X_i, Y_i)}{V(X_i)}$

The Law of Iterated Expectations

As has been seen in Sections 2 and 3, the conditional expectation function is central to the analysis of causation. Underlying its role is the law of iterated expectations. Let $g(x, y) = E(Y\text{: X} = \text{x})$. Then, by (5-16),

5-27 $$E[E(Y|X = x)] = \iint E(y{:}\, X = x)f(x, y)dxdy$$
$$= \int E(y{:}\,\text{X} = \text{x}) \int f(x, y)dydx$$
$$= \int E(y{:}\,\text{X} = \text{x})f_x(x)dx$$
$$= \int [\int yf_y(y{:}\, X = x)dy]\, f_x(x)dx$$
$$= \int \int yf_y(y{:}\, X = x)f_x(x)\, dxdy$$
$$= \int \int yf(x, y)dxdy$$
$$= \int y \int f(x, y)dxdy$$
$$= \int yf_y(y)dy$$
$$= E(Y)$$

Notice that we have used equation (5.14) to get $\int \int yf_y(y{:}\, X = x)f_x(x)\, dxdy = \int \int yf(x, y)dxdy$.

6. Conditional Expectation and Regression

The key usefulness of the law of iterated expectations is that it allows the decomposition of a random variable Y into its conditional expectation function, which is to say a part that is explained by X_i, and a residual, which is the part that is not explained by, and so is independent of, X_i. In particular, it allows us to write:

6-1 $Y_i = E(Y_i|X_i) + e_i$

with $E(e_i|X_i) = 0$, meaning that the residual e_i is independent of X_i; and therefore with the residual e_i uncorrelated X_i or any function of X_i. Observe that since in (6-1) $e_i = Y - E(Y_i|X_i)$, it must be true that

6-2 $E(e_i|X_i) = E\big(Y - E(Y_i|X_i)\big) = E(Y_i) - E\big(E(Y_i|X_i)\big) = E(Y_i) - E(Y_i) = 0$

Here, we appealed to the law of iterated expectations in the last step. Further, write any function $h(X_i)$. By the law of iterated expectations, and using independence of e_i and X_i from (6-2), it would also hold that

6-3 $E(e_i h(X_i)) = E(E(e_i h(X_i)|X_i) = E(h(X_i)E(e_i|X_i) = 0$

So, indeed, (6.1) makes sense and the random variable Y_i can be properly decomposed into the part explained by X_i and the part that is not so explained.

Regression models are built to approximate the conditional expectation function. They also provide an accessible and yet powerful strategy for treating causality. If the conditional expectation function can be given a causal interpretation, so can the regression model that most closely approximates it. And, as we have seen, the conditional expectation can be given a causal interpretation if it properly describes the difference in *average potential outcomes* for a fixed reference population, such as the potential participants in carnival (Angrist and Pischke, 2009: 52). Here, the causal connection between income or well-being and participation in carnival must be understood as the function that describes what an individual would earn on average by participating to degrees in carnival, if we could change their levels of participation randomly so that the *ceteris paribus* conditions can be treated as satisfied and persons with different levels of participation are otherwise comparable. We could then consider a typical person as having the option to choose different degrees of participation, while recognizing that some choices are more likely than others. In this light we can peer more closely at the model in (4-6) and reconsider the notion of participation in carnival as well as the variables that enter as control factors, focusing on their sufficiency. It is a mighty challenge to specify regression models of the sort that could closely approximate the conditional expectation function of interest, especially because of the perceived need to rule out reverse causation. The type of information used is a major issue in all this and the general rule is to employ *pre-existing data* or *experimental data* to infer causality by regression methods.

To see how these fit, we must explore the issue of omitted variable bias in some detail. However, before doing so, we must summarize the gains from our investment in the whole discourse about joint distributions. It is that, as clarified by Pearl (2009: 99), joint distributions and the statistical constructs derived from them are tools for identifying patterns of association between variables. By themselves, they are not tools of causal analysis. When not nonsense distributions, they are "smoking guns" indicating the potential presence of a causal relationship (Morck and Yeung, 2011:2). However, to get to causal relationship from them, it is necessary to invoke assumptions that rely on causal concepts such as "randomization, influence, effect, confounding, holding constant, disturbance, spurious correlation, faithfulness/stability, instrumental variables, intervention, explanation, attribution," which characteristically "cannot be defined in terms of distribution functions" (Pearl, 2009: 100).

Problems of Regression Specification – Omitted Variables

Redefine D as a carnival immersion variable that can take a wide range of values. Some people can ignore the industry completely and have nothing to do with it all year and some can be involved in one way or another all year. We want to probe more thoroughly what allows β to be given a causal interpretation. In particular, we want to consider what person i would earn for any value of D and not just for the realized value, D_i. Nonetheless, D_i would be the observed values of D. Now, consider the model

6-4 $\quad Y_i = \alpha + bD_i + \eta_i$

Underlying (6.4) is the assumption that the coefficient b is typical of all cases. Here, by (6.1), for b to be given a causal interpretation, η_i must at least be the individual-specific and random component of Y_i that also determine individual income and is not explained by the conditional expectation $E(Y_i|D_i)$ approximated by bD_i. It must also be independent of D_i. However, we can usually expect this condition to fail because D_i *causes* Y_i and for this reason may be correlated with η_i which is a component of Y_i.

Next, suppose that independence can be achieved by including the vector of observed covariates X_i. That is, η_i and D_i are only correlated because of the observed characteristics X_i for which we did not control in equation (6-4). That also implies that the list is sufficiently exhaustive. Therefore, we consider the possibility that η_i can be sufficiently explained by the observable characteristics X_i. We write

6-5 $\quad \eta_i = \gamma X_i + e_i$

By equation (6-1), this decomposes η_i into its conditional expectation function, $E(\eta_i|X_i)$, approximated by γX_i, and its random component e_i that is independent of, hence uncorrelated with, X_i. It should be clear that, X_i must be selected explicitly for this purpose, and in general we have to find them by searching for variables that we need to hold fixed under the ceteris paribus conditions needed for comparison at the time when measuring the extent of participation in carnival. It would be preferable to use measures of these variables that are determined *before*D_i itself is determined since they cannot then be outcome variables generated by D_i

(Barnow, Cain and Goldberger, 1981; Angrist and Pischke, 2009: 59, 67, 68). Plugged into (6-4), the outcome model is

6-6 $Y_i = \alpha + \beta D_i + \gamma X_i + e_i$

However, this time, we have tried deliberately to design the analysis so as to ensure that $E(e_i \mathrm{D_i}) = 0$. If the wrong variables are included in X_i we have misspecification in the form of extraneous variables and if X_i is not a complete list, then we run up against misspecification in form of omitted variable bias. It matters what we hold constant. It will therefore pay intellectual dividends to investigate the relationship between the regression coefficients of various models with different control factors specified. Much of the analysis of misspecification as omitted variable bias is about this relationship.

Consider the comparison of (6-4) and (6-6). If we estimate (6-4) in error, then we want to investigate how b and β compare. Is $E(b)$ interpretable and is it a biased or unbiased estimator of β? We can judge by writing

6-7 $$b = \frac{Cov(D_i, Y)}{V(D_i)} = \frac{Cov(D_i, \alpha + \beta D_i + \gamma X_i + e_i)}{V(D_i)}$$
$$= \frac{\mathrm{Cov}(\mathrm{D_i}, \alpha) + \beta Cov(\mathrm{D_i}, \mathrm{D_i}) + \gamma Cov(\mathrm{D_i}, X_i) + Cov(\mathrm{D_i}, \mathrm{e}_i)}{V(\mathrm{D_i})}$$
$$= \frac{0 + \beta V(\mathrm{D_i}) + \gamma Cov(\mathrm{D_i}, X_i) + Cov(\mathrm{D_i}, \mathrm{e}_i)}{V(\mathrm{D_i})}$$
$$= \beta + \gamma \frac{Cov(\mathrm{D_i}, X_i)}{V(\mathrm{D_i})}$$
$$= \beta + \gamma \delta$$

where, $\delta = \frac{Cov(\mathrm{D_i}, X_i)}{V(\mathrm{D_i})}$ arises from regressing X_i on $\mathrm{D_i}$. Equation (6.7) is very helpful in thinking about some important issues of causal analysis.

Usually, informed researchers tend to identify and measure many of the confounding factors in X_i. However, some important ones among these confounding factors might be elusive and unobserved. This is what happens with factors such as business climate and ability. One key issue this relationship raises is that there will be no bias if $\delta = 0$, so that the omitted confounding from the setX_i are unrelated to the included ones. If there is a relationship, so $\delta \neq 0$, we have a problem of omitted variable bias on our hands and b cannot be given a causal relationship. A particularly severe form of this problem occurs when the additional candidate for inclusion in X_i, say X_{1i}, is a potential dependent variable in the regression – in our policy discussion, this would emerge as a result generated by participation in carnival. Such an inclusion in a regression of $\mathrm{Y_i}$ on $\mathrm{D_i}$, X_{1i} and the rest of $X_i$recreates the problem of selection bias and β cannot be given the causal interpretation needed for policy design (Angrist and Pischke, 2009: 65,66). In particular, in the regression, the comparison of outcomes Y_i conditional on X_{1i} will involve selection bias.

Another less severe form of the selection bias problem involves the use of proxy variables $P_i$as controls. These are variables that would partially control for omitted factors that should be in X_i but for which no observations are available. A proxy provides an alternative measure that might be moderately affected by D_i, as exhibited by the δ_p when $\frac{Cov(\mathrm{D_i}, P_i)}{V(\mathrm{D_i})}$ is computed (Wooldridge, 2010).[13] It might be better than no measure but is still polluted along the lines of equation (6-7), though to a lesser degree than indicated by $\gamma\delta$. (Angrist and Pischke, 2009: 68) advise that in the selection of candidates for X_i "timing matters". In particular, "[v]ariables measured before the variable of interest was determined are generally good controls". If there is uncertainty or inadequate knowledge of timing, then "clear reasoning about causal channels requires explicit assumptions about what happened first or the assertion that none of the control variables are themselves caused by the regressor of interest".

Regression Solutions - Instrumental Variables and Instruments

The method of instrumental variables in econometrics is perhaps the foremost method available for addressing omitted and extraneous variable bias, even though it is not much used outside of econometrics. We first provide an intuitive treatment and then extend that with the tools of calculus and OLS.

