Working Paper Series

WORKING PAPER NO. 715

COMPLEXITY, DIVERSITY AND STABILITY DEBATE: EVIDENCE FROM CALIFORNIA COUNTY DATA

by

Andrew Dabalen and George Goldman

WAITE MEMORIAL BOOK COLLECTION
DEPT. OF AG. AND APPLIED ECONOMICS
1994 BUFORD AVE. - 232 COB
UNIVERSITY OF MINNESOTA

DEPARTMENT OF AGRICULTURAL AND RESOURCE ECONOMICS
BERKELEY
CALIFORNIA AGRICULTURAL EXPERIMENT STATION

University of California

DEPARTMENT OF AGRICULTURAL AND RESOURCE ECONOMICS
DIVISION OF AGRICULTURE AND NATURAL RESOURCES
UNIVERSITY OF CALIFORNIA AT BERKELEY

WORKING PAPER NO. 715

COMPLEXITY, DIVERSITY AND STABILITY DEBATE: EVIDENCE FROM CALIFORNIA COUNTY DATA

by

Andrew Dabalen and George Goldman

California Agricultural Experiment Station
Giannini Foundation of Agricultural Economics

JUNE 1994

Complexity, Diversity and Stability Debate: Evidence From California County Data*

Andrew Dabalen and George Goldman
Department of Agricultural and Resource Economics
University of California, Berkeley.

Draft

July 6, 1994

Abstract

A linear statistical model is used to test the hypothesis that higher levels of economic diversity are associated with more stability. We also compare the performance of several proposed indices of diversity in this linkage. The results show little support for the primary hypothesis, and a mixed verdict for the second.

1 Introduction:

Ever since the industrial revolution, interest in regional development has been a persistent theme in development debates. With increasing global competition, commitment to the process has been embraced with single-minded enthusiasm. As an objective, regional development aims at availing good quality of life for citizens. When realized, its defining features are hard to miss. Among them is the notion that *stability* of per capita incomes, employment prospects, etc. are essential to the process.

From a conceptual point of view, however, there is much that is not understood. More to the point just what leads to the realization of a successful

*We would like to thank Dr. John Wagner, Syracuse University, for calibrating all the indices of diversity used in this study.

regional development agenda is not easy to sort out. Although conceptual difficulties are widely acknowledged, that has not deterred regional scientists from proposing plausible bases for a viable and stable economic growth strategy. For sometime now, *diversity* of economic activities, has been commonly believed as a strategy most likely to lead to economic *stability*.

In the literature, the link between stability and diversity often appear as a form of a conjectured super-additivity of productive activities by virtue of the latter's relatedness. This study revisits that supposed linkage, and unlike most previous studies that were either national in scope or comparative across select states, [illegible] it down for counties in the state of California. As a case study, California is suitable for many reasons. Its economy is generally diverse. As a consequence of purposeful pursuit of stable economic growth strategies in the 1950s and 1960s large fluctuations have been rare. Furthermore, diversity across its many counties, from Los Angeles to Alpine, provide reasonably good sample data.

The actual organization of the study proceeds in the following manner. First, in part **2**, we shall give a brief discussion of the conceptual framework. In section **3** we present the data and the variables used, which is followed in **4** by statistical estimation procedures. In section **5** we undertake a discussion of the results. In **6** we conclude.

2 Stability and Diversity:Definitions

A lack of consensus over the association between economic diversity and stability is more than a quarrel over inconsistency of empirical results derived through rank correlations or Ordinary Least Square results. The real dispute revolves around the meaning of these concepts. This pushes the debate into the realm of theory.

A recent contribution to the literature, [9], has pointed out these theoretical difficulties of measuring diversity. These authors posit that an implicit acceptance, in previous literature, of the size of the economy as a proxy for diversity has been misleading in purpose. But if the earlier literature suffered from such naive associations, its more recent replacements which are analogs of concepts from other disciplines are conspicous for their narrow focus. Consider two of the more commonly used measures;*entropy* and *portfolio*.

Entropy is commonplace in the natural sciences. It is typically used to capture the complexity of organization. When it was later picked up by information theorists, it began to be used as a proxy to measure the average amount of information in a channel. The term average here refers to expectations of a random variable. Economists, eager to infer the level of connectedness or synergy in an economy–whatever the scale– and its influence over structure, were naturally attracted to the concept. Entropy, in regional economic models, captures the diversity of economic activity.

Typically, the procedure relies on using the distribution of employment or sales from various industries to infer the distribution of economic activity. Ideally, a diversified economy is assumed to have equal levels of activity, say employment, across industries. By consequence, concentration of activity in a few sectors means less diversity. As Emil Malizia and Shanzi Ke,[8], inform us, "diversity is not only the absence of specialization, *but* the presence of multiple specialization." The theoretical and empirical objections raised against the use of entropy are summarized in John Wagner and Steven Deller, [9].

