A Portrait of the Artist as a Young, Middle-Aged, and Elderly Man[*]

Geoffrey R. Lanyon[†]
Board of Governors
Federal Reserve System
Washington, D.C.

Lones Smith[‡]
Economics Department
University of Michigan
Ann Arbor, MI 48109-1220

September 16, 1999

Abstract

We explore the age-value and age-quantity productivity profiles of fifty-three great Western artists whose work has been auctioned in the last decade. In terms of the average value of their paintings, we find that artists have three distinct phases to their careers: a steep incline, more than doubling their expected value until an age 31 peak; then follows a slower decline until age 47, losing almost half their peak value, and finally a very slow decline. The top quartile of artists distinguish themselves not only by a higher average value, but also in midcareer by losing only a fifth of their peak value in 31–47.

We then compare these value results with the annual quantity age-profiles: Output rises by more than 125% steeply until age 32, and then falls very slowly for the rest of their career. This decline is significantly steeper for the lowest expected value quartile of artists. Finally, we show that artists paint significantly more the higher is their average value.

Our analysis uses both polynomial fits, as well as multiple structure change analysis to deduce the best fits by piecewise linear splines.

[*]An earlier version of this appeared as the first author's 1998 Bachelor's thesis at MIT, supervised by the second author (beginning as a UROP project in 1996). We thank Steve Bergantino, Mark Hooker, Ruth Judson, Athanasios Orphanides, Dan Rodriguez, Brian Sack, Shinichi Sakata, Dave Small, and Eric Swanson for helpful conversations. Jushan Bai and Pierre Perron answered some email queries. Of course, any errors are our own. Smith acknowledges financial support from the National Science Foundation (grant SBR-9711885) for this research.

[†]email: `glanyon@frb.gov`; phone: 202-452-3139

[‡]email: `lones@umich.edu`; home page: `www.umich.edu/~lones`; ph: 734-764-2357; fax: 734-764-army

1 INTRODUCTION

Understanding the nature of worker age productivity profiles is an important problem. For instance, it would help us understand how the overall workforce productivity ought to shift as Baby Boomers age. Many papers over the last decade have studied age profile in various sectors of the economy. (See Kotlikoff and Gokhale (1992) and Tauer (1995).[1]) Very broadly, the focus of this paper is the growing sector of purely creative work. In particular, we study the lifetime age-productivity profile of homo economicus artistic.

While the quantity of output is a natural productivity measure in traditional jobs, an obvious focus of artist productivity is the average value of paintings.[2] Despite a life and career cut short, we feel that Van Gogh should surely be deemed a productive artist, simply because he created art valued by some in the tens of millions of dollars.

We do not consider the two dimensions of artistic productivity as a given, but subject to economic forces. Artists make calculated trade-offs in opting whether to spend four years painting a masterpiece on a chapel cathedral ceiling (Michelangelo, ages 33–37), or one year rifling off at least 333 paintings and sketches (Picasso, age 52). We thus explore both the age-value and age-quantity profiles of fifty-three noted Western artists whose work has been auctioned in the last decade, and for which we have sufficient data. Our main novel finding is comparing the age-value and age-quantity profiles, that value peaks sharply, but output almost plateaus. Our second new finding is a clear distinction between the profiles of the best and mean artists, and the observation that better artists paint more.

In terms of the expected value of their paintings, we find that artists have three phases to their lives: first a steep incline at about 6.4% per annum, peaking at age 31, during which time their work more than doubles in value. Then follows an almost as steep 4.1% annual decline until age 47, losing almost half of their peak value. Their careers end with a weakly significant yearly decline of 0.7%. The top quartile of artists distinguish themselves not only by a higher age profile, but also in midcareer by losing their peak value much more slowly, at only 1.3% annually, and shedding just a fifth of their peak value by age 47.

We next explore the age-quantity profile. We find that expected annual output rises

[1] Tauer found that that farmer productivity increases until middle age, and then declines. Studying a Fortune 1000 firm, Kotlikoff and Gokhale found that productivity falls with age.

[2] As economists and not professional art connoisseurs, our only measure of artistic 'value' will be its monetary value in auction. Galenson (1997) suggests that the economic and aesthetic measures coincide.

steeply at 5.5% of mean (lifetime) output until just after the quantity age-peak at age 32. All told, an artist's annual output rises over 125% as he ages from 20 to his peak at 32. It then very slowly falls at only 0.4% of mean output, with no further break. For the bottom caliber quartile of artists, this decline is much steeper at 1.4%. But the lesson of the data is clear, that value and quantity both rise sharply until an early thirties peak, after which value eventually falls below the age 20 level, while quantity falls very little from peak.

Controlling for the artist in our age-value regressions provides a simple measure of artistic *caliber.* One might imagine that higher caliber artists excelled by sacrificing output. In fact, we find that they instead paint more thoughout their lives, with a 10% increase in average value being associated with a 3% increase in lifetime output.

There has been some work in scientific endeavors, even economics (cf. Hamermesh and Oster (1998)). The literature here generally suggests a concave downward productivity profile, with a peak as early as age thirty. A possible critique of this work is that any measure of journal publications or citations can be misleading: A paper's true impact may only be felt decades later, when it is no longer cited, having entered the canon of knowledge. In contrast, artistic work may be individually prized centuries after their creation.

The artistic age-value profile has been investigated before. Studying the returns to art, Agnello (1994) assumed a quadratic value-profile, and found a peak around age 42. Closely related to our paper is Galenson (1997), who is interested in the secular change in age-value profiles. He separately estimates the best polynomial profiles for forty-two different contemporary artists, and finds that age is significant for many, but not all artists.

