Copenhagen Business School
HANDELSHØJSKOLEN

# Department of Economics

**Copenhagen Business School**

## Working paper 21-2020

# A Simple Theory of Pareto Earnings

Karl Harmenberg

Department of Economics – Porcelænshaven 16A, 1. DK-2000 Frederiksberg

# A Simple Theory of Pareto Earnings

Karl Harmenberg*

December 22, 2020

**Abstract**

I introduce a simple model which endogenously generates a Pareto distribution in top earnings, consistent with empirics. Workers inhabit different niches, and the earnings of a worker is determined by the niche-specific supply of labor and a constant-elasticity labor-demand curve. The highest paid workers are the ones that inhabit a niche with few other workers. A Pareto tail in earnings emerges as long as the distribution of workers over niches satisfies a regularity condition from extreme-value theory, satisfied by virtually all continuous distributions in economics.

## 1 Introduction

> These results are most remarkable. It is absolutely impossible to admit that they are only the result of chance. There must without doubt be a *cause* which produces the tendency for incomes to lie according to a certain curve. The shape of this curve seems to depend to a very small extent on the different economic conditions of the countries under consideration because the results are more or less the same for those countries whose economic conditions are as varied as those of England, Germany, the Italian towns, and even those of Peru.
>
> (Pareto, 1896)

The Pareto property of top earnings is one of the most striking regularities in economics. Since Pareto (1896)'s original discovery, it has been verified across regions and time periods that the top of the earnings distribution closely follows a Pareto distribution, i.e., the share $s$ of individuals with earnings above a certain earnings level $y$ is well approximated by the functional form $s \propto y^{-\alpha}$.

Since top earnings are Pareto distributed under many different institutional settings, and in different eras, an explanation of this fact should ideally rely on robust economic theory. In this paper, I provide such

*kha.eco@cbs.dk, Copenhagen Business School.

an explanation invoking the basic economic idea that scarcity leads to high prices, or in the case of labor, high earnings. Top earners are workers that have found a niche with few other workers. They can be highly specialized professionals such as specialist surgeons or lawyers, pop stars capturing the zeitgeist, or athletes with very few competitors at their level. Exactly how these top earners find their lucrative niches, e.g. , through luck or skill, is not important for the argument.

The model is simple. Workers are distributed over a continuum of different 'niches' and earnings in a particular niche is determined by the number of workers in the niche together with a downward-sloping constant-elasticity labor-demand curve. The top earners are the ones who find themselves in a niche with very few competitors. In this setting, for a large class of distributions of workers across niches, the induced earnings distribution features a Pareto distribution for top earnings.

**Related literature** The predominant framework for generating a Pareto tail in earnings is through random-growth processes, following Champernowne (1953) and Simon (1955).[1] This influential line of research combines a scale-invariant process for earnings with some 'reset' risk (which could be death in a perpetual-youth fashion), thereby accounting for top-earnings inequality in terms of a more primitive earnings process. By contrast, this paper provides a *static* theory of top-earnings inequality leveraging that scarcity leads to high prices. Introducing earnings dynamics to the static model is a straight-forward exercise, and I comment further on this in the discussion after the main result.

Geerolf (2017) also provides a static theory of top-earnings inequality. At a deeper level, although my model leverages variations in scarcity and Geerolf (2017) studies an assignment model generating superstar effects, the models share an approach to top-earnings inequality. Mathematically, both Geerolf (2017) and this paper employ what Sornette (2006) calls a 'power law change of variable close to the origin' to generate a Pareto tail in earnings. Geerolf (2017) shows that an assignment model with a constant-elasticity production function generates a Pareto tail in firm size and earnings, whereas the framework in this paper generates a Pareto tail in earnings through a supply-demand framework.

More broadly, Gabaix (2009) surveys the literature on power laws in economics and finance, and Sornette (2006) surveys a broader literature on power laws in the natural sciences.

