Urban Studies and Planning Faculty Research
Publications

Urban Studies and Planning

8-6-2008

A New Index for Comparing the Diversity of Population Inflows and Population Stocks

George Galster
Wayne State University, GEORGE_GALSTER@WAYNE.EDU

Tatiana Homonoff
MDRC

A New Index for Comparing the Diversity of

Population Inflows and Population Stocks

George Galster*

Hilberry Professor of Urban Affairs

Department of Geography and Urban Planning

Wayne State University

Detroit, MI 48202

Ph.: 313-577-9084 fax: 313-577-0022; email George_Galster@wayne.edu

Tatiana Homonoff

Research Analyst

MDRC

16 West 34th Street

New York, NY 10016

August 6, 2008

* Contact author

ABSTRACT

A New Index for Comparing the Diversity of

Population Inflows and Population Stocks

The paper introduces a new "diversification index" (DIV), which compares the composition of the current or recent population inflow and the composition of pre-existing population stock, with positive (negative) values signifying a process generating more (less) diversity in the stock. Higher absolute values for DIV signify larger differences in the composition of the inflows and the pre-existing stocks of population. DIV is easy to compute and interpret, adaptable to handle population inflows or outflows, and widely applicable to a variety of phenomena.

The paper defines DIV, discusses its properties, and calculates it for several hypothetical cases as a way of showing its intuitive appeal, such as how it would reflect a neighborhood gentrification scenario. DIV indices for both race and income groupings are computed from 1992 to 2006 for three neighborhoods in Chicago to demonstrate how inter-temporal trends in DIV provide insights into neighborhood dynamics. Finally, the paper discusses extensions, potential weaknesses, and other caveats related to the use of DIV in future applied research.

A great deal of social scientific effort has been invested in the development of numerous indices for describing multi-group compositional characteristics of a *stock* of population defined by differences in some demographic, economic, ethnic or other dimension. Examples of such indices include nominal entropy (Theil's Information), ordinal entropy, and Simpson's *D* (Hirschman-Herfindahl); see Reardon and Firebaugh (2002) and Reardon et al. (2006) for evaluative reviews. Less research has been devoted to measuring compositional characteristics of *flows* of population, though the aforementioned indices typically may be applied to measuring flows as well as stocks. No index has yet been developed, however, that *compares the compositions of an inflow of population and the baseline stock of population[1] on a group-by-group basis.* Such a comparative index would be useful for gauging the degree to which the composition of the inflow either matches that of the stock (and thus the degree to which the current stock is tending to remain stable) or differs from it (and thus the degree to which the current stock is tending toward more or less diversity over time due to its inflows).

The ***"diversification index" (DIV)*** introduced in this paper aims to do exactly this. DIV compares the composition of the current or recent population *inflow* and the composition of pre-existing population *stock* on a group-by-group basis, with positive (negative) values signifying a process generating more (less) diversity in the stock. Higher absolute values for DIV signify larger aggregate differences in the composition of the inflows and the pre-existing stocks of population, based on group-by-group comparisons. DIV is easy to compute, easy to interpret, and widely applicable to various phenomena. This approach is quite distinctive from standard measures of inter-group diversity, which indeed could be applied to both a flow and a stock to suggest a difference in their degrees of diversity. These standard measures do not, however,

---

[1] Population can be thought of either as residents of a geographic area or members of a non-spatial collection, such as occupation or standard industrial classification.

compare on a group-by-group basis the inflows to the current stocks to ascertain if the inflows are working to increase or decrease diversity of these stocks by increasing or decreasing the shares of each.

The paper begins by formally defining DIV and discussing its properties. It then proceeds to calculate DIV for several hypothetical cases as a way of showing its intuitive appeal. A further application of DIV to a hypothetical time path of a neighborhood undergoing gentrification-led displacement of population follows. Next, DIV for both race and income groupings is computed from 1992 to 2006 for three neighborhoods in Chicago, to demonstrate how an examination of the inter-temporal trends in DIV provides interesting insights into neighborhood dynamic processes. Finally, the paper discusses extensions, potential weaknesses, and other caveats related to the use of DIV in applied research.