Our analysis uses both polynomial fits, as well as multiple structure change analysis to deduce the best fit by piecewise linear splines. In particular, the lowest degree fit polynomial is a quartic, in contrast to Galenson, who restricts himself to at most cubic polynomial age-value profiles. Moreover, the best piecewise linear spline involves three distinct phases for artists. The multiple structure change models methodology we use has been developed recently by Bai and Perron. Their theory describes how many breaks exist, and where they should go, allowing one to identify differing phases in an artist's career.

The next section summarizes our data sources. Section 3 gives the age-value profile analysis, and section 4 the age-quantity profile analysis, as well as the link between lifetime output and artistic caliber. Some methodological appendices follow.

2 DATA SOURCES

Before plunging into details, we must clarify the nature of our data sets. They are drawn from the 1987–96 volumes of *Mayer International Auction Records on CD-Rom* — which is essentially a compilation of worldwide art auction sales. While *Mayer* contains roughly 80,000 artists and 800,000 observations, our panel datasets includes sales of paintings which span just fifty-three different Western painters from the 17th–20th centuries (see the appendicized Table 5). The Western restriction was a natural choice, while the number of artists was constrained by demanding a minimum of thirty sufficiently documented sales.[3]

Crucially, because different data are available for different paintings, there are really *three* relevant data sets for this paper, all a subset of *Mayer*. The same fifty-three artists figure in all three data sets, but the paintings included vary. The data set with the smallest number of paintings is employed in the age-value profile regression, which demands information about price, date, size, and artistic medium. The age-quantity profile regressions use a larger number of paintings, counting every unique *and dated* painting by an artist. But the largest number of paintings is used for the lifetime quantity against artistic caliber regression, since this merely uniquely counts every painting by an artist in *Mayer*.

3 THE AGE-VALUE PROFILE

3.1 The Age-Value Profile Data Set

Here, we restricted ourselves to *Mayer* paintings for which we knew the year sold, the artistic medium, the size, and of course the price. Prices are recorded in British pounds, one of the common currencies in Mayer, and ranged from £35 to £8,850,000 (Renoir's "La Tasse de Chocolat"). For medium, Mayer classifies art according to: oil, pastel, watercolor, gouache, drawing, print, mixed media on paper, or sculpture. Given the obvious difference between sculpture and other art media, we exclude sculptures from our data set.

In selecting the fifty-three artists in our data set, we originally looked for artists with

[3]Despite the CD rom format, the data essentially cannot be downloaded in spreadsheet format (only 200 data points at a time), preventing simpler computer handling of data sets. Data recovery was very time intensive.

at least 100 observations. This yielded just seventeen artists. Lowering the bar to 30, we only admitted artists who painted at either a young or old age, and thus for whom data at a wide age range was available. The vast majority in fact painted from their 20's into their 60's or 70's. Our last recorded painting for each artist was painted within five years of the death year for all but ten artists. Table 5 (appendicized) lists the earliest and latest "age of execution" for each artist.

In total, we have 4291 observations, where each observation is a specific auction result. The number of sales for each artist ranges from 31 for Jean Auguste Ingres to 137 for Laurence Stephen Lowry. The disparity in quantity of data obtained across artists owes simply to the fact that for some artists there were only 35 well-documented paintings, while for others, a wealth was available. In a few cases, thousands of auction results existed for a single artist, but in the interests of balance, we used at most 140 sales for each artist.

Not surprisingly, there are many cases with multiple observations for the same artist at the same age, while quite often there were no auction results for an artist at a given age. Mayer therefore offers us a panel data set that is extremely unbalanced. In choosing which specific auction results to include (for those artists with over 140), we randomly selected the sales, subject to including at least one observation at every age for that artist.

While many paintings in the data set are dated, many are not. When the date is not specified on the work but one is listed, we take our source's estimate of the date to be accurate. Finally, some sale dates listed in *Mayer* were demonstrably errant (eg. there were several postmortum paintings), and we have excluded them.

3.2 Age-Value Econometric Analysis

We regress the natural log of auction price p_i for painting i on age, as well as other variables which might influence the price fetched by a painting in an auction,[4] specifically:

- $\alpha_{j,i}$, a dummy variable for each artist j ($j = 1, \ldots, J$) of painting i to control for individual effects, such as innate ability, training, motivation, and the preferred art genres.[5] We shall refer to its beta coefficient (β_j below) as the artistic *caliber.*

[4]Being a time series, it is standard to use the logarithm. Moreover, we wish to interpret our beta coefficients to proportionate conclusions about price changes.

[5]Since artists usually remained in the same genre, we did not include a separate regressor for it.

• $\mu_{k,i}$, a dummy variable to control for medium ($k = 1, \ldots, K$) of painting i;

• σ_i, the natural log of the area (in cm^2) of painting i — i.e. doubling the width and height of the painting doubles the log size; Galenson (1997) used this compelling choice;

• $y_{s,i}$, a dummy variable to control the year sold of painting i, amongst $s = 1988$–96.[6] Since our auction results all occurred within a short nine year window, we do not further adjust prices by a consumer price index.

In modeling the age, we explore two different approaches. The polynomial has been previously used to describe age-value profiles (eg. Galenson (1997) and Agnello (1994)); however, other structural forms of age have yet to be examined carefully, and we would really like to identify distinct periods of an artist's career, if any. Thus, we determine the outcomes from a standard piecewise linear spline structural break models.[7]

The basic regression equation is therefore

$$\log(p_{i,j}) = \alpha_0 + A(age_{i,j}) + \sum_{j=1}^{J-1} \beta_j \alpha_{j,i} + \sum_{k=1}^{K-1} \gamma_k \mu_{k,i} + \sum_{s=1}^{S-1} \delta_s y_{s,i} + \eta \sigma_i + \epsilon_i \qquad (1)$$

where $A(age)$ is a polynomial, or piecewise linear spline of the age, denoted age_i of the artist j of painting i when it was executed. Specifically,

• for a polynomial representation, $A(age) = \theta_1 \cdot age + \theta_2 \cdot age^2 + \cdots + \theta_n \cdot age^n$

• for a piecewise linear function with G age cohorts, there are age breaks $b_1, \ldots, b_{G-1}$, and slopes $m_1, \ldots, m_G$, with $A(age)$ the continuous function: $A(age) = m_1 \cdot age$ if $age < b_1$, $\ldots$, $A(age) = A(b_g) + m_{g+1} \cdot (age - b_g)$ if $age \in [b_g, b_{g+1}]$, $\ldots$ and finally, $A(age) = A(b_G)$ $+ m_G \cdot (age - b_{G-1})$ if $age \geq b_{G-1}$.