Intuitively, two sets of considerations arise. One is that the regressor of interest might in fact be endogenous, caused directly by some other set of factors not included in equation (6.6). In general, returning to (6.6), the other problem is that some of the control variables needed in X_i are often unobserved. These would show up in an inflated value of η_i as well as e_i. However, if the data collected from a properly done sample or from some other source includes a (set of) variable(s) Z_i that is sufficiently correlated with D_i but uncorrelated with X_i, then the omitted variable bias can be addressed satisfactorily. The lack of correlation amounts to satisfying the *exclusion* restriction that $Cov(\eta_i, Z_i) = 0$, and hence that $Cov(X_i, Z_i) = 0$ and $Cov(e_i, Z_i) = 0$. This has the important meaning that Z_i cannot normally be a regressor in the model of Y_i in (42). If Z_i and D_i were both regressors and we simply estimated (6.4), then the effect of Z_i would be absorbed in the residual and we would find that $Cov(e_i, Z_i) \neq 0$.

Now, if Z_i exists that satisfies our conditions, then we can go from the correlation $Cov(D_i, Z_i) > 0$ straight to the idea that there exists a line such that:

6-8 $D_i = a + bZ + u$

So, in (6.6), we would find that

6-9 $\beta_{IV} = \frac{Cov(a+bZ_i, Y_i)}{V(a+bZ_i)} = \frac{Cov(a, Y_i) + bCov(Z_i, Y_i)}{b^2 V(Z)}$

[13] Wooldridge, J.M. (2010). **Econometric Analysis of Cross-Section and Panel Data.** Cambridge, MA: MIT Press.

$$=\frac{\mathrm{Cov}(\mathrm{Z_i},Y_i)}{bV(Z_i)}=\frac{\mathrm{Cov}(\mathrm{Z_i},Y_i)}{\frac{\mathrm{Cov}(\mathrm{D_i},Z_i)}{V(Z_i)}V(Z_i)}$$

$$=\frac{\mathrm{Cov}(\mathrm{Z_i},Y_i)}{\mathrm{Cov}(\mathrm{D_i},Z_i)}=\frac{\frac{\mathrm{Cov}(\mathrm{Z_i},Y_i)}{V(Z_i)}}{\frac{\mathrm{Cov}(\mathrm{D_i},Z_i)}{V(Z_i)}}$$

$$=\frac{\beta_{yz}}{\beta_{dz}}$$

Thus, the causal coefficient of interest emerges as the ratio of the coefficient from the regression of Y_i on Z_i, the reduced form regression, and the regression of D_i on Z_i, the first stage regression. Clearly, we require $\beta_{dz} \neq 0$, hence that $\mathrm{Cov}(\mathrm{D_i}, Z_i)$ be sufficiently large and $V(Z_i)$ be sufficiently small. That means the instrument Z_i must have a statistically significant effect of D_i.This significant relation also transmits the effects of Z_i on to Y_i. The big question behind all this is what is embodied in the residual η_i of (6-4). If there are unmeasured factors that turn out to affect D, then we will need to have instruments for D_i as well as the observed conditions X_i in the model. So, in fact, we are looking at the recursive system:

6-10
$$\begin{aligned} D_i &= a + bZ_i + cX_i + u_i \\ Y_i &= \alpha + \beta D_i + \gamma X_i + e_i \end{aligned}$$

The reduced form of this system is

6-11 $Y_i = \theta_1 + \theta_2 Z_i + \theta_3 X_i + \varepsilon_i$

where, $\theta_1 = [\alpha + a\beta]$, $\theta_2 = b\beta$ and is the causal coefficient of interest, $\theta_3 = [c\beta + \gamma]$ and $\varepsilon_i = \beta u_i + e_i$.

The first equation of (6-10) essentially says that the instrument is assigned randomly, conditional on covariates X_i that reflect the ceteris paribus condition. This is its key strength in the modelling process – its causal effect on D_i. The second equation of (6-10) recognizes that the information from Z_i transmits to Y_i through D_i and only through that channel. To find instruments with such properties, the scientist needs considerable institutional knowledge about how participation in carnival is randomized, and in particular about what events precede the decisions to participate to whatever degree. If the residual contains information suggesting that participation is shaped by self-confidence factors, for example, and if self-confidence also shapes income, we are going to need instruments that affect participation in carnival independently of self-confidence. Family wealth might be thought of as a good example of such an instrument but it clearly influences individual income, so it fails the part of the test that the only avenue to Y_i should be through D_i. A better choice might be proximity of residence to one of the main centers of carnival, since persons living far from these centers are less likely to be active participants in any aspect of the industry than persons who live close to them. However, persons living far from the main business centres will also tend to be living in low income areas, so, in the Y_i model, we would need to add a categorical regressor for rural areas.

A Little Calculus and OLS can Increase Understanding

We want to expand our understanding of the problems discussed above with a little calculus and direct reference to the use of ordinary least squares to estimate the parameters of our models. The underlying problem is selection bias, which arises in a model such as (6-4) by changing the residual from η_i to $\eta_i(D_i)$. Persons may have a high level of η_i because of high self-confidence. This would lead to high levels of income, but it would also lead to high levels of participation, since participation in the displays and masquerade or some such is likely to be higher for those with high levels self-confidence. Thus, the model becomes

6-12 $\quad Y_i = \alpha + bD_i + \eta_i(D_i)$

To understand the consequences of $\eta_i(D_i)$, we differentiate Y_i w.r.t D_i:

6-13 $\quad \frac{dY_i}{dD_i} = b + \frac{d\eta_i(D_i)}{dD_i}$

Sound policy analysis requires estimation of the causal effect, b, which is to say the change in the conditional expectation (mean) but instead we get $b + \frac{d\eta_i(D_i)}{dD_i}$. In this simple case, OLS estimates $\frac{d\eta_i(D_i)}{dD_i}$ as the familiar $\frac{Cov(\eta_i, D_i)}{V(D_i)} = \frac{\sum(\mathrm{D_i} - \mathrm{E(D_i)})\eta_i}{\sum(\mathrm{D_i} - \mathrm{E(D_i)})^2} \neq 0$. Put differently, OLS regression estimates will yield the slope $\frac{dY_i}{dD_i}$ as the full set of associated changes in Y engendered by D, including the part that shows up as apparently random changes not explained by the conditional expectation function. So, this full set will be more that the magnitude and direction of causation, $E(Y_i|D_i) = bD_i$, which is what policy needs to work with. This is usually referred to as biased and inconsistent parameter estimation. Instrumental variables provide a way to get the unique change in the magnitude and direction of causation of $E(Y_i|D_i)$, as if $\frac{d\eta_i(D_i)}{dD_i} = 0$.

An experiment is the obvious way out of this situation, but these are not easy to do in social systems and might be prohibitively expensive. So, we turn to instruments as characterized above – instruments that are randomly assigned. Suppose now we want to estimate the response of market demand for fetes to exogenous changes in the price of fetes – the price of tickets, scalping apart. The demand will clearly depend on the price, but prices are not exogenously set because they depend on market demand as well. There is some endogeneity. The way out is to find a variable that affects price but does not affect demand. Some supply side factors are likely to be candidates, because there should exist variables that affect supply, which also affects price, but do not affect demand. Careful search might lead to the idea that certain technological factors make it easier to set up the fete, or background sponsorship not known by the patron until after the decision to buy tickets, would be uncontroversial. They affect the price but do not affect the demand.

Estimation Strategy

Assuming that proper instruments can be identified, equation (6-9) clarifies the estimation strategy. First, as a general matter, one could ignore linearity and write the problem as

6-14 $Y_i = f(D_i, X_i, \eta_i), \;\; D_i = g(Z_i, X_i, e_i)$

We seek β_{IV} in (6-9). This is the same as seeking

6-15 $\beta_{IV} = \frac{\frac{dY_i}{dZ_i}}{\frac{dD_i}{dZ_i}} = \frac{\beta_{yz}}{\beta_{dz}}$

This can be estimated consistently by running separate OLS estimates of the equations in (6-10), and then computing β_{IV} according to equation (6-15). However, only linear specifications guarantee the causal interpretation.

7. Fixed Effects, Random Effects and Panel Analysis

If we use cross-sectional data, then causal inference relies heavily on the ability to find and observe controls for confounding factors, i.e., X_i. Usually, we will almost certainly run up against the problem of unobserved confounding factors. One approach to addressing this problem is to use the instrumental variables method, as long as we can find good instruments. However, another strategy to deal with unobserved confounding factors is to exploit factors that cause variation in the identified units over time and have some effect on their incomes or well-being, while suppressing the underlying unobserved factors that cause variation across units and are correlated with the predictors specified, including the causal predictors of interest. This works if the unobserved factors do not themselves vary over time or vary in their influence on the predictors identified over time, so they cannot cause the changes in income or well-being over time. This amounts to estimating parameters for each unit, albeit taking advantage of data pooling possibilities when studying the variances (Kohler and Kreuter, 2009: 245).[14] The basic data requirements are panel data summarized in Table 3.

Table 3							
Unit ID	**Year**	**Income**	D_i	**X1**	**X2**	**...**	**Xn**
1110	2014	1	0				
1110	2015	2.3	1				
1111	2014	1.5	0				
1111	2015	1.8	1				
And so on							

From a modelling standpoint, we are now interested in Y_{it}, the income of unit i at time t. As before, one gets to observe either Y_{0it} for the non-participants or Y_{1it} for the participants. So, the potential is

7-1 $$Y_{it} = \{{}^{Y_{1it}\ if\ D_i=1}_{Y_{0it}\ if\ D_i=0}\} = Y_{0it} + (Y_{1it} - Y_{0it})D_{it}$$

The vector of confounding covariates are now assumed to be time-varying, written X_{it}. Further, it is assumed that there also exists a vector of unobserved confounding factors that are fixed and labelled U_i, with no time subscript. Call these self-confidence. Assume that the participation status is as if randomly assigned, conditional on U_i and X_{it}. This means that we assuming that the conditional expectation of Y_{0it} subject to U_i, X_{it}, t, D_{it} equals the conditional expectation of Y_{0it} subject to U_i, X_{it}, t. That is,

7-2 $$E(Y_{0it}|U_i, X_{it}, t, D_{it}) = E(Y_{0it}|U_i, X_{it}, t)$$

The general form of the fixed effects model that exploits 'within-group' variation is the linear model:

[14] Kohler, U. and Kreuter, F (2009). Data Analysis Using Stata. San Antonio: Stata Press.