# 2 Model

The labor market consists of a continuum of niches which workers can inhabit. The earnings of a worker in a niche is determined by the number of workers in the worker's niche and the best paid niches are therefore the ones with very few workers. The model does not take a stand on how the distribution of workers over

[1]See, e.g., Nirei and Aoki (2016), Toda and Walsh (2015), Gabaix, Lasry, Lions and Moll (2016), Jones and Kim (2018), Beare and Toda (2020)

niches is generated. Some well-paid niches may require skill to reach while access to other well-paid niches may be regulated, and yet other well-paid niches may require a high degree of luck to reach. Regardless of how the distribution of workers over niches arises, once we have a distribution of workers over niches, we can compute the induced earnings distribution. The main result of the paper is that, for a large class of distributions, the induced earnings distribution features an endogenous Pareto tail.

## 2.1 Setting

**Notation** Workers belong to different niches $x \in \mathbb{R}^n$. The number of workers in a particular niche $x$ is given by the density function $f_x(x)$. All workers in a given niche earn the same amount, $y(x)$.

**Earnings in a niche** The earnings $y(x)$ of a worker in a niche $x$ with mass of workers $f_x(x)$ is determined by an inverse labor-demand curve, $y(x) = D^{-1}(f_x(x))$. That is, the earnings in a niche is determined by the amount of workers inhabiting the niche. The inverse labor-demand curve $D^{-1}$ is shared across niches and satisfies the following properties:

**Assumption 1.** *The inverse labor-demand curve* $D^{-1} : \mathbb{R}_+ \to \mathbb{R}_+$ *is differentiable and strictly monotonically decreasing. The limit elasticity of income exists when the mass of workers* $L$ *approaches zero,*

$$\lim_{L \to 0} \frac{-(D^{-1})'(L)L}{D^{-1}(L)} = \frac{1}{\epsilon}.$$

This is equivalent to assuming that, locally near zero, earnings are determined by a constant-elasticity labor-demand curve, $y \propto L^{-1/\epsilon}$.

**The space of niches** Although the ‘true’ space of niches may be high dimensional, without loss of generality, we order the niches from most common to least common on the real non-negative half-line $\mathbb{R}_+$. Figure 2.1 depicts the basic idea: Think of the density of workers as a histogram and simply sort the individual columns of the histogram from largest to smallest. In Appendix A, I formalize this transformation and show that, with the original density being smooth and Morse and the space of niches being an oriented manifold equipped with a volume form, the resulting density on the positive real line is almost everywhere smooth and strictly monotonically decreasing.

**The distribution of workers over niches** We make some relatively weak assumptions on the distribution of workers over niches, $f_x$. The ordering of niches implies that $f_x$ is monotonically decreasing. If $f_x(x)$ is bounded from below on its support by some $\underline{d}$, then the earnings distribution is bounded from above by

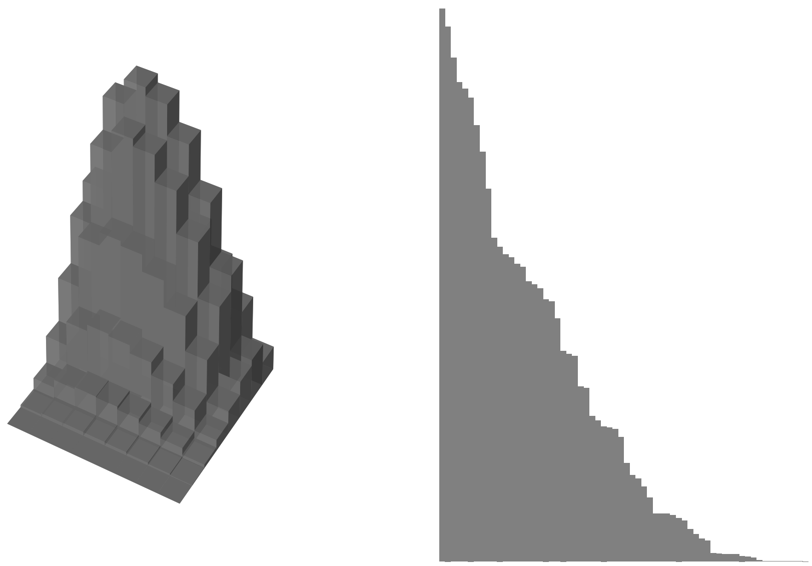

Figure 2.1: We can without loss of generality assume that the distribution of workers over niches is given by a monotonically decreasing density function on the positive real line by first "sorting the histogram". The two-dimensional distribution to the left is transformed to the one-dimensional distribution to the right by taking the histogram and sorting the columns from largest to smallest.