Specification of the Diversification Index (DIV)

For any given population stock arrayed in N different, mutually exclusive and exhaustive groups (including those with zero current members) measured at the beginning of period t (described by percentage shares of the total population stock, $[S_i]$) and an inflow of population measured in percentage shares across the same N groups ($[F_i]$) during period t:

$$DIV_t = \sum_{i=1}^{N} \Delta SF_{it}$$

and the difference in percentages between the inflow F and stock S for a particular group i is:

$$
\Delta SF_{it} \left\{
\begin{array}{ll}
= (F_{it} - S_{it}) & \text{if } S_{it} < 100/N \text{ and } F_{it} \leq 100/N \\
= (S_{it} - F_{it})/(N\text{-}1) & \text{if } S_{it} > 100/N \text{ and } F_{it} \geq 100/N \\
= (100/N - S_{it}) & \text{if } S_{it} < 100/N \leq F_{it} \\
= (S_{it} - 100/N)/(N\text{-}1) & \text{if } F_{it} \leq 100/N < S_{it} \\
= (100/N - F_{it})/(N\text{-}1) & \text{if } S_{it} = 100/N \leq F_{it} \\
= (F_{it} - 100/N) & \text{if } F_{it} \leq 100/N = S_{it}
\end{array}
\right.
$$

DIV is based on the mathematical property that maximum diversity of either a stock or flow occurs when each group represents 100/N percent of the entire population. The intuition behind the scoring of ΔSF_{it} for each of the possible values of F and S above is as follows. With both F and S below their values of what would constitute maximum diversity (100/N), DIV is scored positively to the extent that F exceeds S, and negatively to the extent that F falls below S and thus pulls S farther from 100/N. The opposite logic applies when both F and S are greater than their values constituting maximum diversity. When these two parameters bracket 100/N the explanation is a bit more complicated. When S is less than but F is greater than 100/N, DIV is scored positively only to the extent that the gap between S and 100/N is closed; any "excess" flows greater than 100/N percent are ignored because, if perpetuated, they would push the stock beyond 100/N eventually. The opposite logic applies when S is greater than but F is less than the maximum diversity percentage. When S is at the point of maximum diversity, any F unequal to 100/N reduces ΔSF_{it} to the degree of this inequality.

The N-1 weight is used above to make scales comparable. When N>2 and the F and/or S for the given group is > 100/N, the potential range of variation in │F-S│ is greater than if the elements of this term were < 100/N.

DIV will assume its minimum value of -100 when the least diverse inflow (only one group represented) impinges on a maximally diverse stock (all N groups equally represented). DIV will assume its maximum value of 100 when the least diverse stock

(only one group represented) is impinged upon by a flow in which *all* other groups except the one already present have at least 100/N percent shares).[2] When the composition of the inflow precisely matches the composition of the stock (regardless of what S is), DIV has a value of zero since there is no transformation of the stock by the flow.[3] These DIV values carry no context-free normative content, however, as will be amplified below.

Values of DIV provide an intuitive implication about the direction of adjustment that the population is undergoing as a result of recent flows. Assuming that the population experiences no outflow or that the outflows represent an approximately random sample of the current population stock, a DIV>0 implies that the neighborhood will become more diverse over time were this DIV to continue; the opposite would be implied for DIV<0. When DIV=0, it means that the population will remain stable at its current composition (regardless of whether that is diverse or not), again assuming random outflows from the neighborhood. Note, however, that non-zero values of DIV cannot be interpreted as a measure of the *speed* at which a population will be transformed, since such would require information about the magnitude of both gross inflows and gross outflows of the population, as well as the stock.

Illustrations of DIV Calculations for Hypothetical Neighborhoods

To illustrate how DIV would be calculated in particular situations, consider five cases of hypothetical resident populations in a neighborhood where DIV is calculated for four mutually exclusive groups (N=4).

---

[2] In the case of $S_{it} = 100$, any inflows having $F_{jt} \leq 100/N$ for j=i and $F_{jt} \geq 100/N$ for ALL other j will yield DIV=100. We are grateful to Paul Jargowsky for pointing this out.
[3] As is true for virtually all indices, values of DIV are not unique; identical values can be produced by different profiles of S and F.

Case 1: Least diverse neighborhood being transformed by most diverse flow

This neighborhood would initially have $S_{1t}=100\%$ and all other $S_{it}= 0\%$. The most diverse inflow would have all $F_{it}=25\%$. Using the formulas above: $\Delta SF_{1t} =25$ and the other three $\Delta SF_{it} =25$, yielding $DIV_t=100$, the maximum value for the index.