The number and location of the breaks in the piecewise linear spline forms of age are determined by adapting the technique outlined in Bai and Perron (1998a) and (1998b) (BP1 and BP2). This technique was chosen because it lets the data entirely determine the breaks; it is also relatively easy to implement. Here is a brief summary of this procedure: one starts by choosing a given number of breaks ℓ (say, $\ell = 0$). Then, the optimal

[6] As a side implication, we derive estimates of the return to art for this time span, just after the period in Goetzmann (1993). Consistent with the artist snapshots in Galenson (1997), and the results in Ashenfelter and Graddy (1999), we find a negative nominal return to art 1988–96. Prices for our cohort of artists peaked in 1989, and since have fallen substantially.

[7] Since it is not even obvious that the effect of age on price ought to be continuous, we also pursued a step function form of the age profile, and it was a good close staircase fit to the piecewise linear spline.

$\ell + 1st$ break is determined, so as to minimize the R^2, and thus maximize the explanatory power of the regression. A form of the F-test is then used to determine whether this break was justified. If so, the procedure repeats itself inductively, until stopping. This determines how many breaks can by sequentially justified, by Proposition 8 BP2. Finally, we recompute the optimum location of these breaks using a global maximizer of the R^2. See Appendix A.1 for more detail.

Unfortunately, problems with our dataset render it unsuitable for OLS. First, there is heteroskedasticity: Some artists simply command a wider price range than others. A Glejser's test revealed heteroskedasticity due to the age variable, the size variable, the artist dummies, and the medium dummies, but not the year sold dummies.

Also, it is natural to assume a form of reputational persistence will be reflected in prices. For instance, phases of a great artist's life are often grouped as a unit, like Picasso's 'Blue' or 'Rose' periods, or Renoir's more classical 'Bathers' phase. One suspects that work from the same era might be similarly valued, and thus autocorrelation is possibly an issue. By this, we mean sequential correlation, not necessarily corresponding to real time. For within our data set, the timespan between consecutive execution dates of an artist is not constant.

To determining the presence and extent of autocorrelation, we inductively regressed the error term on the error term of observations for an artist one, two,... all the way to twenty-five paintings back. We found the first, second, third, and fourth lags all to be significant to 99%, and thereafter, the lags were generally quite insignificant.

In addressing the problems discussed above, we implement the Generalized Estimating Equations methodology of Liang and Zeger (1986). This technique, when combined with a robust estimate of the covariance matrix, yields consistent estimates for panel data cursed with both heteroskedasticity and autocorrelation. This is critical, as without consistency, we cannot employ the methodology of Bai and Perron. See Appendix A.2 for details.

To find the best fit polynomial, we sequentially ran linear regressions, then quadratic, cubic, ... until the t-statistic of highest order polynomial coefficient was found to be insignificant. This strategy yielded a quartic best fit. For our best piecewise linear spline, there are just three age cohorts of 0–31, 31–47, and 47+. Table 1 summarizes the results.

Figure 1 contains the graphical depictions. Appendix B describes how we created the graphs in this paper. The two snapshots of the age-value profile essentially tell the same

Table 1: **Effect of Age on Price.** The table summarizes the best polynomial, and piecewise linear functions, as well as the beta coefficients and their z-statistics.

Polynomial			*Piecewise Linear Spline*		
age^n	Coefficient θ_n	(z-statistic)	Age Cohort g	Slope m_g	(z-statistic)
age	0.482	(4.10)	≤ 31	0.0642	(3.92)
age^2	-0.015	(-4.18)	31–47	-0.0412	(-6.38)
age^3	0.000178	(-4.11)	≥ 47	-0.00705	(-2.01)
age^4	-7.66×10^{-7}	(-4.00)			

story: An artist's expected value crescendos upward at 6.4% per annum until age 31. His average value then falls with time, at the fast rate of 4.1% until age 47, at which point this trend slows significantly to 0.7% thereafter. The inflection point required by this slowdown explains why a linear or quadratic polynomial representation of the profile could not possibly suffice. The quartic is intuitively needed because the downward sloping final portion of the career demands an additional inflection point.

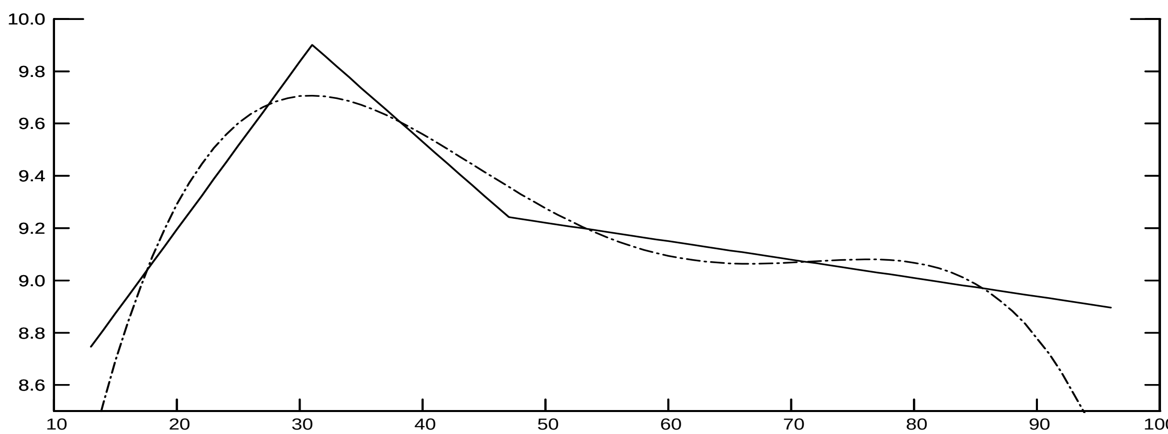

Figure 1: **Two Representations of the Age-Value Profile.** Depicted are the best fit polynomial — a quartic — and piecewise linear spline, having three linear segments. On the vertical axis, is the expected log price (in pounds) for a painting by an artist of the given age, unconditional on all other attributes. The horizontal axis is age.