$D^{-1}(\underline{d})$. To have an unbounded income distribution, we therefore assume that for all $\epsilon > 0$, there exists a niche $x$ such that $0 < f_x(x) < \epsilon$. Finally, we assume that $f_x$ is regular in the sense of extreme-value theory.

**Assumption 2.** *The distribution of workers over niches, $f_x$, satisifies the following properties.*

1. *The density of workers $f_x : \mathbb{R}_+ \to \mathbb{R}_+$ is strictly monotonically decreasing.*
2. *For all $\epsilon > 0$, there exists a niche $x$ such that $0 < f_x(x) < \epsilon$.*
3. *The density of workers $f_x$ is regular, i.e., the following limit exists,*

$$\lim_{x \to \overline{x}} \frac{\partial}{\partial x} \frac{1 - F_x(x)}{f_x(x)} = \xi,$$

*with $\overline{x} = \sup_x\{f_x(x) > 0\} \in \mathbb{R}_+ \cup \{\infty\}$ denoting the supremum of the support of $f_x$. The limit $\xi$ satisfies $\xi > -1$.*

**Remark 1.** *The uniform distribution is regular with $\xi = -1$, the Weibull distribution is regular with $\xi < 0$, the Pareto and Fréchet distributions are regular with $\xi > 0$, and the Gaussian, log-normal, Gumbel, exponential, stretched exponential, and loggamma distributions are regular with $\xi = 0$.*

The limit condition in Assumption 2 stems from extreme-value theory and was introduced to economics by Gabaix and Landier (2008). It is satisfied by virtually all continuous distributions considered in economics and it is therefore a weak restriction on the density function to insist that it is regular in the above sense. By differentiating, Assumption 2 implies that $\lim_{x \to \bar{x}} -\frac{\bar{F}_x(x) f_x'(x)}{f_x(x)^2} = 1 + \xi$. Note also that $\xi > -1$ implies that $f_x$ is eventually strictly monotonically decreasing.

The constant $\xi$ measures the fatness of the tail of the distribution. For the thin-tailed distributions such as the normal distribution, the exponential distribution, and the log-normal distribution, $\xi = 0$. For bounded distributions such as the Weibull distribution, $\xi < 0$. For distributions with fat tails, such as the Pareto and Fréchet distributions, $\xi > 0$. In particular, for a Pareto distribution $\xi = 1/\alpha$.

**What is a niche?** In effect, we have assumed that the demand for labor is constant across niches. This is without loss of generality since we are free to redefine the size of a niche in a way that equalizes the level of demand across niches. Let earnings be given by $y = D^{-1}(a(x) f_x(x))$. Reparametrize the space of niches so that $d\tilde{x} = a(x)dx$. The new density function is given by $\tilde{f}_x(x) = f_x(x)/a(x)$ (such that $\tilde{f}_x(x)d\tilde{x} = f_x(x)dx$) and $y(x) = D^{-1}(\tilde{f}_x(x))$. For example, if demand is proportional to population density and, for intuition, we consider each niche as a discrete geographic bin, then the reparametrization means that the the grid of bins should be very fine on Manhattan but coarse in northern Sweden so that each niche contains an equivalent-sized population.

## 2.2 The earnings distribution

A distribution of workers across niches $f_x$ together with an inverse labor-demand function $D^{-1}$ generates an earnings distribution $f_y$. We say that the earnings distribution $f_y$ has a Pareto tail if

$$\lim_{y \to \infty} \frac{y f_y(y)}{1 - F_y(y)} = \alpha.$$

We can now state and prove the main result of the paper.