7-3 $Y_{it} = a + \lambda_t + \beta D_{it} + \gamma U_i + \theta X_{it} + \varepsilon_{it} = \alpha_i + \lambda_t + \beta D_{it} + \theta X_{it} + \varepsilon_{it}$

where $\varepsilon_{it} = Y_{0it} - E(Y_{0it}|U_i, X_{it}, t)$, and where the key parameters to be estimated are λ_t which is the year effect – the coefficient on year dummy variables (Table 3); $\alpha_i = a + \gamma U_i$ which are the fixed effects to be estimated, operating as coefficients on individual dummy variables; and β which is the causal effect of interest. Clearly, if $D_{it} = 0$, so there is non-participation, then the conditional expectation of the income of non-participants is:

7-4 $E(Y_{0it}|U_i, X_{it}, t) = \alpha_i + \lambda_t + \theta X_{it}$

Further, the model assumes a linear and constant effect of participation, so that

7-5 $E(Y_{1it}|U_i, X_{it}, t) = E(Y_{0it}|U_i, X_{it}, t) + \beta = \alpha_i + \lambda_t + \beta + \theta X_{it}$

Linearity makes model (7-3) more restrictive than the instrumental variables model in (6.11). It is necessary to make the treatment of confounding factors manageable when working with panel data. Its advantage is that equation (7-3) will be a better predictor than equation (6-11) because it will have less bias (Freeman, 1984).[15] The coefficients α_i capture the unobserved factors and hold them constant, each acting like a 'dummy variable for each unit' to measure individual heterogeneity.

However, the presence of α_i leads to serial correlation, so that must be accounted for; otherwise it will lead to incorrect standard errors and inefficient estimation. Further, the coefficient α_i is correlated with the predictors, so it leads to omitted variable bias in the estimates of the coefficients. However, if there are omitted variables that *change within the units* over time, threats to bias remain and random-effects would have to be added to (7-3). As indicated above, coefficient α_i is in effect a dummy variable for each unit. It leads us to the idea that a particularly simple and equivalent way of addressing the problem of having units whose differences must be accounted for is to stick with the original data and add categorical variables (dummy variables) for each of the units, omitting one for a unit that can be treated as a reference unit. The reference unit can also be determined from the model above.

Dummy variables offer one additional opportunity, clarified when it is noted that the coefficients on the predictor variables indicate the average effects of the predictors. Fixed-effects models assume that these average predictor effects are the same across all groups, so the model merely considers the average within-group effects. It is often true, however, that at least some predictors have different slopes effects across groups. Instead of the unwieldy tactic of estimating a different regression for each identifiable group, it would be possible to interact (multiply) the dummy variables and the relevant predictors, so that slope dummies for these key predictors are introduced into the analysis.

[15] Freeman, R. (1984). Longitudinal Analysis of the Effect of Trade Unions, *Journal of Labour Economics*, **3**, pp. 1-26.

What has been said so far about different units can also be said for any other grouping in the data, such as industries, communities, and the like. A substantial problem, however, is that we must always think about unobservables that are not nicely set up as such groups – industries, communities, etc. In these cases, and there might be many such, the problem of omitted variable bias is still present and we will need to find good instruments for all the offending unobservables, otherwise we have the additional problem of failure of the identifying restrictions as some of the predictors become correlated with the model residuals. A solution to this problem is to find a host of over-identifying instruments and use them in a 'method of moments' estimation strategy. This added richness of control variables is one of the reasons the search for psychometric indicators is important in many studies of social phenomena. However, some of the instruments may also capture differences across entities that influence the outcome variables. In such a case we are going to need random effects modelling.

Apart from the need to ensure random selection of the cases on the causal variable of interest, a major issue clearly arises in this context as the sufficiency of data to allow any estimation at all. In the estimation of a model with only cross-sectional variation, all possibilities must be explored to identify the full range of time-invariant unobserved factors and include measures for them in the model, either directly or as instruments of endogenous predictors. A large pool of variables, including dummy variables, increases the chance of measuring the effects of the variables that vary across the businesses and which might explain repayment behaviour. One aspect of this is that the larger pool would facilitate probing of interactions among the variables that might pick up the heterogeneity among the units. For example, in a sample of firms there might be industry differences, management quality differences, and the like, along with their interactions among the variables. This would also lower the chance that an alternative model is needed to the one specified. A large enough pool of observations would also allow out-of-sample testing of the predictive power of the model specified. So the strategy assumed here involves: (i) monitoring a sufficiently large number of cases and monitor a large enough number of variables, based on adequate data collection procedures, to allow minimization of bias; and (ii) including a sufficient number of variables, including dummy variables, that can address endogeneity bias created by unobservables.
Once the individual effects are treated as parameters to be estimated, the results are equivalent to estimating individual deviations from the overall means. Essentially this is like first computing the averages and then subtracting from equation (7-3). The result is:

$$7\text{-}6 \quad Y_{it} - \overline{Y}_{\iota} = \left(\lambda_t - \overline{\lambda}\right) + \beta(D_{it} - \overline{D}_{\iota}) + \theta(X_{it} - \overline{X}_{\iota}) + (\varepsilon_{it} - \overline{\varepsilon}_{\iota})$$

The manoeuvre eliminates the impact of the unobserved individual effects. Moreover, with panel data, one can choose the alternative of differencing from period to period, hence by estimating

$$7\text{-}7 \quad dY_{it} = d\lambda_t + \beta dD_{it} + \theta dX_{it} + d\varepsilon_{it}$$

If there are only two periods involved, then the results are the same for (7-6) and (7-7), except for serial correlation in the case of (7-7).

If the process of data collection over time provides data on the same individual for a sufficient number of accounting dates, then one can add a time trend t if it proves important to do so after the individual series are analyzed. Then, the required model with a time trend is not (7-3) but instead

7-8 $Y_{it} = \alpha + \lambda_t + \gamma_1 t + \beta D_{it} + \theta X_{it} + \varepsilon_{it}$

Now, $\lambda_t + \gamma_1 t$ is a common time trend across individuals.

Further, one might find that the outcome variable, say income, also has persistent effects that affect the current outcomes. Consider the current concern with the effects of individual participation in a program of subsidized training provided by steel orchestras. Causal models can be devised to evaluate their impact. It is likely that the most important omitted variables of the fixed effects model vary with time and have variables effects. People seeking to participate in the program might also have suffered some setbacks in earlier life and want to join these programs as a way of improving their learning outcomes and hence their labor market options. Steelbands tend to attract persons who have not done as well as they might in the labour market and have special earnings histories over h periods. At the individual level, for the two-period case, the model takes the form:

7-9 $Y_{it} = \alpha + \lambda_t + \sum_{h=1}^{n} \gamma_h Y_{it-h} + \beta D_{it} + \theta X_{it} + \varepsilon_{it}$

Here, α could measure the average level of vocational skills among the persons seeking training. The causal effect of the steelband-based training program is potentially captured by β. The persistent effects of the outcome variable are measured by γ_h. These might increase or decline with time, and might be interesting in their own right.

The policy problem might also require consideration of both the unobserved but fixed effects that differ among individuals (α_i), alongside the lagged endogenous variables (Y_{it-h}). Vocational orientations and skills might vary among individuals sufficiently to warrant the introduction of these fixed effects into the model. Here, the unobserved individual characteristics, the lagged earnings, and the observed covariates θX_{it} determine who gets trained. The resulting causal model is:

7-10 $Y_{it} = \alpha_i + \lambda_t + \sum_{h=1}^{n} \delta_h Y_{it-h} + \beta D_{it} + \theta X_{it} + \varepsilon_{it}$

If there are at least 3 observation periods, then differencing can be used and the estimation model becomes

7-11 $dY_{it} = d\lambda_t + \sum_{h=1}^{n} \delta_h dY_{it-h} + \beta dD_{it} + \theta dX_{it} + d\varepsilon_{it}$

So, differencing eliminates the fixed effects and gets a tractable estimation model. However, the resulting changes in the lagged endogenous variables become correlated with the residuals and need to be instrumented. Also, earnings are usually correlated from one year to the next, causing $d\varepsilon_{it}$ to be serially correlated and resulting in major problems of consistent estimation. A wide variety of instruments will have to be considered to escape this problem, and instruments are generally hard to come by.

8. The Differences in Differences Method

The fixed and random effects strategies require panel data on the individuals. Most often, data are only available on groups, like industries and regions, and countries. When considering the causal effects of policies at these levels, the source of omitted variable bias must also be unobserved factors at the same level. If group-level omitted variables can be captured by group-level fixed effects, the result is the 'differences in differences' method of identification. Once again, regression analysis can be used to address the issues involved. We can consider these issues using an important consideration in the carnival industry, the impact of the quality of facilities on the sales of carnival vendors. Each year, a substantial share of the budget of the NCC is spent on the construction of facilities in the main carnival locations of Trinidad. Food and craft vendors appear to rely heavily on the carnival industry for a significant share of their sales. However, not all vending is equal, as some vendors appear to benefit differently from the supply of special facilities during the carnival season, depending on their location. Moreover, the quality of these facilities adjust from time to time depending on the policy stance of the NCC. A significant question then is whether the policy of supplying vending facilities has any significant impact of sales per worker of the vendors. Let Y_{ilt} be the sales per capita (or sales per worker) of vendor i in location l in period t. Further, let Y_{1ilt} be the sales per capita (or sales per worker) of vendor i in location l in the carnival season if vending facilities have been established. Let Y_{0ilt} be the sales per capita (or sales per worker) of vendor i in location l in the carnival season if no vending facilities have been established. Since these are potential outcomes, we only get to see one in practice at any location.