Since the log graph may understate the magnitude of the age profile changes, consider an average twenty year old artist. In just eleven years as he ages to his value peak at thirty one, he gains 102.5% in expected artistic value! By age 47, he has lost 48.3% of his peak value. But his fall does not push him below his age 20 expected value until age 55.

3.3 Age-Value Profiles and Artistic Caliber

We now add an interactive variable to distinguish the artists with the thirteen highest and lowest calibers β_j (i.e. top and bottom quartiles). These are the most and least successful artists in terms of average value, controlling for our previous list of observables. We have found that the deviations from average are insignificant for the lower quartile, and so we omit them from the discussion below. We therefore focus just on the *high caliber* artists.

To ensure a more transparent comparison of the age profiles, we dwell entirely on the piecewise linear spline depiction. For just as in Figure 1, the omitted polynomial tells a similar story — but the curves are harder to compare. So let $A(age)$ be the piecewise spline form described on page 5. Next, let $A_{HC}(age)$ equal a dummy *times* the same piecewise linear spline form as $A(age)$. The dummy in question equals one if the artist is considered high caliber, and zero otherwise. Then the regression equation becomes:

$$\log(p_i) = \alpha_0 + A(age) + A_{HC}(age) + \sum_{j=1}^{J-1} \beta_j \alpha_j + \sum_{K=1}^{k-1} \gamma_k \mu_{k,i} + \sum_{s=1}^{S-1} \delta_s y_{s,i} + \eta \sigma_i + \epsilon_i \qquad (2)$$

Determining structural breaks for two variables is much more problematic than for one, due to a degree of freedom in choosing the order of breaks for A and A_{HC}. In fact, we have found that the breaks and resulting profiles differ substantially depending on the choice. We simultaneously avoid such problems, and achieve a transparent comparison, by asking a simpler question: How do the age-value profiles of high- and average-caliber artists compare, *assuming the same age breaks*? Thus, $A_{HC}(age)$ has possibly different cohort slopes m_g^{HC}, but the same breaks $\hat{b}_1, \ldots, \hat{b}_G$, as estimated already.

Results are summarized in Table 2, and Figure 2 gives the graphical representation. The top quartile artist's age-value profiles differ insignificantly both early and late in life, but taper off much less in the middle age cohort 31–47. The expected value of the high caliber artists falls 3.6% more slowly than the average artist, i.e. at about 0.5% annually rather than 4.1%. Compounded, this means that the rich artist loses 19.3% of his peak value by age 47, as compared to the average loss of 48.3%. In other words, the higher caliber artist do not merely have an upward-shifted age-value profile; rather, they can lose their peak value significantly more slowly. His decline is so much slower that in contrast to the average artist, he never again has an expected value as low as his age 20 level.

Table 2: **Effect of Age on Price, Distinguishing the Top Quartile of Artists.** The table summarizes the best piecewise linear spline regression results, where there is an interactive variable, capturing profile effects specific to the top quartile of artists. The middle cohort rich interactive variable 0.0365 for the high caliber artists is significant to 97%.

	Piecewise Linear Spline (Top Quartile)			
Age Cohort g	Slope m_g	(z-statistic)	Interactive Slope m_g^{HC}	(z-statistic)
≤ 30	0.060	(2.95)	0.0169	(0.37)
31–47	-0.0499	(-7.02)	0.0365	(2.21)
≥ 47	-0.00589	(-1.48)	-0.00386	(-0.46)

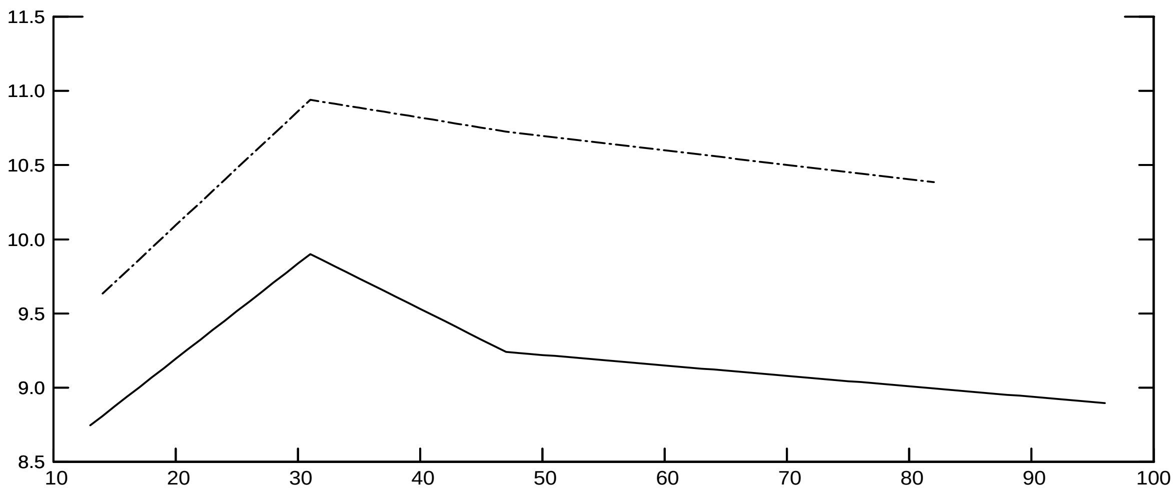

Figure 2: **Piecewise Linear Spline Age-Value Profile for Top Quartile of Artists.** On the vertical axis is the expected log price in pounds; the horizontal axis is age. The top graph is the age-value profile spline for the upper quartile of artists, and the bottom graph is the same profile of the average artist, i.e. the same as in Figure 1. Also, the spline for the best artists ends at age 82 because we do not have data for those artists past that age.