Case 2: Most diverse neighborhood being transformed by least diverse flow

This neighborhood would initially have all $S_{it}= 25\%$. The least diverse inflow would have $F_{1t}=100\%$ and all other $F_{it}=0\%$. Using the formulas above: $\Delta SF_{1t} =-25$ and the other three $\Delta SF_{it} =-25$, yielding $DIV_t = -100$, its minimum value.

Case 3: Moderately diverse neighborhood being transformed by most diverse
flow

This neighborhood would have, e.g., $S_{1t}= S_{2t}=50\%$ and both other $S_{it}= 0\%$. The most diverse inflow would have all $F_{it}=25\%$. Using the formulas above: $\Delta SF_{1t} = \Delta SF_{2t} =8.333$ and the other two $\Delta SF_{it} =25$, yielding $DIV_t = 66.67$. In comparison to case 1, this shows how DIV will grow progressively smaller, even with the most diverse inflows, as all S_i get closer to 25% (or, more generally, 100/N).

Case 4: Any neighborhood not being transformed by flow identical to its stock

This neighborhood would have all $S_{it}= F_{it}$ for all i. Using the formulas above: all $\Delta SF_{it} =0$, yielding $DIV_t=0$.

<u>Case 5: Not diverse neighborhood being transformed by not diverse flows</u>

<u>tending to transform one type of homogeneity into another.</u>

This is a case where the inflows are mirror opposites of the current stock, such that the neighborhood will eventually switch from one sort of not diverse place to a different type of not diverse place, were this in-flow to persist indefinitely. As illustration, consider a neighborhood with S_{1t}=80%, S_{2t}=20% and all other S_{it}=0%. Imagine an inflow that has the opposite composition of the stock: F_{1t}=20%, F_{2t}=80% and all other F_{it}=0%. Using the formulas above: ΔSF_{1t} =18.33, ΔSF_{2t} =5, and the other two ΔSF_{it} =0, yielding DIV_t = 23.33. This shows how DIV "penalizes" inflows that "overshoot" the long-run maximum diversity flows (25%), even though for a period (until S adjusts sufficiently) they may tend to improve diversity (as suggested by the positive value for the DIV).

An Illustration of the Time Path of DIV in a Hypothetical Gentrifying Neighborhood

Many community advocates express concern that the in-migration of higher-income groups into a formerly low-income neighborhood will lead to market processes that will lead to the involuntary displacement of many of the original, low-income residents. While not necessarily opposing the initial in-migration of those with more disposable income into a previously disinvested, disadvantaged neighborhood, they often would prefer to have the population stabilize with some substantial mix of income groups. The time path of DIV potentially provides a clear indicator of both possibilities.

Consider a homogeneously lower-income neighborhood that only has lower-income households moving in; DIV would register approximately zero. But, as the leading edge of the gentrifying group began to flow in, the DIV would begin to rise as the inflow became more diverse; it would reach its highest value when (and if) the gentrifier group flowing in reached its percentage share of 100/N. However, if gentrifiers become

the dominant share of in-mover, DIV will begin to fall as the inflow registers as less diverse. If this inflow of gentrifiers persists over a sufficient period, the stock itself will be transformed (probably through displacement processes) to one with gentrifiers constituting the dominant share, whereupon a continued inflow of predominantly gentrifiers will produce a negative value for DIV. As the succession of gentrifiers and the displacement of the original residents is nearly completed, DIV will again approach zero, as the (gentrifier-dominant) composition of stock and inflow aligns. Thus, the time path most feared by opponents to displacement is one where DIV approximates a sine wave, such a portrayed in Figure 1 by line A.

[Figure 1 about here]

By contrast, imagine that an intervention of some sort occurs in this hypothetical gentrifying neighborhood that either (1) locks in a substantial share of the low-cost housing stock for perpetual use by lower-income households, and/or (2) thwarts the dominance of gentrifiers in the inflowing population once a substantial mix in the population stock has occurred. In such a scenario the DIV would not fall into the negative range, instead reaching zero when the neighborhood population stock still possessed a substantial income diversity. This alternative scenario is portrayed in Figure 1 by line B.

The foregoing implicitly shows how the evaluative interpretation of DIV is highly contextualized. Though the interpretation of a DIV value is unambiguous in all contexts, its normative content is not. If one's goal is stable, diverse neighborhoods, for example, a DIV=0 in a diverse neighborhood is desired, but a similar value in a homogeneous neighborhood is not.