Let γ_l be the coefficient on a dummy variable for location l and λ_t be the coefficient on the period of the year when there are no facilities – the non-festive period. This is assumed to be the same in all locations. Further, let D_{lt} be a dummy for the location and festive season when the facilities are constructed. The differences in differences method rests on the assumption that the observed sales model can be written as the linear additive model:

8-1 $Y_{ilt} = \gamma_l + \lambda_t + \beta D_{lt} + \varepsilon_{it}$

for $E(\varepsilon_{it}|l,t) = 0$. Thus, for $D_{lt} = 0$, the conditional expectation for the no-facilities model is the linear and additive form:

8-2 $E(Y_{0ilt}|l,t,) = \gamma_l + \lambda_t$

That is, in the absence of the constructed facilities, sales are determined by the sum of the location effect that is time-invariant and the period effect that is common across locations. The conditional expectation for the facilities model is

8-3 $E(Y_{1ilt}|l,t,) = \gamma_l + \lambda_t + \beta$

The key identifying assumption in (8-1) is that there is a common sales trend in both the facilities locations and the no-facilities location. The construction of the facilities creates a deviation from the common trend, measured by β. The difference between the location and facilities effect is picked up by the location fixed effect, γ_l.

Estimation Strategy for the Differences in Differences Method

As noted above, the tactic of using dummy variables and interactions with relevant predictors often proves helpful in estimation of a differences in differences model in (8-1). In particular, instead of estimating a different regression for each identifiable vendor, the interactions introduce new slope dummies for these key predictors. Suppose there are two locations, say Savannah (l_2) and Carapichima (l_1). One would be captured in the intercept and we would write equation (8-1) as:

8-4 $Y_{ilt} = \alpha + \gamma l_2 + \lambda D_t + \beta l_2 D_t + \varepsilon_{it}$

The model corresponds to equation (8-1) with $l_2 D_t = D_{lt}$. This formulation immediately allows addition of new locations and periods nationally. It also allows investigation of policies other than those described by dummy variables. For example, one might consider all preparatory construction activities by the communities beyond mere physical facilities. Indeed, one might find that construction by the NCC is likely to have more influence on vendor sales in communities with lower levels of construction activity. One might even find that the fraction of young people who get jobs (FE_l, the fraction of young people in total employment) might be good measure of the importance of NCC construction. Let D_t be a dummy for the year when the NCC adopted a policy of nationwide preparatory construction for every carnival, rather than mere main centre interventions, choosing some minimum level of spending in each community. The influence of the expenditures could be tested with a model such as:

8-5 $Y_{ilt} = \alpha + \lambda_t + \beta FE_l D_t + \varepsilon_{it}$

Then, the FE_l variable would measure the impact of the NCC expenditures on sales before the increase while $FE_l D_t$ would capture the impact of the increase in expenditure.

If we have data for only two periods, then on taking differences the model in (8.5) reduces to:

8-6 $dY_{ilt} = \alpha + \beta FE_l + d\varepsilon_{it}$

where, dY_l is the change in sales in location l and $d\varepsilon_{it}$ is the residual of the differenced equation.

The differences in differences method can accommodate covariates X_{ilt} that control for confounding factors. In this case the confounding factors would be location and time-varying factors that do not respond to changes in the basic level of NCC spending on facilities. That is, the model in (8.1) can be extended to

8-7 $Y_{ilt} = \gamma_l + \lambda_t + \beta D_{lt} + \theta X_{ilt} + \varepsilon_{ilt}$

Thus, for $D_{lt} = 0$, the conditional expectation for the no-facilities model is the linear and additive form:

8-8 $E(Y_{0ilt}|l, t, X_{ilt}) = \gamma_l + \lambda_t + \theta X_{ilt}$

That is, in the absence of the constructed facilities, sales are determined by the sum of the location effect that is time-invariant, the period effect that is common across locations, and the effects of the confounding factors, θX_{ilt}. The conditional expectation for the facilities model is then

8-9 $E(Y_{1ilt}|l, t,) = \gamma_l + \lambda_t + \beta + \theta X_{ilt}$

Using Explicit Group Time Trends

If in the process of data collection respondents provide data for different accounting dates to others, then one can add a time trend t if it proves important to do so because one or more years stand out. This also serves to emphasize that other covariates can be easily added to a differences in differences model, just as for a standard fixed effects causal model. The required differences in differences model with a time trend is

8-10 $Y_{ilt} = \gamma_{0l} + \lambda_t + \gamma_{1l} t + \beta D_{lt} + \theta X_{ilt} + \varepsilon_{ilt}$

Here, γ_{0l} will provide a set of intercepts each of which is specific to a location. The coefficient γ_{1l} measures how the location effects adjust over time, while and λ_t continues to be the common coefficient across locations on the period of the year when there is no facilities construction – the non-festive period. This strategy allows different locations to follow different trend effects on sales. In general, the method requires at least 3 years and more so a sufficient number of years to adequately pick up the location-specific trends and allow β to isolate the causal effects of the policy intervention (Besley and Burgess, 2004).[16]

Here again, instead of a time trend, a dummy variable can be added for all but one of the years. The dummy variables effectively hold constant or fix the average effects of each unit, with the standard constant terms doing the same for the one unit that was left out to serve as standard reference. So, they control for average differences across units found in any observable or unobservable predictors whose effects are constant over time, eliminating the cross-unit variations and leaving only the within-unit variation. The dummy variables method works as long as there are enough observations for each unit; the more observations the better. Otherwise it can become impractical.

Carnival Industry-Level Issues

[16] Besley, T. and Burgess, R. (2004). Can Labor Market Regulation Hinder Economic Performance? Evidence from India. *Quarterly Journal of Economics*, 113: 91-134.

For the purpose of this study, one of the important applications of the group fixed effects model is to various aggregates in the carnival sector, even though it is also appropriate to treat the sector as a unit in its own right, for comparison with other sectors. At the carnival sector level, the policy issues surround the idea being widely promoted that the carnival industries are beginning to play a significant role in the transformation of the Trinidad and Tobago economy. It appears to be an avenue for dynamic entrepreneurship and growing productivity as well as export expansion. Thus it might be able to play a significant in building national foreign exchange earning capacity, raising employment and growing per capita income If the sector is outgrowing the national economy, it must also be creating investment capital at a faster rate than any other sector of the economy while generating externalities in the sense of wider and more effective linkages among different sectors. This is effectively a process of transforming the services sector but might also result in some impetus in related ICT activity and manufacturing if it also increases employment of domestic intermediates and capital. What is also interesting is that these tendencies seem to be emerging despite a rising real foreign exchange rate rather than devaluation to make non-oil exports more competitive. This policy stance is linked to fear of inflation that might result from rapid escalation in input costs in an import dependent economy. There is also growing fear that there is an emerging pattern of massive importation of finished carnival goods that is contributing to reduced creativity. It also appears that some of the gains have generated through primitive accumulation, which is to say by relying primarily on retained earnings rather than long term funds made available by a banking sector that is creating new instruments for high-risk sectors. Banking in Trinidad and Tobago still errs in the direction of stringent and restrictive credit guidelines to the carnival sector. Many of these concerns and claims can be addressed scientifically.

Output and Efficiency in Carnival

In this context, an important assignment is to investigate the drivers of sector output per worker using information on sector inefficiency and the extent of reliance on imports to run its production process. One can construct indicators for either of these variables. One might simply use the number of workers employed in the sector in a measure of output per worker (y_{it}). If data are available, it is possible to construct measures of sector efficiency (μ_{it}) using stochastic frontiers.[17] It is also possible to find data on investment in the sector (I_{it}). For sector i in period t we can then write:

$$8\text{-}11 \quad y_{it} = \alpha_i + \lambda_t + \gamma_1 t + \beta y_{j,it} + \theta \mu_{it} + \pi I_{it} + \delta y_{j,it}\mu_{it} + \varphi X_{it} + \varepsilon_{it}$$

The inferences made in the model are only about the sectors of the economy, including those that embody carnival as a satellite. To stress the role of the carnival sector, other services, and manufacturing, we can exclude agriculture and core mining. Here, α_i measures fixed but unobservable macroeconomic and sector institutional influences that remain constant over time and differentiate sectors from each other in terms of productivity; $\lambda_t + \gamma_1 t$ measure the common time trend across the sectors, including a year-specific constant (λ_t) measure of productivity shocks across industries; μ_{it} is an alternative way to consider profitability which is

[17] The method of stochastic frontier modeling was pioneered by Aigner, Lovell and Schmidt (1977) and Meeusen and van den Broeck (1977) and Jondrow, et al (1982).

especially useful when industry operating surplus and interest rate data are difficult to obtain; $y_{j,it}$ is the measure of output per dollar of imports used in the sector, with room to further distinguish imports of intermediates, machinery and equipment from imports of services; and $y_{j,it}\mu_{it}$ is a measure of the possible impact of import dependence (or competition from imports) on efficiency. The variables X_{it} continue to be strictly exogenous confounding factors measured at the sector level. Candidates here include (i) the level of external reserves of the economy, (ii) the average rate of interest, (iii) the broad money supply, and (iv) the going real exchange rate. We would normally expect $\theta > 0$ in all sectors; $\beta < 0$ in some sector but not all, and $\theta < 0$ in those sectors in which competition from imports lower efficiency and hence productivity. We can expect that (y_{it}) and (μ_{it}) and I_{it} are endogenous, reflecting underlying selection bias, but that good instruments are not usually available. Weak instruments would make estimation with OLS inconsistent, so, we could instead use the strategy of dynamic panel estimation by Arellano and Bond (1991)[18], which relies on data-generated weights and GMM estimation (Hansen, 1982).[19] The dynamic panel-data version of the model is:

8-12 $y_{it} = \alpha_i + \sum_{h=1}^{n} \delta_h dy_{it-h} + \beta y_{j,it} + \theta\mu_{it} + \pi I_{it} + \delta y_{j,it}\mu_{it} + \varphi X_{it} + \varepsilon_{it}$

The α_i and ε_{it} must be independent for each i and t, but GMM can yield asymptotically consistent estimates of the causal effects of interest even if α_i are correlated with the covariates.

Investment in Carnival

The panel fixed effects approach in (8.12) can be put to another important policy use. There is much discussion among stakeholders about whether there is a sufficient flow of investment into the carnival industries. The data on gross fixed capital formation in the economy is typically what is used in empirical studies but, being aggregated, it is quite smooth relative to sector investment patterns and perhaps also subject to aggregation bias. So, it is not a good guide. If sector-level data are available, it might be possible to use the historical evidence to study the patterns, without losing the volatility of the sector data. So, if available from the CSO, such data might allow us to understand what drives investment dynamics in the carnival sector. Perhaps influenced by declarations in the 2014 budget, some stakeholders have expressed the view that investment in carnival requires considerable complementarity between public and private investment. Some are also concerned that bank practices and prejudices as well as the high real interest rates and the general economic climate prevailing in the Trinidad and Tobago market affect the private sector's investment in carnival, sometimes negatively. It is also a longstanding claim that high real exchange rates can be a constraining factor.