**Theorem 1.** *Given Assumption 1 on labor demand and Assumption 2 on the distribution of workers across niches, the earnings distribution features a Pareto tail. The Pareto tail coefficient is given by* $\alpha = \epsilon/(1+\xi)$.

*Proof.* Write $y_0 = D^{-1}(f_x(x_0))$. Since both $D^{-1}(\cdot)$ and $f_x(\cdot)$ are strictly monotonic, the correspondence is bijective and we have

$$1 - F_y(y_0) = 1 - F_x(x_0).$$

Differentiate both the left-hand side and the right-hand side with respect to $x_0$. We get

$$-f_y(y_0) \cdot (D^{-1})'(f_x(x_0)) \cdot f_x'(x_0) = f_x(x_0).$$

or

$$f_y(y_0) = -\frac{f_x(x_0)}{(D^{-1})'(f_x(x_0))f_x'(x_0)}.$$

Therefore,

$$\lim_{y_0 \to \infty} \frac{-y_0 f_y(y_0)}{1 - F_y(y_0)} = \lim_{x_0 \to \overline{x}} \frac{-D^{-1}(f_x(x_0))f_y(y_0)}{1 - F_x(x_0)} = \lim_{x_0 \to \overline{x}} \frac{-f_x(x_0)^2}{(1 - F_x)(x_0)f_x'(x_0)} \frac{-D^{-1}(f_x(x_0))}{f_x(x_0)(D^{-1})'(f_x(x_0))} = \frac{\epsilon}{1 + \xi}.$$

□

# 3 Discussion

Although Theorem 1 covers both the case where the distribution of workers over niches $f_x$ has bounded support ($\xi < 0$) and where the distribution of workers over niches features a fat tail ($\xi > 0$), it is instructive to first discuss the case of a thin-tailed distribution with unbounded support ($\xi = 0$) (such as the the normal distribution, the log-normal distribution, and the exponential distribution). For this class of distributions, Theorem 1 gives the sharp result that the Pareto tail parameter is equal to the elasticity of labor demand, independent of the particular distribution of workers across niches. Figure 3.1 shows graphically how the earnings distributions generated by an exponential, a normal and a log-normal distribution of workers over niches all generate asymptotic Pareto tails.

Although the log-normal distribution has, loosely speaking, many more outliers than the normal distribution, top-earnings inequality as captured by the Pareto coefficient is the same under both distributions of workers across niches. Why does the distribution of workers across niches not matter more for the Pareto tail parameter? Imagine you could distribute workers across niches and that your objective was to decrease top-earnings inequality. Would you lower the number of hedge fund managers? In partial equilibrium, reducing the number of hedge-fund managers lowers the number of highly-paid executives, reducing top-earnings inequality. However, in general equilibrium, reducing the number of hedge-fund managers increases the earnings of the remaining hedge-fund managers, thereby increasing top-earnings inequality. Theorem 1 shows that these two effects offset each other and leave top-earnings inequality intact, as long as the change in distribution does not change the fatness of the tail $\xi$ of the distribution of workers.