An Illustration of DIV in Selected Chicago Neighborhoods

Up to this point, our illustrations have been based on hypothetical situations. Does DIV yield anything of interest using real data? Our prototype investigation using selected Chicago community areas suggests an unequivocal "yes."

We computed DIV separately for both four racial groups (Hispanic, non-Hispanic white, non-Hispanic black, non-Hispanic other) and three income groups (high, middle, and low)[4]. For estimates of the annual population stocks we used 1990 and 2000 census statistics, estimated statistics for 2004 from Easy Analytic Software, Inc. (EASI), and linear interpolation for values in intervening years. For estimates of the annual inflows we used Home Mortgage Disclosure Act (HMDA) data on the characteristics of originated mortgage loans to owner-occupants for 1992-2006. We recognize that HMDA data are unlikely to serve as accurate proxies for the composition of inflows of renters, but we were unaware of any sources of annual data on such inflows. As such, the following DIV estimates should more accurately be viewed as a comparison of the composition of new homebuyers and the existing population. In calculating DIV we used the estimated stock for year t and the flow during the year t+1.

Our first illustration, Woodlawn, represents an archetypical disadvantaged, heavily black-occupied neighborhood with nascent recent signs of revitalization (Moore, 1973). In 1991, two-thirds of its residents were categorized as low-income, only 7% were high-income and 95% were black; the inflows almost matched the income and racial composition of the stock, producing DIV near zero. But throughout the period the flows steadily changed to diversify both dimensions of Woodlawn. In Figures 2 and 3 we

---

[4] The three income groups are specified relative to the Chicago metro Area Median Income (AMI). The "high" category refers to income levels greater than 120% of AMI, "middle" refers to income levels that fall between 80% and 120% of AMI, and "low" refers to income levels below 80% of AMI.

show the shares of the individual racial and income groups that comprise the annual flows into Woodlawn, so the reader can gain a more intuitive understanding of the components contributing to DIV. By 2004, the low-income share of the neighborhood population had fallen 3 percentage points and the high-income share had risen an equal amount; the percentage of black residents had fallen a percentage point. Indeed, by 2006 the inflows were notably more diverse than the stocks, with 42 percent high income and 26 percent non-black homebuyers. As befits this situation of growing income and racial diversity, the DIVs for both income and race evince higher values over time; see Figures 2 and 3.

[Figures 2 and 3 here]

Uptown represents a very different sort of neighborhood: notably diverse racially and economically (Nyden, Maly, and Lukehart, 1997). In 1991, 39 percent of the population was white, 23 percent was black, and 22 percent was Hispanic. Fifty-six percent were low-income, 27 percent were middle-income, and 17 percent were high income. During the ensuing decade, the inflows worked to reduce racial diversity, however, as white homebuyers predominated (reaching 87 percent of the inflow in 2006). Thus, the racial DIV for Uptown registered negative values, as shown in Figure 4. By contrast, the neighborhood became even more diverse economically, as the inflows were disproportionately high-income (in 2006 the homebuyers were 48 percent high-income and only 6 percent low-income), which yielded a positive income DIV; see Figure 4.

Our final illustration is Lincoln Park, a privileged neighborhood where, in 1991, 85 percent of the population was white and 65 percent were high-income. In the early 1990s the inflows had lower proportions of high-income homebuyers but even higher

proportions of whites than the current stock, producing positive DIVs for income but negative DIVs for race; see Figure 5. Though this racial pattern of modestly reducing diversity persisted during the period, the income inflows became mirror images of the stock, dropping the income DIV to zero. Indeed, high-income groups comprised 75 percent of the inflow and 77 percent of the stock by 2006.

[Figures 4 and 5 here]

Although the foregoing cases suggest very different dynamics and corresponding variations in DIV patterns over time and space, we emphasize that these three cases represent only a small fraction of the wide variety of DIV patterns that can be observed in Chicago and, presumably, in other cities as well. This suggests to us that DIV will be a useful tool for potentially clustering neighborhoods by dynamic typologies and revealing heretofore obscured insights.