Entropy indices of diversity achieve a maximum where there are many heterogeneous economic activities whose distribution is even. The other alternative measure to entropy, portfolio analysis draws upon a different analogy. Its construction imagines a social planner who is entrusted the task of making decisions about an ideal social industrial structure. The analogy is to an intelligence—individual decision maker, city planning commission, county administration, etc.— that makes optimal decisions through minimizing risk. Given the objective of the planner, it is not surprising to conclude that the optimal and preferred choice will be a mean-variance efficient portfolio,[1] a result well-known in portfolio finance literature.

Entropy measures are additive, intuitive and easily computable. Unfortunately, they are also mechanistic and therefore problematic in the following ways. First, aggregation is limited only to production relatedness. Restricting similarity to shared raw materials or manufacturing technology wishes away complex horizontal and vertical integration processes common among industrial firms. Second, businesses that are assumed to be related across 2–digit SIC codes are equally assumed to have the same level of synergy– a fact which is not evident, because some relatedness ought to be more synergistic

[1] A portfolio is mean-variance efficient compared to others if either, for the same mean it has less variance or for the same variance it has a higher mean.

than others, [3].

Portfolio analysis captures interindustry linkages well, because it explicitly incorporates covariances in the portfolio. Still, the complexity of the phenomenon renders formulation of any single measure daunting. It is why, for instance, we have several functional specifications of entropy measures.[2] Recently, Wagner and Deller, [9], proposed a new way to capture the "inherent interdependence" between industries by using information revealed by the Input-Output (I-O) tables.

The measure of diversity proposed by these authors is a composite multiplicative index with three elements:

- *size,* or the number of endogenous industries located in region i relative to the number of same in a reference economy–the nation or the state.

- *density,* or the number of non-zero elements in the **I-A** matrix relative to the square of endogenous industries, and

- *condition number,* a ratio of the largest to smallest condition numbers of the **I-A**. Of the three components that make up the composite, it is this index that captures interindustry linkage. The larger the condition number, the more diverse the economy.

In one sense, preoccupation with a precise and clear construction of a measure of diversity, though often cast as a problem in theoretical comprehension, belies the real motivation of the controversy. Imprecision over the meaning of diversity compromises the requisite strategies development planners need to achieve their real objective : stability in growth.

Like diversity, unanimous agreement over the definition of instability has proven difficult. While several methods have been proposed, here we mention, briefly, three common ones. The structural method postulates the dependence of employment or output on a set of random and trend variables. Employment sequence is then decomposed into separate components due to random and permanent shocks, [1].

[2]See **Appendix A** for more details.

[3]For a detailed discussion on this matter see [9].

The second method is due to Stephen Beveridge and Charles Nelson, [2]. They too decompose the level of activity into transitory and permanent components. Let $z(t)$ be the permanent component and $c(t)$ the random part. Suppose $z(t)$ grows at a rate $m+\epsilon$, where m is deterministic and ϵ is stochastic. Then,

$$c(t) = z(t+k|t) - z(t) - km$$

for a suitably large k. Notice that $c(t)$ is the forecast random k periods ahead.

The suitability of this method is that it avoids the empiricism of regression on time, discussed below. However, the procedure still must invoke a particular decomposition —numerous forms exist— of the underlying time series.

The third method is also the simplest and it is one of the measures of instability adopted in this study. To separate what is permanent from what is cyclical, simply regress the level of activity against time. Formally,

$$e_{it} = \alpha + \beta t + \epsilon,$$

where, e_{it} is employment in region i in time period t. Instability is then measured as,

$$INSTAB_i = \left[\frac{\sum_{t=1}^{T}(\frac{e_{it}-\hat{e}_{it}}{\hat{e}_{it}})^2}{T}\right]^{\frac{1}{2}}$$

where,
e_{it} = total employment for county i in period t.
$\hat{e}_{it}$ = linear approximation of long run employment trend, that is, $\hat{\alpha}+\hat{\beta}t$, for each i.

To measure and isolate cyclical component, first we purge the data of seasonality and randomness through use of mean quarterly data adjusted for any random variation. We then regress this mean annual data on time. Subtraction of the mean, $\hat{e}_{it}$ from the seasonally adjusted e_{it} approximates the cyclical component. To account for differences in scale among regions we divide the difference by $\hat{e}_{it}$. The summation completes approximation of overall measure of regional economic instability.

3 Data Sources and Variable Description:

3.1 Data Sources.

The sample data is for 58 California counties. Our interest is to determine the connection between diversity and structural instability among California's rural counties. According to the state's classification, areas with an urban population of over 50,000 would be considered metropolitan. We ignored the official determination of the county status and instead reclassified, arbitrarily, metropolitan counties as those with a population of 1 million and more except in the rare cases as that of San Francisco county, with a population of 750,000 but having no rural population.