4 THE AGE-QUANTITY PROFILE

4.1 The Age-Quantity Profile Data Set

To capture the annual output of each artist, we eliminated duplicates,[8] and restricted ourselves to sales for which the sale date was known. For many of the works, multiple dates are listed — for instance, a work may be dated 1884–6. While we simply chose not to include these observations in data set A, we do include them here, using the middle year (rounding up when fractional). Since so many works were not dated, this yielded just 14283 total paintings, from 32 for Millais to 3828 for Picasso. Picasso had the most output in any year (333), while empty years were quite common for our artists.

For all of our quantity analysis, we do not pretend that we have an exhaustive list of the collected works of any artist. Surely many or even most have remained ensconced in museums and private collections during our decade span. Rather, our working assumption in using *Mayer*'s auction sales as a proxy is that no artist's output is systematically excluded any more than any other. As we only care about proportionate differences in output, we only require that there is some (unknown) fraction $\phi \in (0, 1)$ such that *Mayer*'s list is an unbiased estimate of ϕ times the actual output. Our group of artists all being more or less notable, and all Western, there is no a priori reason to suspect that art auction sales would under or over-represent their work. Still, this remains the key potential critique of our findings — and for that matter, of any study that touches on artistic output. For we could find no data set confidently listing all output by each of our fifty-three artists, and from conversations with sources, we are reasonably confident that none exists.

Since not all of the works in the *Mayer* auction catalog are dated, our analysis also assumes that for any particular artist, the fraction of dated works painted at a given age is the same as the fraction of undated works painted at that age. Section 4.4 provides one internal test of the coherence of the dated and undated data sets.

[8] *Mayer* does not flag works represented in multiple sales. We thus used not only the title and date, but also size, medium, dimensions, and other descriptors, to identify such duplicates. Indeed, several paintings with the same name and artist were painted in the same year. In just a few cases, this was problematic, since different auctions may give slightly different dimensions for the same work. In such cases, two paintings by the same artist were deemed a multiple sale if they had the same title, date, medium, whether signed status, and the dimensions differed by at most $0.5\,cm$. Fortunately, there existed relative few multiple sales for each artist, and so this should hardly affect our findings.

4.2 Age-Quantity Econometric Analysis

We very much wish to represent quantity just as we have done with the price, so that coefficients can be interpreted in proportionate terms. For as noted in section 2, our data set clearly does not include anywhere near the entire output of the artists studied, so that the absolute numbers per se are not relevant. But using log-quantity is simply not an option (as we did with the price), because an annual output level of zero accounts for 27.5% of all quantity observations. While we could add a small ε-increment to all quantity levels, we have verified that the age profiles are somewhat sensitive to the selected ε — obviously because of the large number of zero output years. We instead use *normalized quantity* — that is, annual output divided by the artist's lifetime average output.

We again consider both polynomial and piecewise linear splines for our age-normalized quantity profiles. An 'observation' $q_{j,t}$ now aggregates all auction sales of paintings in year t of artist j's life, provided he is still painting — namely, if t lies between the first and last years with positive output in our data set. Altogether, there are 2918 such observations in the annual quantity data set. Obviously, controlling for the year sold, size, and medium of amalgamated paintings is not possible, and thus we run the following simple regression:

$$(q_{j,t}/\bar{q}_j) \quad = \quad \kappa_0 + AGE(t) + \sum_{j=1}^{J-1} \kappa_j \alpha_j + \epsilon_{j,t} \tag{3}$$

where $q_{j,t}$ is quantity produced by artist j at age t, and $\bar{q}_j$ is his mean quantity produced in a given year; $AGE(t)$ is a polynomial or piecewise linear spline function of age t, as usual. The coefficient κ_j is captures any artist-specific output effects.

The best fit polynomial and piecewise linear spline age-quantity profiles are found just as in section 3. While we find that autocorrelation is not a problem with this data, heteroskedasticity still is, and thus we thus employ a least squares regression with a robust estimate of the covariance matrix. We find that the lowest order polynomial best fit is cubic, while the best piecewise linear spline had just with one break just at age 32. Note that this peak quantity year is just one year delayed from the peak value year found. Until age 32, average artistic output rises at about 5.5% of his lifetime average output per year. Thereafter, it falls more slowly at 0.4% per annum. Again, the compounded results are more dramatic: a typical artist's mean annual quantity rises 127% from age 20 until his

Table 3: **The Effect of Age on Annual Output.** The table summarizes the best polynomial, and piecewise linear functions, as well as the beta coefficients and their t-statistics.