Discussion, Caveats, and Extensions

Unlike most social indicators, there is no consensual normative content to DIV, nor is its normative interpretation static over time. For example, a conventional indicator of neighborhood crime rates is universally seen as a measure of a negative attribute, and a higher reading at any moment in time would thus be undesirable. By contrast, higher values of DIV may be interpreted as either good or bad, depending on the goals of the observer and the particular dynamic neighborhood context. As illustrated by Figure 1, the potential normative implications of DIV are best revealed by an examination of inter-temporal changes in DIV values (coupled with initial neighborhood composition), not a static one.

Furthermore, we stress that DIV makes no implicit judgment about the feasibility of achieving stability or diversity. We recognize that achieving maximum values for DIV in all neighborhoods or non-spatial sets simultaneously is likely to be infeasible because of limits to the diversity of the overall population pool in question (e.g., the entire metropolitan area). In this vein, one might reasonably consider modifying the specification of DIV so that "maximum diversity" of the inflow is measured relative to the diversity of the larger population pool from which the flow is drawn, instead of the absolute standard of 100/N for each group.

We note several limitations to DIV. First, it is insensitive to the magnitude of the inflow relative to the size of the stock. It thus does not give a clear sense of the rate at which the inflow is modifying the composition of the stock. We have been unable to devise a weighting scheme that does not introduce its own set of confounding and potentially misleading elements into the index.[5] A second limitation is that DIV can change its value discontinuously when one or more of the $S_{it}>$ change from being slightly greater than to less than 100/N or vice versa. A third potential pragmatic limitation is than in some current applications the analyst will have no annually updated measures of stocks for residential population from the U.S. Census Bureau. This limitation forced us to use interpolation in our foregoing example of DIVs for Chicago. Fortunately, this limitation should be lifted with the emergence of annually updated, five-year moving average population estimates for census tracts provided by the American Community Survey beginning in 2010.

We close with two potential extensions of the DIV index. First, DIV is easily adaptable should an analyst wish to employ a different standard for maximum diversity.

---

[5] Moreover, to provide any realistic estimate of the degree to which inflows change neighborhood population composition, one would need to have information about inflows of not only homebuyers but also in-moving renters, which is infeasible with current data sources. This constraint may be less binding in other applications.

For example, in the case of racial diversity one could substitute the percentage of each racial group i in the encompassing population pool (like metropolitan area) for 100/N in the DIV formula. Second, it should be apparent that DIV is symmetric for measuring either inflow *or outflow* composition compared to stock composition (though in the case of outflows all signs would need to be reversed to preserve the same intuitive interpretations of DIV). We hope that researchers in the future can find a variety of additional creative uses for DIV and that the result will be valuable insights into neighborhood dynamic processes.

References

Moore, W. (1973). "Woodlawn: The Zone of Destruction." *The Public Interest* 30: 41-59.

Nyden, P., Maly, M. and Lukehart, J. (1997). "The Emergence of Stable Racially and Ethnically Diverse Communities," *Housing Policy Debate* 8(2): 1051-1482.

Reardon, S. F., and Firebaugh, G. (2002). "Measures of Multigroup Segregation." *Sociological Methodology*, 32(1), 33-67.

Reardon, S. F., Firebaugh, G., O'Sullivan, D., & Matthews, S. (2006, August). *A new approach to measuring socio-spatial economic segregation*. Paper presented at 29th general conference of the International Association for Research in Income and Wealth, Joensuu, Finland.