The Cash Management literature has shown that where banks tend to lack the quality of functions that are dedicated to forging new conceptual approach to the creative sectors of the economy, much is not achieved in terms of the scale of investment needed for structural transformation. New practices, especially ICT-based crowd-funding, are emerging in the model of fundraising but in this setting, it is up to the banks to develop a

[18] Arellano, M. and Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations, *The Review of Economic Studies*, **58**, pp. 277 – 297.

[19] Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica* 50: 1029–1054.

deeper understanding of the dynamics of the various sectors in order to create financial products that foster sustainable creative entrepreneurial activity and related investment growth especially among the SMEs that dominate carnival. Basic responsibilities, such as minimizing financial risks and operational costs, and maximizing cash returns, tend to be the responsibilities of bank Cash Management functions in this process, but banks in the Caribbean tend to remain focused on conservative assessment through the evaluation of cash flow, liquidity, banking management, risk analysis, payment capacity and the associated use of information technology. Matters such as these can be reliably addressed by causal analysis with a panel model such as (8.12), in which the carnival sector, or sectors closely related to it, are identified. To explain the sectorial private investment in a way that also addresses the carnival sector, y_{it} can refer to investment per worker or per dollar of imports used; and the exogenous drivers can be wider government investment (G_{it}), actual sector interest rates (i_{it}), credit availability ($Cred_{it}$), national forex reserves ($RES_{j,it}$), and the real exchange rate (REX_{it}). Then the equation for panel analysis can be:

8-13 $$y_{it} = \alpha_i + \sum_{h=1}^{n} \delta_h dy_{it-h} + \beta G_{j,it} + \theta i_{it} + \pi Cred_{it} + \delta RES_{j,it} + \varphi REX_{it} + \varepsilon_{it}$$

Granger Causality with Multi-period Data

Assuming a sufficient number of years of data on the policy interventions, it is also possible to judge the validity of a policy stance by using the special case of causality testing advocated by Granger (1969)[20] in the context of time series analysis. The key concept involved is a test of whether, once suitable randomisation is otherwise established, causes (causal policy variables) also happen before consequences, and not after. The basic idea was formulated by Wiener (1956)[21] who suggested that if the prediction of one time series is improved by incorporating information from a second time series, then the second has a causal influence on the first.

Reconsider the current concern with the effects of individual participation in a program of subsidized training provided by steel orchestras mentioned above. At the individual level, with panel data available, the Granger causality model takes the dynamic form:

8-14 $$Y_{it} = \alpha + \lambda_t + \sum_{h=0}^{n} \delta_{-h} Y_{it-h} + \sum_{\tau=0}^{m} \beta_{-\tau} D_{it-\tau} + \sum_{\tau=1}^{q} \beta_{\tau} D_{it+\tau} + \theta X_{it} + \varepsilon_{it}$$

Here, α continues to measure the average level of vocational skills among the persons seeking training. However, the causal effect of the steelband-based training program is potentially captured by $\beta_{-\tau}$ but not by β_{τ}. This is the hypothesis of Granger Causality. If past earnings predict current earnings, and in addition, at least some $\beta_{-\tau}$ are statistically non-zero, while all β_{τ} are statistically zero, then we say that at D_{it} Granger causes Y_{it}.

Granger causality can also be used to probe causality at the group level. In the context of our earlier policy concerns at the group level, past values of D_{lt} should contain information that helps predict Y_{ilt} above and

[20] Granger, C.W.J. (1969). Investigating Causal Relations by Econometric Models and Cross-Spectral Methods, *Econometrica*, 37: 424-38.

[21] Wiener, N. (1956). The theory of prediction. In E. F. Beckenbach (ed), **Modern Mathematics for Engineers**, vol. 1. New York: McGraw-Hill.

beyond the information contained in past values of Y_{ilt} alone. Assume that the policy variable D_{lt} changes at different times (say years) in different locations. The NCC might start its upgrade work in POS and then migrate its efforts to other locations around the country in sequence. Practically, Granger causality tests whether, conditional on location and year effects, past D_{lt} always predict Y_{ilt} while future D_{lt} does not predict Y_{ilt}. So, it also tests that Y_{ilt}, as outcome, never causes D_{lt} as an outcome in its own right. Consider the model:

8-15 $Y_{ilt} = \gamma_l + \lambda_t + \sum_{\tau=0}^{m} \beta_{-\tau} D_{lt-\tau} + \sum_{\tau=1}^{q} \beta_\tau D_{lt+\tau} + \theta X_{ilt} + \varepsilon_{ilt}$

The component $\sum_{\tau=0}^{m} \beta_{-\tau} D_{lt-\tau}$ is a basket m lags or post-treatment (policy) effects, essentially measuring the effects of earlier policy interventions undertaken over m periods before time t; and $\sum_{\tau=1}^{q} \beta_\tau D_{lt+\tau}$ is a basket of leads or futuristic (pre-treatment) anticipatory effects of policy changes undertaken after time t. Granger causality requires that $\beta_\tau = 0$, for all τ. If at least some $\beta_{-\tau}$ are not equal zero, then they are also likely to exhibit patterns of increase or decline in the influence of the policy interventions over time (Autor, 2003).[22]

[22] Autor, D. (2003). Outsourcing at will: The contribution of the unjust dismissal doctrine to the growth of employment outsourcing, *Journal of Labor Economics*, 21: 1-42.

9. Country-level Applications

The analysis presented in the previous sections are applicable to any units and aggregations of units, including countries. At the country level, causal analysis can be done using a time series of country-specific data or using a panel of which the country is one case. In either case, the broad effort at causal analysis tends to assume the absence of reverse causation and to use instrumental variables that are known to be unaffected by the outcome variable and therefore remove any prospect that reverse causation might show up through endogenous regressors. In particular, let Y_{it} be the economic outcome of country i, S_{it} be a potentially endogenous (simultaneous) regressor of causal interest, X_{it} summarise all the other exogenous regressors, and e_{it} is the residual error containing all the combined causal information (variables) that are however known to be uncorrelated with X_{it}. Then, for the country-level panel data, we write,

9-1 $Y_{it} = \alpha_{it} + \beta S_{it} + \gamma X_{it} + e_{it}$

Simultaneity, sometimes generated by concurrent events, will raise the prospect that S_t is potentially endogenous, conveying feedback effects. Investment is usually treated as one such endogenous variable. The general cause of the endogeneity is still a version of selection bias generated when data is being collected and aggregated or when the outcome variable is inherently calculated to include the S_t, as is the case with value-added and investment. One way to find an instrument for S_{it} at the national level is to find non-economic variables that influence it but do not influence the outcome at the same time. Geography is one such variable that is widely used; colonial history another (Krugman, 1980;[23] Krugman, 1989;[24] Krugman and Obstfeld, 2000;[25] Henry and Miller, 2009;[26] Asemoglu, Johnson and Robinson, 2001[27]). Another strategy is to find economic instruments that occur so far back in history that they could not possibly be affected by or affect the outcome except through the S_{it} channel. Summarily, at least one instrument in the chosen set must have the property that it "cannot possibly 'be caused' by things going on today." Then, "if paths of causation can be traced through such factors, the direction of causality can be inferred" (Morck and Yeung, 2011: 2). These considerations amount to searching for useful natural experiments at the country level. However, a crucial condition is that the instrument must be measured without error and surely without undue noise and that becomes a substantially increasing challenge the farther back in time one has to go for the instrument. This is the problem encountered by Acemoglu, Johnson and Robinson (2001) when they used mostly concocted data

[23] Krugman, P. (1980). Scale economies, product differentiation, and the pattern of trade, *American Economic Review* 70, 950-959.
[24] Krugman, P. (1989). Differences in Income Elasticities and trends in Real Exchange Rates, *European Economic Review*, 33, 1031-1054.
[25] Krugman, P.R. and Obstfeld, M. (2000). *International Economics. Theory and Policy*, Fifth Edition. Cincinnati: Addison-Wesley Publishing Company.
[26] Henry, P., and Miller, C. (2009). Institutions versus Policies: A Tale of Two Islands. *American Economic Review*: Papers & Proceedings, 99:2, 261–267.
[27] Asemoglu, D., Johnson, S., and Robinson, J.A. (2001). The Colonial Origins of Comparative Development: An Empirical Investigation. American Economic Review, 91: 1369-401.

on colonial settler mortality rates to instrument quality of governance (land appropriation risk) and thus to sort countries by the propensity to use good institutions of governance (to protect private property rights) (Albouy, 2008)[28].

This brings us back to our earlier discussion that a social scientist searching for a natural experiment to ascertain the effect of a certain approach to policy – say democratic versus autocratic policy-making – must be able to sort units, here countries, as if working with randomly assigned treatment and control groups of countries. This is sorting that will adequately distinguish heavily affected countries from lightly affected countries in a way reminiscent of scientific randomization. However, the groups of countries must be identical in all other ways. The only permissible difference between them must be that the policy generates greater effect on one set and significantly lesser effect on the others. One of the great observations of Best (1968) suggests that the colonial powers could not simply be sorted by settler and non-settler groups. Such groups did not simply differ initially on governance grounds. There is also a sharp difference of initial conditions between those non-settler colonies that were colonies of conquest, as in Latin America, and those non-settler that were colonies of exploitation, as in Caribbean countries. European non-settlers are not to be treated as a homogeneous group; Spanish conquistadors pursuing El Dorado's bullion and planters from the British upper class were different breeds altogether. So, the ceteris paribus condition is not satisfied and this compounds the doubts raised by Albouy (2012) about the usefulness of the settler mortality instrument.

What is more, at least one valid instrumental variable must be found for each policy variable of interest. Instruments are hard to come by. As Morck and Yeung (2011: 4) put it, instruments must only vary "in response to exogenous factors, that is factors determined by nature, God, or people whose actions do not depend on the dependent variable in the model". Claims that instruments satisfy these criteria must also be properly validated by the available evidence, with the ceteris paribus condition fully satisfied. The slightest non-random initial difference between the countries that becomes evident over time, or the slightest difference in how a random shock affects countries would "confound the natural experiment into presenting a false picture of what causes what" (Morck and Yeung, 2011:5).