Polynomial			*Piecewise Linear Spline*		
Age^n	Coefficient θ_n	t-statistic	Age Cohort	Slope m_g	t-statistic
age	0.117	(5.139)	≤ 32	0.0545	(8.273)
age^2	-0.00196	(-4.104)	≥ 32	-0.00384	(-2.16)
age^3	0.0000102	(3.279)			

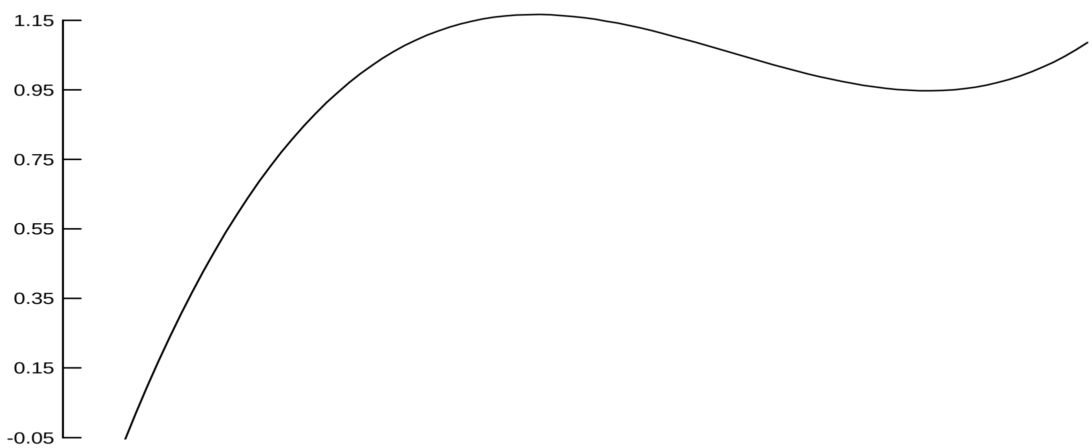

Table 4: **Age-Quantity Profile, Distinguishing the Bottom Quartile of Artists.** The table summarizes the results for the best piecewise linear spline age-quantity profile regression, where there is an interactive variable capturing profile effects specific to the bottom quartile of artists.

	Piecewise Linear Spline (Bottom Quartile)			
Age Cohort g	Slope m_g	(t-statistic)	Interactive Slope m_g^{HC}	(t-statistic)
≤ 32	0.0574	(7.70)	-0.0192	(-1.15)
≥ 32	-0.00108	(-4.95)	-0.0104	(-2.90)

We employ the piecewise linear spline form for $A(age)$. Let $A_{LC}(age)$ be a piecewise linear spline with the same breaks as already computed for $A(age)$, times a dummy equal to one if the artist is low caliber, and zero otherwise. The regression is:

$$(q_{j,t}/\bar{q}_j) = \alpha_0 + AGE(t) + AGE_{LC}(t) + \sum_{j=1}^{J-1} \beta_j \alpha_{j,t} + \epsilon_{j,t} \tag{4}$$

Regression age-quantity profile results are summarized in Table 4 and Figure 4. The significant effect is that low caliber artists lose their peak annual quantity 1% faster (as a fraction of mean output) than average — i.e. at about 1.4% of mean output annually rather than 0.4%. The more striking compounded effect is that the low caliber artists return to their age 20 output levels by age 76, while at that age, the average artist is still producing double his age 20 level.

4.4 Annual Quantity versus Caliber

The motivation for including artistic caliber in the quantity regressions is that value and quantity are joint decisions. We conclude our results with a straightforward test of this link, asking whether higher caliber artists also produce more than lesser caliber ones.

A. Annual Quantity Analysis. A first approach is to use the data set from our piecewise linear spline age-quantity profile regression (§4.2). Of course, since we wish to get at the effect of caliber on quantity, we cannot employ a regression of *normalized* quantity, as we have done. Instead, we regress log quantity *plus* small ε on $AGE(age)$ (for the spline), and the previously estimated artistic caliber $\hat{\beta}_j$; that is, the caliber measures replace the

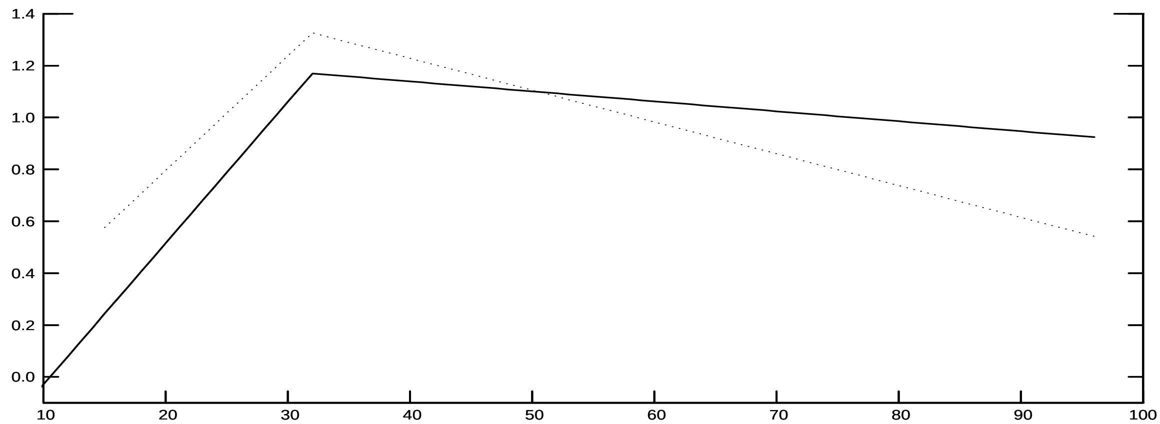

Figure 4: **Age-Quantity Profiles and Low Caliber Artists.** On the vertical axis is annual quantity divided by lifetime average quantity; age is on the horizontal axis. The dotted line is the age-quantity profile for the low caliber artists, and the solid line, for average artists.

artist dummies. While we rejected this before because the age profile was too sensitive to the chosen ε, we can use it to bracket our desired caliber effect.

$$\log(q_{j,t} + \varepsilon) = \alpha_0 + AGE(t) + \theta\hat{\beta}_j + \epsilon_{j,t} \tag{5}$$

The result of this regression yields $\hat{\theta} = 0.278$ when $\varepsilon = 0.1$ (with $t = 11.627$), and $\hat{\theta} = 0.346$ when $\varepsilon = 0.01$ (with $t = 9.63$). Thus, higher caliber artists paint more, with a 10% higher average value associated with an increase in lifetime output between 2.78% and 3.46%.