The data was compiled from three main sources;

1. California Statistical Abstract, for population density ($POPDEN$) and per capita income ($PECAINC$).

2. various departments of the state of California[4],

3. MicroIMPLAN data bases for some diversity measures.

3.2 Variable Description.

3.2.1 Dependent Variables.

Four dependent variables were used in this study to capture the notion of structural instability. Unemployment, $UNEMPYT$, and instability, $INSTAB$, whose measurement was mentioned in section **2** were chosen because they are common measures of instability. Both concepts convey to us, in an intuitive sense, the extent of resource allocational distortions.

Structural changes in any economy can also be captured by the long-run time series data of per capita income. Ideally, a time-series on income distribution for each county would more accurately summarize structural changes. However, gini coefficients, which such a data would reveal, are not easily available. Still, lack of a better data series is not the excuse for our use of per capita income level as an alternative measure of structural change. Instead, it is included here because it is a good approximate quantification

[4]For more specificity, see the **Appendix A**

of the efficiency of resource transformation, just as much as unemployment captures its lack.

Finally, we used deviation of a county's per capita income from that of the state's, $DEPEINC$, so as to gauge how successful a structural change a county is undergoing compared to the state. Note that implicit in this argument is the notion that California state is a diversified economy. Calculation of $DEPEINC$ is simple;

$$DEPEINC_i = \frac{PECAINC_{st} - PECAINC_{it}}{PECAINC_{st}}.100$$

where,
$PECAINC_{st}$ = per capita income of the state (s) at time t.
$PECAINC_{it}$ = per capita income of county i at time t.

3.2.2 Independent Variables.

A descriptive summary of all the independent variables used appear in **Appendix A.** The industrial sector classification data was obtained from *California's Department of Finance.* The 1969 – 1974 uses 1967 SIC—or the Standard Industrial Code. Shared SIC are usually used as measures of relatedness of business activities. The most common grouping of industries using SIC is through their production relatedness. The 1975 – 1987 is based on 1972 SIC. While the 1988 – 1991 data is based on 1987 SIC.

Income by source and major industry is aggregated for two reasons:to avoid high multicollinearity, and; to account for the major sources of diversity. The resulting variables are $PERMAN$, or percent of income source due to manufacturing, $PERPRI$, percent of income source from primary industries, $PERSERV$, percent of income source from service sector, and $PERGOVT$ or income source from government sector. The actual computation is shown in **Appendix A.**

$POPDEN$ is included because it is assumed to be important in determining instability. Other factors, excluded from the four aggregates above that would be considered important are, wages and salaries, $OTHLABIN$, of county residents employed by international organizations, and foreign embassies and consulates based in the U.S. Other than wages and salaries, whose distribution is accounted for in the four aggregates, the other important source of income is proprietory ($PROPRINC$).

4 Statistical Methods

We do not know of any general model of diversity. Nor is there yet a way to theoretically choose between the different diversity measures. This phase of our work therefore aims at demonstrating which of the indices performs better empirically. In particular, we compare how well IND, Wagner and Deller measure of diversity computed from **I-O** tables, explains structural stability(or its lack thereof) compared to the long-standing entropy measures.

The statistical model itself is the standard simple linear model.

$$Y_i^v = X\beta + \epsilon$$

where, Y_i^v =The Nx1 dependent variable for county i, and
$v = UNEMPYT$, $DEPEINC$, $PECAINC$, or $INSTAB$.
X = the NxM independent variables tabulated in A.
$\beta = M$x1 vector of coefficients, and
ϵ = the error term.

In carrying out the estimation, we were concerned that the metropolitan counties would bias the sample, in the sense that they will affect significantly, the level and the distribution of stability. Two ways to detect such a bias involve running a dummy variable model, where all the metropolitan counties are dummied or running the model without the metropolitan counties. First of all, the DUM–the dummy variable which takes on the value of 1 for metropolitan counties and 0 otherwise– was not significant in any of the models reported below.[5] Neither was there any meaningful qualitative changes to the results when the metropolitan counties were ommited. For purposes of maintaining more degrees of freedom, the dummy variable models with all the counties is the one we chose to report.

The actual estimation procedure took three steps. In each stage, the dependent variables remained the same. All the independent variables, except the diversity measures were also maintained throughout.

1. In the first stage, we ran a model with all the entropy indices– except

[5]There is an economic reason why metropolitan counties are not any more stable than the other counties not classified as such. Tight linkages between industries in a county may contribute to greater diversification, but they also tend to have large positive covariances. On average, in a stability sense, they are not necessarily more diversified.