Finding a good 'natural experiment' at the country level is only one task along the road to a causal interpretation. The choice also has the major challenge of being strongly correlated with the policy variable of concern – what scientist call the treatment. This is because the instrument, not the scholar, must be the sorting mechanism mentioned above, essentially randomly assigning countries to the treatment group, those with comparable systems of decision-making in our example, and to the control group, those without such systems laws. An instrument cannot achieve this goal unless it is strongly correlated with the presence of such systems among countries. Once a basket of instruments is proposed, the Stock and Watson (2010)[29] standard for their usefulness is that a joint F statistic below ten in a regression explaining the relevant treatment variable tends to

[28] Albouy, D. (2012). The Colonial Origins of Comparative Development: An Empirical Investigation: Comment. *American Economic Review* 2012, 102(6): 3059–3076.

[29] Stock, J. and Watson, M. (2010). *Introduction to Econometrics, 2nd ed.* Addison Wesley.

have a *weak instruments problem*. Noisy measurement – measurement error to put it bluntly –, such as might be the result of nonrandom sampling, is a major source of weak instruments.

In time series analysis, there is considerable appeal to the power of the principles that (i) the cause must *predate* the outcome, and that (ii) the cause makes unique changes as an intermediate flow of information retained in the outcome. This leads to Granger causality as characterized above but usually exhibited in vector autoregression models. A great challenge here is to know how much precedence to accommodate in the analysis but these tests are ways to learn from the tendency of human history to repeat itself to some degree. However, even in such tests there are pitfalls to be avoided if acceptable causality is to be the outcome. It is well known that for the Granger causality test to be valid, the variables in the model must all be stationary, adjusting without a common trend. However, beyond the matter of the common trend, the unobserved factors must also be addressed, and hopefully they do not show a trend. The problem is that if a third variable causes the set to move together, even if without a trend, then the Granger result is still a false positive.[30] A common trend is not the only way for a latent variable to cause the others. Even more important perhaps is the underlying assumption that reverse causation is not allowed by the method. Economists know well that when, as in the human condition, decisions are made under conditions of uncertainty, expectation about the future might shape current behavior, and when they are realized it can be said that the future is/was the cause of the present. The nexus of the money supply, inflation, growth, and central bank expectations of both inflation and the rate of growth can cause Granger-causality tests to yield false causation by the money supply of both inflation and the rate of growth. Further, these tests can also yield false results when the data are inaccurately measured, when the time series is too short, as they can be when the Arellano and Bond class of panel estimators are used, and when policy revisions based on expectations are timed to confound the independence of the confounding factors.

An Application of Interest to Carnival Enthusiasts

There is some evidence that Caribbean economies have been undergoing structural change, some of it is due to the rise of sectors such as carnival. Policymakers should be interested in whether such structural change might also explain why Caribbean economies are experiencing widespread slowdown in the rate of growth, for example because of reduced rates of indigenous innovation and reduced rates of technological spillovers; or because of inadequate exporting of the main service outputs, specifically education and healthcare. Set in the context of causal modelling as specified above, we can probe the specific question of the causes and growth-effects of structural change, in particular of changes in the shares of sectors to GDP (per capita), employment or trade. We can include the carnival sector among the comparison sectors if suitable work is done to define and measure it in the set of countries studied. Classical economists (Smith, Ricardo, Marx) rooted their understanding of economic development in the forces driving changes in sector shares with cumulative feedback to growth. Their approach was updated by Lewis (1954)[31] and by Kaldor (1957; 1966; 1967; 1978)[32].

[30] This observation was first drawn to my attention in 1982 by David Gordon, my econometrics teacher at the New School for Social Research, New York.

[31] Lewis, W. A. (1954). Economic Development with Unlimited Supplies of Labor. *Manchester School of Economics and Social Studies* 22:417-419.

In these models, understanding of the changing shares is based on comparison of the effects of a number of factors that differentiate sectors, with considerable attention to the skill-intensity of production. These are usually treated as

1. Productivity of resource use, in models in which the intercept picks up the influence of domestic capital volume and residuals pick up the influence of technology or so-called total factor productivity. The primary factor productivities are the arguments in these models
    a. Labour productivity
    b. Import productivity – all types of imports
2. Location of the business operations relative to the production frontier, or sector efficiency, which depends on business climate, defined by creation of
    a. structural capital that favours/elicits work
    b. intellectual property
3. The rate of accumulation of necessary capital
    a. Domestic
    b. Imported

However, to model Caribbean economies with such variables, we need an adequate conception of aggregate capital, and technological heterogeneity of sectors must be assumed from the start. This raises the need for some detailed theoretical thinking. Moreover, as can be gleaned from the discussion above, more updated statistical procedures are now available to treat causation than were available to Kaldor. In the background, we assume that microeconomic data are available from which sector details can be generated for the countries in the panel.

Productivity of Resource Use

Manufacturing and services technology differ and this has implications for understanding how sectors contribute to the size and growth of GDP, employment, or trade. Technology represents how the primary inputs of imports and labour are combined using domestic capital. The central problem of the classical model is the failure to impose developmental ordering on the amount of capital in the economy, by considering its domestic/import structure. Lewis (1954) began that process by pointing to the key importance of the society using surplus labour to create domestic capital. This actually tended to happen throughout the colonial and post-colonial periods, manifested mainly in asset creation by the persons occupied outside the plantations, and later by persons who continued that tradition through local business development. When considering the total capital stock of an economy or sector, it is important use this dichotomy to impose an ordering while recognizing that both domestic and imported capital matter in the production process. We can use an appropriate mean of the quantities to combine domestic capital; one which takes account of the different roles and significance of domestic and imported capital as well as their properties as sums of values. From a local

[32] Kaldor, N. (1957). "A Model of Economic Growth", *Economic Journal* 67:591-624; (1966). Causes of the Slow Rate of Growth in UK. Cambridge UP; (1967). **Strategic Factors in Economic Development**. Cambridge UP; (1978). "Causes of the slow rate of growth of the UK" in **Further Essays on Economic Theory**. London. Duck Worth.

perspective both domestic capital and imported are growing exponentially when development is occurring, so we can use a logarithmic mean of financial values to combine them. This will also have the effect of recognizing that for development to occur both types of capital are needed and that a situation in which neither type is present is tantamount to having no developmental capital and no development. In particular, let K_j be the real value of imported capital and K_d the real value of domestic capital. Then, for the quantity of developmental capital, κ, we represent the sector development capital function as:

$$9\text{-}2 \quad \kappa = \begin{cases} 0 \; if \; K_j = 0 \; or K_d = 0 \\ K_j \; if \; K_d = K_j \\ \frac{K_j - K_d}{ln(\frac{K_j}{K_d})} \quad otherwise \end{cases}$$

Now, to relate these estimates of the asset pool to the sector production process, we assume that when accumulating assets and adding liabilities, the component firms are always looking around for ways to improve the assets and accumulate its capital component. So, they are always trying to upgrade technology, and accordingly to create increasing returns as an intended consequence of the investment undertaken. Success is another matter, but in some sectors at least some firms get better at this sort of thing (learning by trying and doing, so to speak), goodwill and intellectual property grow, and their growth drives the overall relative growth of those and linked sectors (Frankel, 1962).[33] Moreover, the successful firms tend be become important on the global stage when they are also global leaders of the changing business model. This is currently happening in the carnival sector.

Overall resource productivity still matters for firm success but this depends in part on (i.e., is proportional to) the overall pooling of domestic and imported capital that firms accumulate in the sector. Let the total capital of a sector be

$$9\text{-}3 \quad K_s = \sum_{j=1}^{n} \kappa_j$$

For η a positive exponent that summarizes the amount of augmentation due to knowledge diffusion among firms, and α the productivity of capital, we can write the relationship between sector productivity of capital and this pool of capital as:

$$9\text{-}4 \quad A = \alpha(\sum_{j=1}^{n} \kappa_j)^{\eta}$$

At the level of the firm, let κ_{jl} be the firm's optimal capital per worker. Its optimal output rate is an interaction of its optimal capital and labour with the efficiency of sector capital and the scale of resources it has available. However, its actual output rate also depends on the efficiency with which it employs the technology available in

[33] Frankel, M. (1962). "The Production Function in Allocation of Growth: A Synthesis". American Economic Review 52, 995-1022.

the sector and that depends on how closely it operates to the frontier productivity defined by A. Let that level of efficiency be represented by $\xi \leq 1$, on appropriate scaling. As indicated above, efficiency in this sense can be independently measured using the tools of stochastic frontier analysis and available microeconomic data. Crucial here are tools developed to take advantage of panel/longitudinal data and allow the target measures of inefficiency and hence risk to adjust over time.[34]. The output rate of the firm is then given by

9-5 $Y_j = \xi_j A {\kappa_{jl}}^{\theta} L^{1-\theta}$

Firms operate on the technological frontier only if $\xi = 1$. However, all firms still face the same sector frontier technologies and factor prices. If at the level of the firm we set $L = 1$ for convenience the typical firm's frontier capital per worker is given by

9-6 $\kappa_j = \kappa_{jl} = \frac{K_s}{N_s}, for\ all\ j$

where K_s is defined by (1.2). Then,

9-7 $\sum_{j=1}^{n} \kappa_j = K_s$

and (1.3) becomes

9-8 $A = \alpha {K_s}^{\eta}$

It follows that output per firm is given by

9-9 $Y_j = \xi_j \alpha {K_s}^{\eta} (\frac{K_s}{N_s})^{\theta}$

Thus, the level of output in the sector is

9-10 $Y = \alpha {K_s}^{\eta} (\frac{K_s}{N_s})^{\theta} \sum_j \xi_j$

$$= N_s \alpha {K_s}^{\eta} (\frac{K_s}{N_s})^{\theta} = \alpha {N_s}^{1-\theta} {K_s}^{\eta+\theta}\ iff\ \xi_j = 1\ for\ all\ j$$

One can move from an equation such as (9.10) to use as dependent variable the growth in the sector share in GDP, the growth in the sector share of employment, or labour productivity growth. The rate of growth of real GDP per capita (using mid-year or end year population and unadjusted for degrading of energy resources) can also capture the economy-wide forces affecting a sector. This variable also corresponds well with the national concern with development, equity and poverty reduction. In such a dynamic setting the investment rate in the

[34] (Battese and Coelli, 1992; 1995; Kumbhakar and Lovell, 2000).

sector emerges as an explanatory variable alongside the growth of sector efficiency; and either or both of these can be the focus of causal policy analysis and monitoring.