B. Analysis Using the Aggregate Quantity Dataset. Our annual quantity dataset excluded undated observations from *Mayer*. Obviously, if we are merely interested in lifetime output, we should include all observations, excluding merely the multiply counted paintings. After carefully pruning duplicate sales by comparing the particulars of each painting sold for each artist, this yielded 22860 paintings — from 53 for Millais to 4418 for Picasso. We then ran the following simple OLS regression on the fifty three data points (i.e. one for each artist)

$$\log Q_j = \theta_0 + \theta\hat{\beta}_j + \epsilon_j \tag{6}$$

where $\hat{\beta}_j$ is the caliber of artist j estimated in section 3, and Q_j is his lifetime output.

This time, we find the coefficient $\hat{\theta} = 0.299$ with a t-statistic of 3.51. In other words, a 10% increase in artistic caliber is associated with a slightly lower yet still very significant estimate of 2.99% increase in lifetime output. We remark that this lies within the previous estimated range of $[2.78\%, 3.46\%]$, based on the smaller annual quantity data set. We thus conclude that higher caliber painters do not achieve their greater average value by sacrificing output. The optimally selected quantity and quality are complementary goods.

5 CONCLUSION

A. Summary. This paper carefully analyzes the age-productivity profile of a classic creative profession: artists. We identify the two most compelling elements of this profile — output value and quantity — and show that both rapidly rise to peaks at age 31 and 32, respectively. The fall from peak value substantially slows after age 47, while the decline in annual quantity is both without significant breaks after the peak, and also extremely slow. To be sure, artists forget how to paint as well as they did in their prime, but hardly lose a brushstroke in their flow output of paintings.

Taking the value profile as a focal measure of artistic *caliber,* we have also explored how these profiles differ as artistic caliber changes. We find that until age 47, the per painting value of the best caliber artists suffers less than the average artist. Likewise, the annual output of the worst caliber artists suffers more than the average artist during the post peak decline. Thus, upper caliber artists are distinguished by their ability to sustain value of their output, and the lower caliber artists by their inability to sustain their flow output. We concluded with the discovery that the best artists by caliber paint the most.

What we have ventured about inferring quantity from *Mayer* records — representative sampling — is clearly our most loaded assumption. While complete collected works listings are likely unavailable, we hope that future research here using larger data sets that also include works in the museums of the world will reconfirm our age-quantity profile findings.

B. Other Effects. We have shied away from discussions of tenure effects. One might imagine that since not all artists begin painting at the same time, perhaps those who start earlier are better at the same age than the late beginners. While there are obvious econometric multicollinearity problems in answering this, one can simply separately run our regressions for the late starters. The results from this were not significant.

C. Towards a Theory. In a separate work, we sought to sought to explain our results by a coherent maximizing story of artistic decision-making. It is obvious that there are two elements to any such story. First, learning-by-doing can provide the upswing to the age-value profile, as the artist becomes more proficient at his creative activity. To understand the value downswing, one really must focus on what is unique to the creative process. One can imagine that an artist is either endowed with a stock of ideas, or has a flow that is age-biased towards youth. While overcoming some technical hurdles associated with this nonsteady-state analysis (namely, two dimensional stocks), we were unable to find a robust set of preferences and production functions to rationalize the joint decisions for quantity and value. We consider this an intriguing open question in this field, as it offers a natural and important dynamic choice problem that underlies the creative process.

Appendices

A STRUCTURAL BREAKS PROCEDURE

A.1 Bai-Perron Technique

BP1 and BP2 develop a theory to determine the number and locations of structural breaks in a regression model. Equation (12) in BP1 or BP2 describes an F-test for $m = k$ breaks vs. $m = 0$ breaks, even when a robust variance covariance matrix is needed.[9] It is of the form:

$$\{F_T(\lambda_1, \ldots, \lambda_k; q) = \left(\frac{T-(k+1)q-p}{Tkq}\right)(\delta' R'(RV(\delta)R')^{-1}R\delta\} \tag{7}$$

where T is their sample size; p and q are the number of variables

(1986). See section A.2 for a more detailed discussion of this technique.

BP1 and BP2 also describe a sequentially-applied test $F_T(\ell+1|\ell)$ to test whether an additional break is present. This is simply the difference in the sum of squared residuals (SSR) of the ℓ break model, minus the *minimum possible* SSR of the $\ell+1$ break model (over all possible locations of the $(\ell+1)$st break), with this difference divided by the regression variance under the null hypothesis.

BP2 then shows how this can be inductively applied to determine consistently the number and location of structural breaks when they are unknown (Proposition 8). However this sequential approach relies on an accurate estimation of SSR, and since our robust estimation procedure does not grant us this, we must modify (7) for the degrees of freedom. To this end, observe that in (7): q is the number of variables with possible structural breaks; in the numerator, k is the number of breaks under the alternative hypothesis; and in the denominator, k is the difference between the number of breaks under the null and alternative hypotheses. Therefore, the appropriate F-test for ℓ vs. $\ell+1$ breaks, is of the form

$$\{F_T(\lambda_1,\ldots,\lambda_k;q) = \left(\frac{T-(k+1)q-p}{Tq}\right)(\delta' R'(RV(\delta)R')^{-1}R\delta \tag{8}$$

So once we have found the optimum break points, this provides a suitably consistent *test* for them, given our inability to use standard OLS.

The next issue is how to actually *compute* the break points, absent OLS assumptions. Here, we still rely on the Bai-Perron algorithm of minimizing the SSR, or equivalently maximizing the R^2. This approach was suggested to us by Jushan Bai in personal communication, as a likely consistent approach for the best fit regression. We hope that applications of their methods to problems such as ours will lead to a fleshing out of this theory to fully handle common deviations from the OLS world, such as we have. The close match with the best polynomial fit surely argues for the validity of this approach.