$NDIV^6$– and IND, all defined more precisely in **appendix A**. The results appear in **table 1**.

2. Next we looked at the various components that make up IND. The results of that model is **table 2**.

3. To check for heteroskedasticity, which is a common problem with cross-section data, we plotted the residuals of the linear models in stages (1) and (2) against the independent variables. We then weighted each of the models above with $POPDEN$. The results are reported in **tables 3** and **4** respectively.

5 Discussion of the Results:

Except for $INSTAB$ model, the $F-$statistic and the R^2 point to overall significance of the models.

(a)$UNEMPYT\ MODEL$:

Reliance on natural resource sectors, $PERPRI$, is destabilizing while services, $PERSERV$, are not. That on the surface of things seem to be the surprising result. The coefficient on $PERPRI$ is persistently positive while that of $PERSERV$ is negative. This is true whether the model is weighted by $POPDEN$ or not.

In order to contemplate the plausibility of this, it is worthwhile to undertake a thought experiment. Consider two economies that are the same in everything except their levels of service and natural resource sector. For simplicity, suppose one economy has all the natural resource sector and no service sector and vice versa. Imagine now that both economies experience a shock at the same time. The shock could be a disastrous forest fire, a devastating fish disease, etc., or a stock market crash.

Starting from the same point, shocks on natural resources resemble waves with long cycles while those of services look shorter. If we postulate stability to mean the frequency of an economy to be on the long-run equilibrium mean, the shorter cycles of the service sector come back to the mean more often than those of natural resources for any shock and specified time interval.

[6] We ommited $NDIV$ throughout the estimation because as a concept it contributes nothing to our understanding of the problem at hand.

Table 1: Models With Aggregate IND.

Variable	$UNEMPLYT$	$DEPEINC$	$PECAINC$	$INSTAB$
$INTERCEPT$	-146.6097 *	620.8034 *	149963 *	0.2484
DUM	1.4488	-13.516	-2812.078	0.0013
$PERMAN$	0.0912	-1.985	-143.124	0.00074
$PERPRI$	0.845*	-1.472	-306.29	0.00036
$PERSERV$	-0.3662*	3.174*	660.536*	0.00033
$PERGOVT$	0.0179	-1.169*	-243.315*	-0.000099
$POPDEN$	0.0011*	-0.005*	-1.12*	-0.0000011
$OTHLABIN$	0.0000049	-0.0000010	-0.0002	-1.29×10^{-9}
$PROPRINC$	-0.0000020	-0.0000027	-0.00057	8.55×10^{-10}
$TRANSPRT$	-0.0000022	0.0000033	0.0006	4.30×10^{-10}
$LDIV$	41.034*	-156.66*	-32594*	-0.034
$HDIV$	293.09*	-2278.88*	-474121*	-0.287
$ODIV$	1.195*	2.747	571.69	-0.00063
$PDIV$	-0.5603*	3.130*	651.28*	-0.0010
IND	-0.0812*	0.1069	22.24	0.000034
R^2	0.719	0.748	0.748	0.175
Adj. R^2	0.578	0.623	0.623	-0.23
$Prob > F$	0.0001	0.0001	0.0001	0.95
D.F.	28	28	28	28

$* =$ Significant at 0.05 or less

$\star =$ Significant at 0.10

Table 2: Models When Components of IND. Are Used

Variable	$UNEMPYT$	$DEPEINC$	$PECAINC$	$INSTAB$
$INTERCEPT$	-116.6097*	385.30	100968	0.076
DUM	1.486	-13.373	-2782.27	0.001
$PERMAN$	0.115	-2.22*	-463.67*	0.0007
$PERPRI$	0.7712*	-0.775*	-161.26	0.0004
$PERSERV$	-0.340*	3.045*	633.65*	0.0002
$PERGOVT$	-0.0086	-0.981*	-204.18*	-0.000039
$POPDEN$	0.0010*	-0.0051*	-1.077⋆	-0.0000
$OTHLABIN$	0.0000034	0.000012	0.0025	1.419×10^{-9}
$PROPRINC$	-0.0000011	-000010	-0.0021	-6.73×10^{-10}
$TRANSPRT$	-0.0000018	-0.000000	-0.000026	-3.46×10^{-10}
$LDIV$	35.944*	-117.95*	-24540*	-0.023
$HDIV$	205.04	-1599.01*	-332674*	-0.095
$ODIV$	1.35*	1.386	288.36	-0.0009
$PDIV$	-0.446	2.15	447.89	-0.0012
$SIZE$	-110.81	699.69	145572	0.123
$COND$	-4.066	31.09	6469.85	0.0089
DEN	30967	-283907	-59066834	-49.39
R^2	0.735	0.774	0.774	0.205
Adj. R^2	00.572	0.635	0.635	-0.28
$Prob > F$	0.0003	0.0001	0.0001	0.96
D.F.	26	26	26	26

$*$ = Significant at 0.05 or less

$\star$ = Significant at 0.10

Table 3: Weighted OLS With *IND*.