Closure

In the model above, there is little doubt that capital is endogenous, and will cause related bias, so we need an identifying equation for it. This is sometimes referred to as a 'closure' in economic analysis. It is the mechanism through which the model embodies specific assumptions about economic causation. This mechanism informs about the primary drivers that are critical to the behaviour of the model and economy. Not any assertion will do, but at the sector level Frankel gave considerable attention to realistic competition in the investment process, much as Kaldor had done. Different methods of addressing endogeneity bias – different closure rules – lead to different model behaviours. For the model closure to emphasize the causal primacy of competition, observe that the standard identity of accumulation applies, so for $g(K_s)$ the regeneration and expansion process and δ the rate of depreciation, we can write:

9-11 $\frac{dK}{dt}_s = g(K_s) - \delta K_s$

Now, to accommodate sectors such as carnival, two sets of issues must be accommodated in the specification of $g(K_s)$ in the sectors. The first is that firms compete directly, often in cutthroat fashion. So logistic type functions are admissible as $g(K_s)$. The assertion of a quadratic is an assertion that competition has a primary role in the model and is a key to understanding the behaviour of the sectors. The second is that the underlying savings rate adjusts with success in capital accumulation. So, the savings rate, s_s, is endogenous. Summarily, one can then write

9-12 $\frac{dK_s}{dt} = s_s(t)K_s(1 - K_s) - \delta K_s$

At the same time, sector investment depends on the inflation rate, the exchange rate and the rate of interest, which together shape the extent of Dutch disease. Use of an equation such as (9.12) makes it necessary to find a suitable instrument for $s_s(t)$ and that reopens issues we have already covered.

Estimation

In probing the causes and consequences of changing sector shares, the analyst can use data on the growth of output and the GDP sectors shares from national statistical organisations or from websites such as the UNSD's or the World Bank's. Some series available for at least 1970-2012. Typically, because of the assertion that the long-run dynamics of sector shares and economic growth are interdependent, tools such as panel co-integration, error correction and Granger Causality are favoured in probing the specified relationships even though some preliminary cross-sectional work can be done for comparison with the panel results (Dutt and Lee, 1993[35]; Necmi, 1999[36]). Set in the context of prior determination of whether the variables are all I(1), the co-

[35] Dutt, A. K. and K.Y. Lee (1993). The Service Sector and Economic Growth: Some Cross-Sectional Evidence. *International Review of Applied Economics* 7: 311-329.

integration analysis will allow determination of whether long run relationships exist between the sector shares. Panel unit root tests will be needed to assure consistency and Stata reports a good assortment of these tests, in particular the test of Levin, Lin and Chu (2002)[37] and the Fisher PP (Choi, 2001)[38], on the other.[39] The error-correction analysis will allow determination of whether there exists long-run structural adjustment between the various pairs of sector shares in the economy. In particular, we could consider whether shocks in the shares of services self-correct. Assuming relevant steps to create I(0) series, the Granger causality tests would tell us whether there is a feedback relationship between sector share and growth, or whether it is a one way drive from sector share to growth. They would also tell us whether services are indeed emerging as engines of growth. A side effect is clarification for professional economists as to whether the relationships are equilibrating phenomena.

On Causal Poverty Monitoring

Once we can get the analysis of output per worker right, we can use the information to do causal monitoring of social impact and there is a rich literature pointing to how this might be done, taking account of the problems and challenges we have identified so far. Much of it focuses on poverty and inequality, and there is considerable consensus that if we can get the sectors right and grow the economy, then the per capita incomes of the poor will also tend to improve. Here, the reference set of indicators is the MDG indicators, which were designed to allow cross-country comparison of poverty performance on commonly agreed global initiatives. A coherent analytical framework will tend to use data on these indicators to provide a general comparable interpretation of how poverty is reduced, and by how much. Differentiation of performance on the MDG indicators will tend to lead only to likely differentiation on poverty performance. While much of the literature focuses on country-level aggregates, much can also be done with sector data in a multi-country setting. The poor, like all households, are identified with the industries in which they earn a living, even when that industry is 'general government'. It is still appropriate to assume that in Trinidad and Tobago poverty is the main social problem that the national community wants to eliminate decisively. This follows in part from its participation in the 1992 Rio Declaration[40] and the Rio+20 outcome document "The Future We Want"[41]. In the Rio Declaration, the principle was recognized that "eradicating poverty and reducing disparities in living standards in different parts of the world are essential if we are to achieve sustainable development whilst meeting the needs of the majority of the people." Focus on poverty reduction is also a direct reflection of the Millennium Declaration.[42] The commitment to poverty reduction is a basis for achieving the rights defined in the Declaration (dignity, freedom, equality, a

[36] **Necmi, S. (1999)**. Kaldor's Growth Analysis Revisited. *Applied Economics* 31:653-660.

[37] Levin, A., Lin, C.-F., and Chu, C.-S. J. (2002). Unit root tests in panel data: Asymptotic and finite-sample properties. *Journal of Econometrics* 108: 1–24.

[38] Choi, I. (2001). Unit root tests for panel data. *Journal of International Money and Finance* 20: 249–272.

[39] See also, Hadri, K. (2000). Testing for stationarity in heterogeneous panel data. *Econometrics Journal* 3: 148–161; Maddala, G. S., and Wu, S. (1999). A comparative study of unit root tests with panel data and a new simple test. *Oxford Bulletin of Economics and Statistics* 61: 631–652.

[40] http://www.unep.org/Documents.Multilingual/Default.asp?documentid=78&articleid=1163.

[41] http://www.uncsd2012.org/content/documents/727The%20Future%20We%20Want%2019%20June%201230pm.pdf.

[42] http://www.un.org/millennium/declaration/ares552e.htm.

basic standard of living that includes freedom from hunger and violence and encourages tolerance and solidarity), with feedback.

While the other rights (dignity, freedom, non-violence, tolerance and solidarity, etc.) have proven to be more difficult to address empirically, viable efforts to monitor poverty outcomes across countries in ways that are sensitive to inequality date back to broad claims by Kuznets (1955).[43] The Kuznets hypothesis was that growth is not inherently poverty-reducing, especially in developing countries, because the poverty outcome is linked to inequality and, in the early stages of development, inequality necessarily tends to increase to facilitate transformative investment. Since Kuznets, the data to investigate such a hypothesis has improved dramatically, especially because of the work of the World Bank on the measurement of poverty and the efforts by countries to monitor performance on the MDGs (Deaton, 2003)[44]. From the viewpoint of a method that can address selection bias, multi-country 'mixed-effects' models that take account of inequality are now being estimated that relate the poverty rate (or its rate of reduction) to GDP per capita (or its rate of growth), depending on whether the levels are cointegrated across the data (Balakrishnan, et al, 2013;[45] Bourguignon, 2004;[46] Dollar and Kraay, 2002;[47] Beck, et al, 2005[48]). In most such models, growth is the country-specific indicator expected to generate the randomly varying effects among countries, while inequality is the indicator that is expected to generate common effects. The models can also be represented in ways that allow for applications of Granger-causality tests, simultaneously characterizing the outcomes of growth of per capita income, change in inequality and change in poverty. A good example of what can be done in a general structural framework can be found in the work of Beck, et al (2005), which extends the established frameworks of poverty analysis described below to investigate scientifically how the development of the small business sector can lead the process of poverty reduction via the process of economic growth. Inclusion of inequality in these multi-outcome models is particularly powerful from the standpoint of the MDGs, because many of the initiatives under MDG2 through MDG8 can be viewed through the prism of efforts to spread the benefits of growth while it reduces poverty. For example, MDG2 achieves this through upgraded access to education; MDG3, by bringing women more adequately into the wealth stream and improve governance; and MDG4-MDG6, by improved access to health care.

Cross-country numerical monitoring will require a common standard of success in poverty reduction across the globe, to say nothing of sufficiently common periods of measurement. One weak option is to use the fraction of persons living below the country absolute standard of poverty – the headcount ratio defined with the country

[43] Kuznets, S. (1955). Economic growth and income inequality. *American Economic Review*, **45**(1): 1-28.
[44] Deaton, A. (2003). MEASURING POVERTY IN A GROWING WORLD (OR MEASURING GROWTH IN A POOR WORLD. NBER WORKING PAPER SERIES, Working Paper 9822, NATIONAL BUREAU OF ECONOMIC RESEARCH. Available at http://www.nber.org/papers/w9822.
[45] Balakrishnan, R., Steinberg, C., and Syed, M. (2013). The Elusive Quest for Inclusive Growth: Growth, Poverty, and Inequality. IMF Working Paper, Asia and Pacific Department, IMF.
[46] Bourguignon, F. (2004). The Poverty-Growth-Inequality Triangle. The World Bank. Available at http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.58.469&rep=rep1&type=pdf.
[47] Dollar, D. and Kraay. A. (2002). "Growth Is Good for the Poor." *Journal of Economic Growth* 7(3): 195-225.
[48] Beck, T., Demirgüç-Kunt, A., and Levine, R. (2005). SMEs, GROWTH, AND POVERTY. NBER WORKING PAPER SERIES Working Paper 11224. Available at http://www.nber.org/papers/w11224.

poverty line (Bourguignon, 2004: 1; Fosu, 2011: 4-9;[49] Balakrishnan, et al., 2013: 9). The drawback here is that different countries set different standards, so adjustments over time do not necessarily refer to the same populations. Nevertheless, the most straightforward models use the headcount measure (Balakrishnan, et al., 2013: 9). Let P be the headcount poverty rate, y per capita income, and $GINI$ be the measure of inequality. Also, let i refer to country and t year, with ρ a dummy for the relevant decade. The updated multi-country mixed effects poverty model takes the basic form

9-13 $$\ln P_{i,t} = \alpha_i + \beta_{i,d} \ln y_{i,t} + \gamma \ln GINI_{i,t} + \rho_d + \varepsilon_{i,t}.$$

Here, α_i measure the fixed (intercept) effects that differentiate country performance, while the $\beta_{i,d}$ measure the (random) variable slope (response elasticity) effects that differentiate country growth performance on poverty. The ρ_d pick up whether different decades of effort have made any difference in the per capita income of the poor.