Figure 1

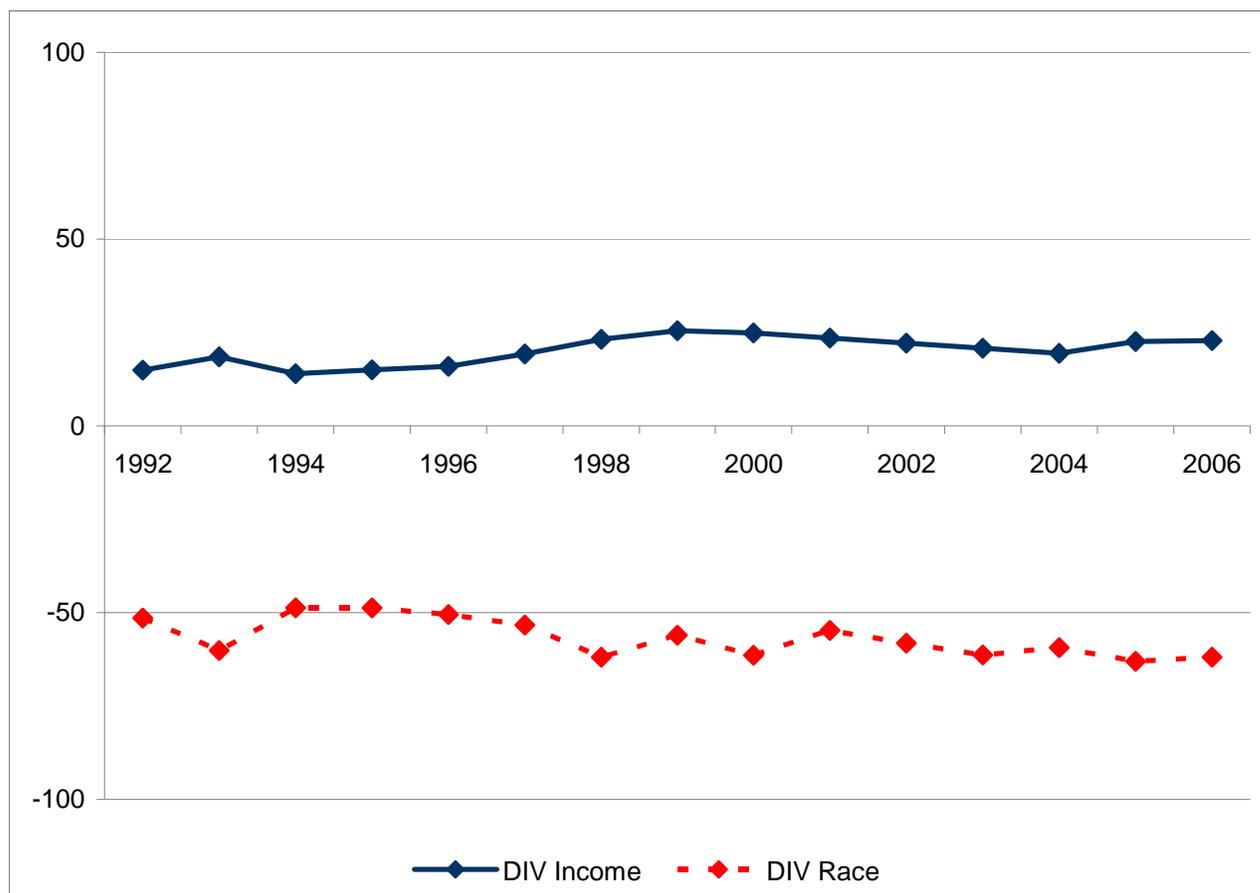**Alternative Hypothetical Time Trends of DIV in a Gentrifying Neighborhood**
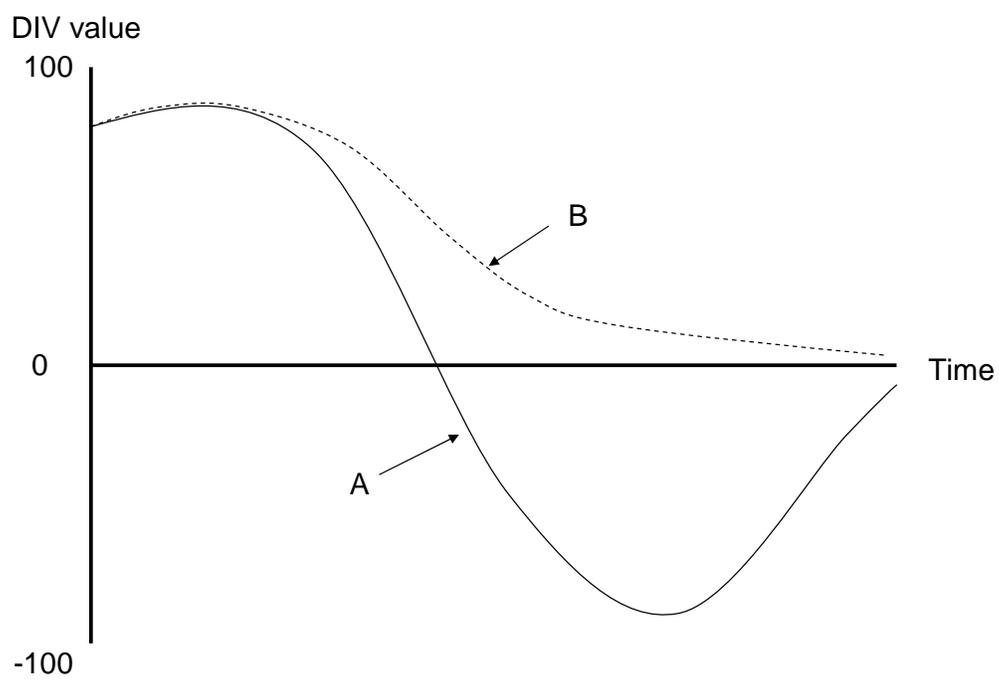
Figure 2