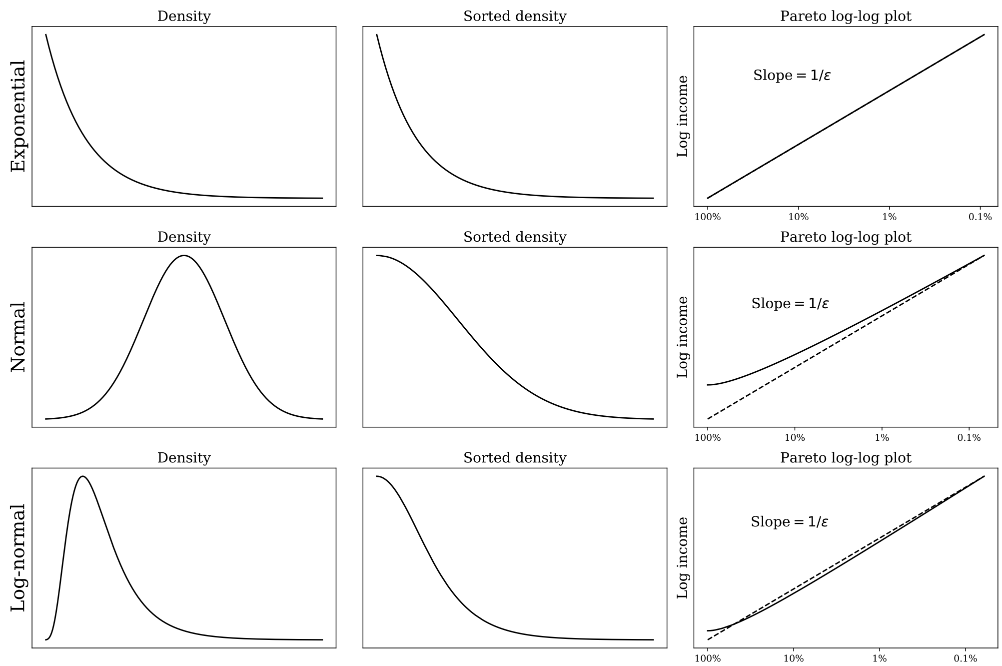


Figure 3.1: A constant-elasticity demand curve, $y = L^{-1/\epsilon}$, generates an exact Pareto distribution of income when workers are distributed according to an exponential distribution. For other distributions such as the normal and log-normal distribution, the income distribution asymptotically approaches a Pareto distribution.

**The missing assumption on supply** The model introduced in this paper makes structural assumptions on the demand side of the economy but remains vague with respect to the supply side of the economy. If labor supply is determined in a friction-less fashion that implies wage equalization, then the density of workers is either uniform or degenerate, violating Assumption 2. It is therefore important that there is some hetereogeneity or friction inducing a non-degenerate non-uniform density of workers over niches. For a concrete structural model of the supply side with heterogeneity in innate ability and optimal human capital accumulation, see Heathcote, Storesletten and Violante (2017). In their framework, it is costly to reach the outer niches and the individual-specific cost depends on 'innate ability'. By assuming an exponential distribution in innate ability, Heathcote, Storesletten and Violante (2017) generate an exact Pareto tail. This paper shows that the earnings distribution in Heathcote, Storesletten and Violante (2017) features a Pareto tail for a generic distribution of innate ability, not only for the analytically convenient exponential distribution.

**Implications for earnings dynamics** The model introduced in this paper can be used as a 'recipe' for designing income processes which are automatically consistent with the Pareto tail in earnings.

1. Let earnings in a niche be given by the inverse demand function $D^{-1}(L) = AL^{-1/\epsilon}$

2. Postulate any reasonable process in niche space

3. Compute the implied stationary distribution

4. Given the stationary distribution and the process in niche space, back out the earnings process

For example, let the process in niche space be a simple $AR(1)$, $x_{t+1} = \rho x_t + \epsilon_{t+1}$ with $\epsilon_{t+1} \sim N(0, \sigma^2)$. The stationary distribution in niche space is then $x \sim N(0, \sigma^2/(1-\rho^2))$. Write $\tilde{\sigma} = \sigma/\sqrt{1-\rho^2}$. Earnings in a given niche is

$$y(x) = A\left(\frac{1}{\sqrt{2\pi}\tilde{\sigma}}\exp\left(-\frac{x^2}{2\tilde{\sigma}}\right)\right)^{-1/\epsilon} = A\left(\sqrt{2\pi}\tilde{\sigma}\right)^{1/\epsilon}\exp\left(\frac{x^2}{2\epsilon\tilde{\sigma}^2}\right).$$

Use the rescaling $\hat{y} = y/y_{\min}$ with $y_{\min} = A\left(\sqrt{2\pi}\tilde{\sigma}\right)^{1/\epsilon}$. Then a few lines of algebra gives that the implied earnings process is

$$\log\hat{y}_{t+1} = \rho^2\log\hat{y}_t + \frac{\rho}{\sqrt{2\epsilon}\tilde{\sigma}}\sqrt{\log\hat{y}_t}\epsilon_t + \frac{1}{2\epsilon\tilde{\sigma}^2}\epsilon_t^2.$$