Another approach, adopted by some leading scholars (especially linked to the IMF and World Bank), is to use the average income of the bottom quintile of income distribution as the main indicator of achievement in poverty reduction (Dollar and Kraay, 2002: 195; Beck, et al, 2005: 13; Balakrishnan, et al., 2013: 11). The work by Dollar and Kraay has proven to be quite contentious because of its methodological flaws. For example, Bourguignon (2003)[50] observes that a significant omitted variable is the change in the distribution of income attending growth and change in the poverty rate. Another is that the time between measurements in different countries is so irregular, from 5 to 35 years, as to make it impossible to specify the population in relation to which inference is drawn (Ashley, 2007).[51] From a conceptual viewpoint, perhaps the main disadvantage is that the poverty measure lacks the multidimensional characteristics represented in the MDGs, so these indicators cannot be treated as its linear decomposition.

The main advantage of the approach is that it inherently addresses inequality, if related to the rate of economic growth. In particular, the inclusiveness of growth, which is to say whether or not growth occurs with falling inequality, can be considered without estimates of inequality, along the lines of Dollar and Kraay (2002) and Balakrishnan, et al (2013). Let $yp1$ be per capita income in the bottom quintile and y per capita income. Also, let i refer to country and t year, with ρ a dummy for the relevant decade. The Dollar and Kraay (2002) model is essentially a random effects model that can be written with categorical time indicators as

9-14 $$\ln yp1_{i,t} = \alpha_i + \beta_{i,d} \ln y_{i,t} + \rho_d + \varepsilon_{i,t}.$$

---

[49]Fosu, A.K. (2011). Growth, Inequality and Poverty reduction in Developing Countries. UNU Working Paper #11. Available at http://www.wider.unu.edu/publications/working-papers/2011/en_GB/wp2011-001/.

[50] Bourguignon, F. (2003). The Growth Elasticity of Poverty Reduction: Explaining Heterogeneity across Countries and Time Periods. In T. S. Eicher and S. J. Turnovsky (Eds.), *Inequality and Growth* (Cambridge, MA: MIT Press.

[51] Ashley, R. (2007). Growth May Be Good For the Poor, but Decline is Disastrous: On the Non-Robustness of the Dollar-Kraay Result. Economics Department (0316), Virginia Tech, Blacksburg VA 24061. Telephone: (540) 231 6220; fax: (540) 231
5097; e-mail: ashleyr@vt.edu.

Here, α_i measure the fixed (intercept) effects that differentiate country performance, $\beta_{i,d}$ measure the (random) variable slope (response elasticity) effects that differentiate country growth performance on poverty, and ρ_d pick up whether different decades of effort have made any difference in the per capita income of the poor. The whole model can be reconstructed in terms of income shares (Balakrishnan, et al., 2013: 11). Models such as these typically require good instruments for income, and as we have seen above these are difficult to come by. Balakrishnan, et al (2013:9) use lags of real per capita income from the national accounts (adjusted by data from the Penn World Tables) to instrument the measure of per capita income (consumption) obtained from living standards measurement surveys.

One of the advantages of using the headcount indicator of poverty to consider the growth effects is that it is also straightforward to measure the amount of change in the extent of poverty that have occurred over a given period, say a decade or 15 years. Then, instead of using time dummies, these measures can be used to determine directly the main factors explaining the progress over the relevant period. The fixed-effects form of the model by Dollar and Kraay (2002) is the best-known of this type. Let $yp1$ be per capita income in the bottom quintile and y per capita income. Also, let i refer to country and t year. Then, considering the achievements from say 2000 to 2014, and whether growth is inclusive in the process, one would write,

9-15 $$ln\frac{yp1_{i,2014}}{yp1_{i,2000}} = \alpha\, ln\, yp1_{i,2000} + \beta\ ln\frac{y_{i,2014}}{y_{i,2000}} + \varepsilon_{i,t}.$$

The initial conditions $\alpha yp1_{i,2000}$ monitor the divergence (convergence) among countries (Beck, Levine and Loayza, 2000)[52]. Clearly, growth is good for the poor, and is also inclusive in that process, if $\beta > 0$.

All of these methods are illustrative of the kind of causal analysis that is possible as well as the challenges encountered. The application by Beck, et al (2005) is interesting in this context because it illustrates that other non-income explanatory factors can be used to explain progress on poverty; in their case the contribution of small enterprises to employment. Here, for S the share of small enterprises in employment, one would write,

9-16 $$ln\frac{yp1_{i,2014}}{yp1_{i,2000}} = \alpha\, ln\, yp1_{i,2000} + \beta\ ln\frac{y_{i,2014}}{y_{i,2000}} + \gamma S_i + \varepsilon_{i,t}.$$

It is straightforward to see that all of this work can be done at the sectoral level within a single country. As noted, cross-country numerical monitoring will require a common standard of success in poverty reduction across the globe. While the other issues of causal analysis will not disappear, that problem of a common standard of poverty will not plague cross-sectoral analysis within a country, so in considering the methods above a country can resort to national cross-sectoral data only and still be reasonably well-guided.

[52]Beck, T., Levine, R. and Loayza, N.. (2000). Finance and the Sources of Growth. *Journal of Financial Economics* 58, 261-300.

10. Summary - the Return to History and Democracy

All of the considerations raised so far suggest that considerable scientific work has to be put into the design of the studies on which the NCC will make its policy decisions. Sound policy design requires prior determination or timely post-implementation monitoring of the causal relationship(s) of interest to the policymaker – determination of what causes what effect by considering what would happen in the counterfactual worlds with and without its investments and interventions. Once this responsibility is accepted, the NCC must then consider what data-gathering process would yield the information about the causal effect of interest.

The basic threat is selection bias, which can sometimes overwhelm the causal effect, and the basic call of science is for experimental or quasi-experimental data to eliminate it. These ideals amount to random assignment to treatment groups that are otherwise identical and they are often expensive and sometimes impractical. They are increasingly being used in policy analysis and the NCC can push to use them as much as possible. It makes sense for the NCC to insist on random or quasi-random assignment of policy groups whenever studies are being done with the resources it commits to develop its policy framework. In any event, in practice the NCC should push analysts to come as close as possible to conditions that mimic such experiments. Random sampling by, or in collaboration with, the CSO is one major way to achieve this, especially if care is take to ensure that the confounding factors are properly balanced across the treatment groups of interest to the NCC. The reason is that it can be shown that the effect of randomly assigning to the treatment and non-treatment groups is the same as the effect of applying the treatment to randomly selected cases (Angrist and Pischke, 2009: 15). Proper statistical comparison of means on the confounding factors is crucial to achieving such assurance and this can also be addressed by the CSO. Where randomization cannot be achieved directly, good instruments are needed to approximate it and sound econometric and other statistical inference techniques are needed for identification. Granger causality tests provide one good strategy of identification and inference. However, good instruments are hard to come by and Granger causality cannot deal with reverse causation.

Policy questions that cannot be answered by such techniques are normally treated as unanswerable (Angrist and Pischke, 3009: 5). Fortunately, there are other ways to proceed. This situation cries out for adequate and detailed analysis of the historical context of policy and the underlying claims of causation – a return to both history and democracy as suggested by Best (1968; 1971[53]) and Best and Levitt (1969)[54]. However, this must be history dedicated to connecting events with both internal and external logical consistency of one sort or another and with adequate effort to ensure that sound data is used in the process (Morck and Yeung, 2011: 11). Even more important, the NCC could insist on the value of the independent exercise of free will and the

[53] Best, L. (1971b) Size and Survival. In Girvan and Jefferson (1971), *Readings in the Political Economy of the Caribbean*. Kingston: New World. pp. 7-28pp. 29-34.
[54] Best, L.A. and Levitt, K. (1969). Externally Propelled Industrialization and Growth in the Caribbean: Selected Essays, Vols. I-IV. McGill University Unpublished Manuscript. Available at the University of the West Indies Library, Mona.

institutionalization of democratic choice as a path to determination of cause (Best 1971).[55] The key point is that if independent thought exists, then its outcomes can be exogenous and hence causal in the most important senses of the notion of cause. If scientific outcomes are starting conditions, and if all data domains and interests are in the dialogue about proximate cause, then the resulting policy interventions are properly treated as the origins of causal chains in their own right, producing effect and cause in some sequence.

For *a-priori* policy design with concern for monitoring causal effects, some formal guidelines can be found in the Holland (1988) encouragement design: (i) deliberate experimental manipulation of the doses of an incentive regime or encouragement to perform an activity of importance to the NCC and its stakeholder community; (ii) measurement of the subsequent amount of the encouraged activity targeted by the incentive regime, such as feting in the fete-worker discipline example; (iii) measurement of a final outcome or response variable, such as worker discipline in the example; and (iv) an abiding joint interest in the interface of science and policy, and hence in measuring the causal effect of the encouraged activity on the ultimate desired outcome. This is not exactly how Best's conception of free will or democracy is normally interpreted in Caribbean analytical circles, so clarification with the example is worthwhile here. In that case, the key feature of free will is the self-selection of the amount of feting in response to the incentives. This is completely under the control of the randomly selected worker. From the viewpoint of science, fete in this case is potentially both a cause of worker discipline and an effect of the policy incentives.

Encouragement or incentive designs can particularly useful to the NCC, because they go directly to its responsibility to design and introduce a suitable range of incentives and opportunities for democratized choice to underpin development of the carnival industry. Analogous to the amount of feting in this example, the treatments or causes of interest ultimately must be voluntarily applied by the stakeholders of the industry, and this is boosted to various degrees by the process of democracy itself. Training can be provided to vendors to engage in healthy practices, but ultimately the practices are heavily influenced by the choice of vendors to cooperate with society. Inducements of various sorts might be given to foreigners to visit during the carnival season, but ultimately, the spending behavior reflects the individual interpretation of the utility of response. Effective application of such a design would require considerable understanding of the society in which the application is being done.

[55] Independent Thought and Caribbean Freedom. In Girvan and Jefferson (1971), *Readings in the Political Economy of the Caribbean*. Kingston: New World. pp. 7-28; Best, L. (1975).