Racial Group Stocks (2000), Flows, and DIV for Woodlawn Neighborhood,
1992-2006
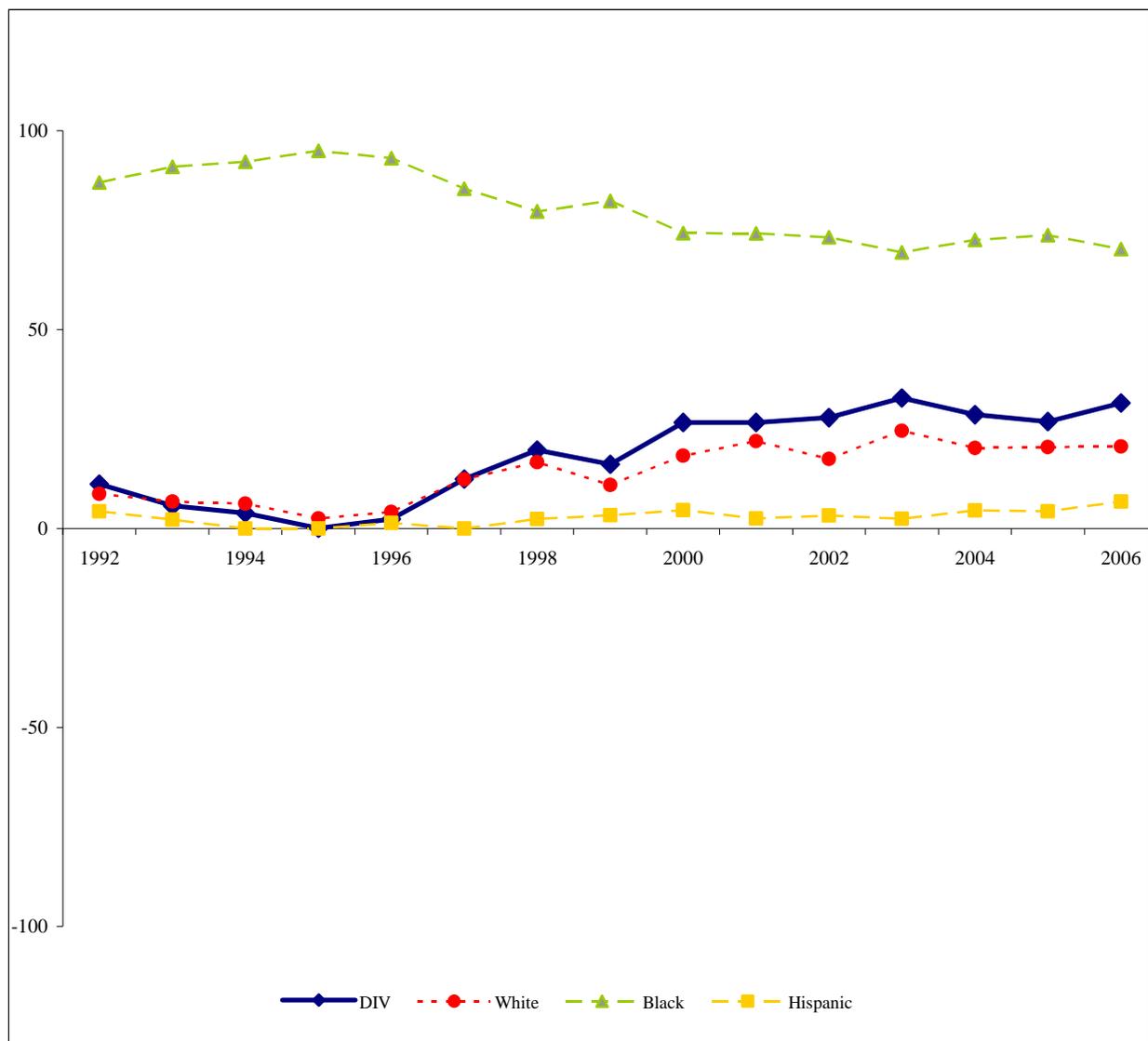
Figure 3
Income Group Stocks (2000), Flows and DIV for Woodlawn Neighborhood,