Finally, we note that we have essentially combined two portions of BP2. We use their sequential approach (just described) in determining the number of breaks — i.e. *given* the current breaks, is an additional one justified? But once this number ℓ of breaks is determined by the sequential procedure, we have opted to run the global algorithm to recompute the best location of ℓ breaks, over all choices.

A.2 The GEE Technique

We make use of Liang and Zeger's *Generalized Estimating Equation* (GEE) technique that is implemented in STATA statistical software, as it is specifically designed for panel data. This yields consistent estimates of the regression parameters and their variances under mild time dependence – in particular, our problem of AR4 autocorrelation. GEE requires that one provide the correlation matrix to indicate the specific nature of the autocorrelation. We again use the STATA option provided, based on an algorithm in Newton (1988).

On top of the above, we have additionally added the STATA White robust adjustment to handle heteroskedasticity.

B DIAGRAMMATIC TECHNIQUE

For all four graphs, we plot age on $Y(age)$ — equal either to a normalized natural log of price, or normalized ratio of quantity over mean quantity. We wish the ordinate of the graphs to accurately measure the expectation of Y, given only age, and unconditional on all other right side regressors. Accordingly, we replace the regressors by their mean values, and multiply by the computed beta coefficients. For instance, in Figures 1 and 2, the following equation was used to estimate $Y(age)$.

$$Y(age) = \alpha_0 + AGE + \sum_{j=1}^{J-1} \beta_j \bar{\alpha}_j + \sum_{k=1}^{K-1} \gamma_k \bar{\mu}_k + \sum_{m=1}^{M-1} \delta_m \bar{y}_i + \eta\bar{\sigma} + \epsilon_i \tag{9}$$

where $\bar{\alpha}_j$ is the expected value of the artist dummy variable, $\bar{\mu}_k$ and $\bar{y}_m$ are the respective mean values of the medium and year sold dummies, and $\bar{\sigma}$ is the mean log painting area.

Table 5: **Artists and their Age Ranges.** This first of all lists all artists common to our three data sets. It also provides the artists' age ranges for the paintings used in our age-value profile data set.

Artist	Ages	Artist	Ages
Andreas Achenbach	15–87	Maximilien Luce	16–81
Rosa Bonheur	22–70	Mino Maccari	34–85
Frank William Brangwyn	21–88	Reginald Marsh	21–55
William Callow	20–91	André Masson	26–84
Moshe Castel	21–71	Henri Michaux	33–84
Paul Cezanne	18–61	Giuseppe Migneco	29–83
Edward William Cooke	18–67	Sir John Everett Millais	17–64
Thomas Sidney Cooper	30–96	Claude Monet	14–80
Salvador Dali	21–77	Edvard Munch	17–80
Charles-Francois Daubigny	14–60	Adriaen Jansz van Ostade	22–74
Honoré Daumier	21–67	Max Papart	35–82
Edgar Degas	22–70	Jules Pincas (Pascin)	18–45
Eugene Delacroix	19–64	Pablo Picasso	18–66
André Derain	17–74	John Piper	27–83
Otto Dix	17–78	Camille Pissarro	33–73
Charles Edward Dixon	20–62	Pierre Auguste Renoir	31–78
Raoul Dufy	19–74	Diego Rivera	22–70
Jean Auguste Ingres	22–82	Henriette Ronner-Knip	13–85
Louis Gabriel Eugéne Isabey	18–73	Thomas Rowlandson	20–85
Augustus Edwin John	17–76	Aligi Sassu	17–76
Wilfredo Lam	25–80	Gino Severini	20–80
Albert Lebourg	18–73	Paul Signac	19–72
Tamara de Lempicka	22–77	Chäim Soutine	22–49
André Lhote	18–75	Henri de Toulouse-Latrec	15–36
Bruno Liljefors	21–77	Maurice Utrillo	18–69
Carl Liner	22–59	Andrew Wyeth	20–69
Laurence Stephen Lowry	19–85		

References

Agnello, Richard J., "Price Determinants and Investment Returns for Art: Evidence from Paintings' Auctions," August 1994. University of Delaware Working Paper 94-3.

Ashenfelter, Orley and Kathryn Graddy, "An Empirical Study of Sales Rates and Prices in Impressionist and Contemporary Art Auctions," 1999. Princeton Unversity Working Paper, May.

Bai, Jushan and Pierre Perron, "Computation and Analysis of Multiple Structural Change Models," August 1998. M.I.T. mimeo [BP1].

– and – , "Estimating and Testing Linear Models with Multiple Structural Changes," *Econometrica*, January 1998, *66*, 47–78 [BP2].

Galenson, David W., "The Careers of Modern Artists: Evidence from Auctions of Contemporary Paintings," December 1997. NBER Working Paper No. 6331.

Goetzmann, William N., "Accounting For Taste: Art and the Financial Markets Over Three Centuries," *The American Economic Review*, December 1993, *83*, 1370–1376.

Hamermesh, Daniel S. and Sharon M. Oster, "Aging and Productivity Among Economists," *The Review of Economics and Statistics*, February 1998, *80* (1), 154–156.

Kotlikoff, Lawrence J. and Jagadeesh Gokhale, "Estimating a Firm's Age-Productivity Profile Using the Present Value of Workers' Earnings," *Quarterly Journal of Economics*, November 1992, pp. 1215–1242.

Liang, Kung-Lee and Scott L. Zeger, "Longitudinal Data Analysis Using Generalized Linear Models," *Biometrika*, 1986, *73*, 13–22.

Mayer International Auction Records on CD-Rom, London: Acatos, Lausanne and Digital Media Resources Ltd., 1997.

Newton, Joseph H., *TIMESLAB: A Times Series Analysis Laboratory*, Belmont, CA: Wadsworth and Brooks/Cole, 1988.

Tauer, Loren, "Age and Farmer Productivity," *Review of Agricultural Economics*, 1995, *17*, 63–69.