Variable	*UNEMPYT*	*DEPEINC*	*PECAINC*	*INSTAB*
INTERCEPT	-81.324*	550.89*	135420*	0.134
DUM	1.322	-12.577	-2616.70	0.003
PERMAN	0.266	-2.96*	-615.88*	0.0009
PERPRI	0.881*	-2.14	-446.65	0.00032
PERSERV	-0.219*	4.232*	880.48*	0.00037
PERGOVT	0.059	-1.86*	-388.04*	-0.000068
POPDEN	0.00055*	-0.0084*	-1.76*	-0.0000015*
OTHLABIN	-0.0000	0.000037	0.0077*	$-5.95\text{x}10^{-9}$
PROPRINC	0.0000	-0.000025*	-0.0052*	$3.22\text{x}10^{-9}$
TRANSPRT	-0.0000	-0.0000	-0.0017	$1.81\text{x}10^{-9}$
LDIV	23.877*	-139.81*	-29089*	-0.037
HDIV	141.34	-1551.98*	-322890*	0.18
ODIV	0.706	-0.005	-1.232	-0.0033
PDIV	-0.287	0.393	81.94	-0.0011*
IND	-0.064*	0.160	33.30	0.00010
R^2	0.779	0.936	0.936	0.355
Adj. R^2	0.668	0.905	0.905	0.03
$Prob > F$	0.0001	0.0001	0.0001	0.397
D.F.	28	28	28	28

$* =$ Significant at 0.05 and less.

$\star =$ Significant at 0.10

Table 4: Weighted OLS With Components of IND.

Variable	$UNEMPYT$	$DEPEINC$	$PECAINC$	$INSTAB$
$INTERCEPT$	-43.46	308.54	84998	0.09
DUM	1.49	-14.44*	-3004.28*	0.003
$PERMAN$	0.272	-3.28*	-684.24*	0.0011*
$PERPRI$	0.731*	-0.522	-108.63	0.00022
$PERSERV$	-0.15	4.000*	832.33*	0.00023
$PERGOVT$	0.064	-1.85*	-385.86*	-0.000090
$POPDEN$	0.0003	-0.0078*	-1.62*	-0.000001
$OTHLABIN$	0.0000014	0.000047*	0.0098*	-6.45×10^{-9}
$PROPRINC$	0.0000015	-0.000032*	-0.0067*	3.40×10^{-9}
$TRANSPRT$	-0.00000	-0.0000095	-0.0019	2.23×10^{-9}
$LDIV$	18.861*	-117.17*	-24378*	-.027
$HDIV$	-23.18	-232.94	-48463	0.259
$ODIV$	0.908	-3.36	-700.14	-0.002
$PDIV$	-0.213	-0.231	-48.12	-0.0012*
$SIZE$	-130.38	1140.94*	237373*	-0.012
$COND$	-6.74*	48.49*	10089*	0.005
DEN	42957	-439306*	-91397628*	17.764
R^2	0.815	0.949	0.949	0.377
Adj. R^2	0.701	0.918	0.918	-0.005
$Prob > F$	0.0001	0.0001	0.0001	0.95
D.F.	26	26	26	26

$* =$ Significant at 0.05 and less

$\star =$ Significant at 0.10

Back to the results. *POPDEN* leads to higher levels of unemployment. Theil's entropy measure of diversity, which is the key diversity index used in Malizia and Ke,[8] gives a result that is contrary to what conventional knowledge expects. If diversity is stabilizing, we should expect a negative coefficient on *LDIV*. Another entropy measure, *PDIV*, is significant and theoretically consistent only in the unweighted model. Although the numbers suggest that *IND* has much less influence over unemployment than *LDIV*, the measure is significant and theoretically consistent. Among its constituent elements, only *COND* is consistent–with respect to theoretically expected sign of coefficient– and significant after the model is corrected for heteroskedasticity.

(b)*DEPEINC MODEL*: Recall that *DEPEINC* is a percentage measure of relative performance of a county's economy to that of the state. Put simply, we want to measure the distance or the difference, $d(x_{st}, x_{it})$, between per capita income of the state (s) in time t, x_{st}, and that of a county i in time t, x_{it}. If the initial difference was large but narrowing over time, we say that the county is doing better because we are assuming that the state economy is diversified. Such a narrowing of difference implies that the county's economy is converging, in a diversity sense, to that of the state. Similarly, if the difference was small initially but increasing over time, we say that the county's economy is not diversifying as much as that of the state. All this makes sense only if we make a further assumption that no county's economy has been or is now more diversified than that of the state.