1992-2006
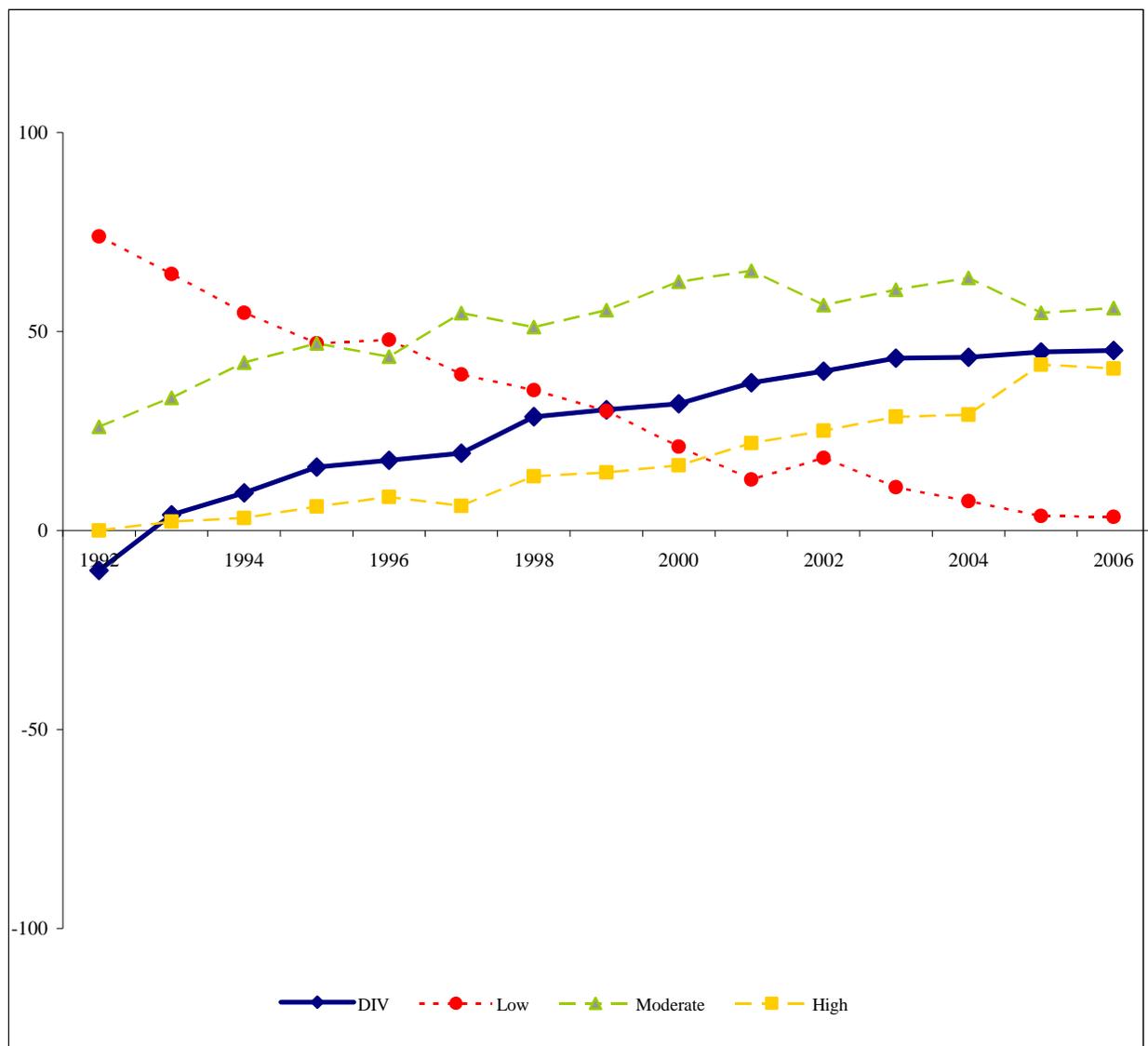
Figure 4
Racial and Income DIVs for Uptown Neighborhood, 1992-2006

Figure 5
Racial and Income DIVs for Lincoln Park Neighborhood, 1992-2006