PERGOVT, in both the weighted and unweighted versions of the model, reduces the deviations of counties' per capita income from that of the state. What this information suggests is that decentralization of government services to the local level has a stabilizing effect. *PERPRI* loses its significance when the model is weighted. On the other hand, with weighting *PROPRINC* becomes significant and confirms our intuition– that more proprietory ownership has a tendency to reduce the gulf between the state and a county.

The sign on *POPDEN* is at first surprising. It suggests that *POPDEN* has the tendency to decrease the stability gap between the state and the counties. Our intuition is that increasing populations strain a region's resources and may be destabilizing. While such reasoning may be true, the results here still make sense as long as the variation that is explained is that of convergence. If a county is as big as Los Angeles, the results are obvious.

But this is even more true for smaller less populous counties.

PERSERV is positive and significant throughout. The interpretation of this result ultimately falls victim to the conceptual problematics of *DEPEINC*. Ideally, we would like the state's per capita income to be a stationary reference to which county incomes are measured. But the state's per capita income is not held constant in calculating *DEPEINC*. Besides, it is difficult to imagine that cycles of the state's economy have a negligible impact on a county's economy.

The biggest contribution towards reducing structural differences between the state and the counties are *LDIV* and *HDIV*. Again the surprise is that the contribution of *HDIV* to reducing *DEPEINC* is greater than all other variables. This is surprising because, all our theories are more sympathetic to a view that condemns concentration of economic activity into a few firms. Part of the condemnation stems from received theory that argues that diversity is unarguably better.

(c) *PECAINC MODEL*: As our intuition would predict, *PERSERV* add to structural stability and *POPDEN* decreases it, according to this sample data. Another consistently significant influence over per capita change is *PERGOVT*. The negative sign on the coefficient supports a standard argument, albeit controversial, which claims that increases in government expenditures (running a budget deficit) do lead to long-term crowding out effects and less growth.

Among diversity measures, *IND* is not significant in any of the models. The entropy measures do, however, explain this model very well. *HDIV* is not only significant but has the intuitively correct sign. It does lose its significance when the model is weighted and only the constituents of *IND* are used. Theil's entropy measure (*LDIV*) is– despite its significance–tricky to interprete. In particular, the unexpected negative sign may be a pointer to other subtle issues that are not explicitly included in the formula. More specifically, aside from the positive linkages, intersectoral linkages may have large cyclical covariances. So that, the information we may be getting from this variable is that, for the duration of the study, the negative covariance dominate. As we should expect, *PDIV* has the expected sign and is significant.

The aggregate *IND* measure is not significant in either the weighted or unweighted model. However, when we disaggregate it, all three of the components become significant–which raises suspicion over the formulaic strcuture of $IND. COND$ which is supposed to capture endogenous industrial composition of a county yields results that are contrary to our intuition. It suggests, as a first guess, that having a relatively high number of industries located in a region is not a sure way of predicting positive per capita growth.

(d) $INSTAB\ MODEL$: This model's results are the most surprising and puzzling. They are confounding because in principle, what it measures is approximated by $UNEMPYT$. And yet the results here bare no resemblance to those we obtained in **5**. As a general remark, none of the variables is significant, except $PERMAN$ and $PDIV$–and even these only when the model is weighted.

But what is really surprising is that, $INSTAB$ as is commonly defined and popularized in the literature is not so easily explainable. Unlike other studies, such as [8], [9], that have claimed some real causal relations between $INSTAB$ and diversity using national data, this semi-micro data for a single state reveals no such association. Save for $PDIV$, no index of diversity is statistically significant in this model. In fact, the model specification itself is suspect. The $F-$ statistic is insignificant at 0.10.

The preceding discussion on the results are particular to a model. For a summary, let us list some common results that we find interesting.

- Some common biases that hold value-adding extractive industries – natural resources, manufacturing–in a higher esteem than services, is suspect. The latter is stabilizing and the former are not.

- Where they explain variation in any of the dependent variables, the performance comparisons between entropy indices and IND or its elements, is mixed. None has any overwhelming consistent explanatory power over another. Note that we are not hereby saying that $PDIV$ explains structure as well as IND or $LDIV$ or any of the other representative indices. Rather, we are saying that between an entropy measure such as $LDIV$ or $HDIV$ on the one hand and IND on the other, no clear indisputable explanatory power of one over another can be claimed.

- Some indices give counter-intuitive results. That is to say, some of the formulaic constructions of diversity do not lead to signs (qualitative statistical interpretation) that we expect of them.

- The more popular diversity indices do not explain $INSTAB$. We have claimed that this last result is particularly interesting if for no other reason but as a caution against unequivocal acceptance of received theories.

6 Conclusion

This study tests the linkage between diversity of economic activities and successful development. Using a simple linear model, we tested this relationship using four commonly accepted proxies for measuring successful economic transformation (these are the four dependent variables) on California counties. Our results reveal two important insights. First, there is no statistical support for the proposed positive linkage between diversity and stable growth. Second, none of the two classes of diversity measures— the Wagner-Deller measure and entropy indices— distinguish themselves as superior.

From one perspective, the results of the $INSTAB$ model are disappointing. This is so if our intention was to use the data simply as a witness, a procedural step to confirm our deeply held theories. The problem though could be that we are asking too much of our simple constructions. In particular, the failure of the diversity indices to explain $INSTAB$ may be due to various reasons. Below we highlight two potential problems and suggest how we intend to resolve them. Such an undertaking will inform the next phase of our research.

1. Complexity and agglomeration have attracted a lot of attention from regional economists recently. This literature on agglomeration confirm that non-linear models provide a better fit to the pattern of spatial industrial structure that we observe than linear models. In one study, [6], an assumption of initial distribution of spatial activity as even – a common assumption in constructing entropy indices of diversity– still led to a concentration of activity. In fact, an even distribution of manufacturing activity is unstable. What this paper suggests is that perhaps our linear specification of the model may hide more than it reveals. Formulation of more flexible functional forms will be part of next stage of this study.

2. Besides the linearity of the specified models, the other missing dimension in most of the existing studies is time. Whether the measures of diversity are drawn from **I-O** tables or from employment data, they are static. Yet as Kraybill and Dorfman, [5], argue

   > An improved understanding of the dynamics of intersectoral income and expenditure cycles is important for evaluating the timing and magnitude of local economic impacts .....

   The advantage of using dynamic models rest on the fact that intra- and interindustry relationships can be explained easily. They are the short-run deviations of the state-space component of the model,[7]. And intuitively, the deviations are due to regional capacity constraints, labor supply shifts, changes in relative prices and as Fafchamps, [4], recently posited, localized externalities –both positive and negative. The use of state-space dynamic models to define diversity more meaningfully, will be a topic for further study.

A Appendix1

Formulations of Entropy Measures of Diversity.
Altogether five different specifications of diversity measures have been suggested. In all cases, except $HDIV$, greater value of the indices indicates greater relative diversification, while lower values indicate more specialization. The suggested formulas are;

$$PDIV_i = \frac{e_{id}}{e_i}.100 \tag{1}$$

$$ODIV_i = \sum_{s=1}^{S} \frac{\left[\frac{e_{is}}{e_i} - \frac{1}{S}\right]^{\rho}}{\frac{1}{S}} \tag{2}$$

$$HDIV_i = \sum \left[\frac{e_{is}}{e_i}\right]^2 \tag{3}$$

$$NDIV_i = \sum_{s=1}^{S} \frac{\left[\frac{e_{is}}{e_i} - \frac{e_s}{e}\right]^{\rho}}{\frac{e_s}{e}} \tag{4}$$

$$LDIV_i = -\sum_{s=1}^{S} \left[\frac{e_{is}}{e_i} \ln \frac{e_{is}}{e_i}\right] \tag{5}$$

where,
e_{is} = activity, typically employment, in county(region) i in industry s.
e_i = is the total activity in region i.
e = is total activity in the base economy, here the state.
e_{id} = is the activity in durable manufacturing within region i.
ρ = a positive constant taking the value of 1 or 2.

The other important contribution to the literature on diversity measures that have been used in this study is the Primary Diversity Measure, (PDM), and here simply denoted as IND proposed by John Wagner and Steven Deller. Using **I-O** tables, these authors characterize a diversity measure of the form;

$$PDM_i = SI_i * DEN_i * CN_i$$

where,

$$SI_i = \frac{N_i}{N_{be}}$$

And,
$SI_i =$ is the relative size of the **I-A** matrix of region i.
$DEN_i =$ is the density of **I-A** matrix and is furthermore defined as

$$DEN_i = \frac{NON - ZERO_i}{N_i * N_i}$$

$N_i =$ the number of endogenous industries identified by MicroIMPLAN.
$N_{be} =$ consists of the number of endogenous industries in the reference economy,
$NON - ZERO_i =$ the number of non-zero elements in **I-A** matrix, and,

$$CN_i = |(I - A)||(I - A)^{-1}| = \frac{\delta_1(I - A)}{\delta_n(I - A)}$$

where,
$|(I - A)| =$ is the norm of the (I-A) matrix,
$\delta_1(I - A) =$ is the largest value of the (I-A) matrix, and
$\delta_n(I - A) =$ is the smallest value of the (I-A) matrix.

The greater the deviation of CN_i from 1, the more diverse the economy. Also there is a positive association between density, DEN_i, and diversity. SI_i does not capture the network or linkage of industries as conveyed in DEN_i or CN_i. However, there is an implicit admission in this measure, as in the previous ones that it seeks to supersede, that the larger the size the better the economy, especially with regard to ability of the latter to weather shocks.

A Appendix2

Data Preparation and Description: Our data is limited to social and economic variables. Since our main interest is in explaining structural shifts, emphasis was placed on samples of variables relevant to structural economic issues. Consequently, besides unemployment, employment and per capita income, we also collected data on various sources of personal income:(1) Nonfarm or farm; (2) By type of industry–manufacturing(nondurables, durables), construction, mining, etc.

Even with the structural variables, our focus was directed at their distribution. One way to highlight this entails compiling the proportion of employment activity by industry. This is the approach taken by Malizia an

Ke, [8]. Our procedure is different only in the data we used–county income from each industrial type. Note that, in principle, these two approaches should tell the same stories.

In keeping with the persistent theme of specialization or diversification, we then aggregated the sources of income for each county into four major sources;

- **Primary Industry**–Agriculture, Forestry, Fisheries, Mining.
- **Manufacturing**–Durables, Nondurables.
- **Services**–Finance, Insurance, Retail and Wholesale.
- **Government**–State, Local and Federal governments' payrolls.

The variables used and their description are tabulated below.

Variable	Description
Dependent Variables	
$UNEMPYT$	Unemployment data compiled from the documents of *Economic Development Department*(EDD)
$DEPEINC$	Deviations of a county's per capita income from that of the state.
$PECAINC$	County per capita income compiled from *California Statistical Abstract*(CSA)
$INSTAB$	Instability as measured in section **2** from employment data (EDD).
Independent Variables	
DUM	Dummy variable for the metropolitan counties.
$PERMAN$	Percent of a county's income source from manufacturing; $= \frac{DUR+NONDUR}{TPERSINC}.100$, where $TPERSINC$ is total personal income for each county.
$PERPRI$	Percent of a county's income source from natural resources $= \frac{AGRICUL+FORESTRY+FISHERIES}{TPERSINC}.100$
$PERSERV$	Percent of income source from service sector, $= \frac{WHOLESALE+RETAIL+FIN+INS+SERV}{TPERSINC}.100$
$PERGOVT$	Percent from government payroll. $= \frac{STATE+LOCAL+FED}{TPERSINC}.100$
$POPDEN$	Population density for each county.
$PROPRINC$	Proprietor income.
$TRANSPRT$	Income from transportation.
$LDIV$	Theil's entropy measure of diversity.
$HDIV$	Herfindahl's index measure of concentration.
$ODIV$	Ogive measure of diversity.
$PDIV$	Percent share of durable goods industrial activity as a measure of diversity.
$SIZE$	Size of (I-A) matrix.
$COND$	Condition number of the (I-A) matrix.
DEN	density of the (I-A) matrix.
IND	Aggregate $PDM = SIZE * COND * DEN$.

References

[1] Peter Berck, Diana Burton, George Goldman, and Jacqueline Geoghegan. *Instability in Forestry and Forestry Communities*, chapter Fourteen, pages 315–338. Emerging Issues in Forest Policy, edited by Peter N. Nemetz. University of British Columbia Press, Vancouver, BC, 1992.

[2] Stephen Beveridge and Charles Nelson. A new approach to decomposition of economic time series into permanent and transitory components with particular attention to the measurement of the "business cycle". *Journal of Monetary Economics*, 7(2), 1981.

[3] Rachel Davis and L.G.Thomas. Direct estimation of synergy: A new approach to the diversity-performance debate. *Management Science*, 39(11):1334–1346, 1993.

[4] Marcel Fafchamps. Localized externalities and the pattern of economic development. Draft paper delivered at ARE-UCBerkeley Seminar, January 1994.

[5] David S. Kraybill and Jeffrey H. Dorfman. A dynamic intersectoral model of regional economic growth. *Journal of Regional Science*, 32(1):1–17, 1992.

[6] Paul Krugman. Fluctuations, instability, and agglomeration. Seminar Paper delivered at UCBerkeley, Spring, 1994.

[7] James P. LeSage. Modeling dynamic interindustry regional growth in the presence of structural shifts and outliers. *Journal of Regional Science*, 33(3):365–385, 1993.

[8] Emil E. Malizia and Shanzi Ke. The influence of economic diversity on unemployment and stability. *Journal of Regional Science*, 33(2):221–235, 1993.

[9] John E. Wagner and Steven C. Deller. A measure of economic diversity: An input-output approach. Staff Paper 93.3, September 1993.