Aside from the $\epsilon_t^2$, this process is a Cox-Ingersoll-Ross process in log earnings (in continuous time, the $\epsilon_t^2$ term instead becomes a drift term). Through an $AR(1)$ in niche space, we have thus recovered the well-known fact that a Cox-Ingersoll-Ross process in log earnings generates a Pareto tail in the stationary earnings distribution. The result of this paper suggests a broader research agenda: Consider a broad class of processes in niche space. Which of these processes best accounts for earnings dynamics at the micro level? These processes will, by design, generate a Pareto tail in earnings.

Ultimately, any theory of the Pareto property of top earnings needs to rely on some functional-form assumption. The functional-form assumption necessary for Theorem 1 is the assumption that a stable elasticity of the labor-demand curve exists. Mathematically, the earnings distribution inherits its power-law functional form from the power-law functional form of labor demand. The result in this paper is, as evidenced by the length of the proof of Theorem 1, not difficult from a mathematical perspective. The contribution of the paper is a simple *economic* theory generating an earnings distribution with a Pareto tail. Compared to the previous literature, we do not need assumptions on labor supply aside from the technical assumptions in Assumption 2. The model in this paper is therefore really a *class* of models, each particular model having a different description of labor supply. Furthermore, the assumption on labor demand, a stable elasticity, is at the very least a common assumption in economics. It is analogous to the operational assumption from the top-income taxation literature (following Saez (2001)) that there exists a stable top-income earnings elasticity.

To conclude, this paper introduces a simple theory of why a Pareto tail in earnings is generically observed. In a simple model relating low supply/scarcity to high earnings, a Pareto tail in earnings emerges under the assumption of a shared labor-demand curve with a limit elasticity. Through the lens of the model, the increase in top-earnings inequality observed in the last decades is accounted for by a fall in the elasticity of labor demand for the professional class. Although an increase in top-earnings inequality can also be accounted for by a change in the distribution of workers across niches, such a change must be coarse and qualitative—a mere change from a normal distribution to a log-normal distribution is not sufficient to change top-earnings inequality.

## References


Beare, B.K., Toda, A.A., 2020. Geometrically stopped Markovian random growth processes and Pareto tails. `arXiv:arXiv:1712.01431v3`.

Champernowne, D., 1953. A Model of Income Distribution. Economic Journal 63, 318–351.

Gabaix, X., 2009. Power Laws in Economics and Finance. Annual Review of Economics 1, 255–294. URL: `http://www.annualreviews.org/doi/abs/10.1146/annurev.economics.050708.142940`, doi:10.1146/annurev.economics.050708.142940.

Gabaix, X., Landier, A., 2008. Why has CEO pay increased so much? Quarterly Journal of Economics 123, 49–100. URL: `http://www.nber.org/papers/w12365`.

Gabaix, X., Lasry, J.M., Lions, P.L., Moll, B., 2016. The Dynamics of Inequality. Econometrica 84, 2071–2111. doi:10.3982/ecta13569.

Geerolf, F., 2017. A Theory of Pareto Distributions. Working Paper , 1–49.

Heathcote, J., Storesletten, K., Violante, G.L., 2017. Optimal tax progressivity: An analytical framework. Quarterly Journal of Economics 132, 1693–1754. doi:10.1093/qje/qjx018.

Jones, C.I., Kim, J., 2018. A schumpeterian model of top income inequality. Journal of Political Economy 126, 1785–1826. doi:10.1086/699190.

Matsumoto, Y., 2002. An Introduction to Morse Theory. American Mathematical Society, Providence, Rhode Island.

Nirei, M., Aoki, S., 2016. Pareto distribution of income in neoclassical growth models. Review of Economic Dynamics 20, 25–42. URL: `http://dx.doi.org/10.1016/j.red.2015.11.002`, doi:10.1016/j.red.2015.11.002.

Pareto, V., 1896. Cours d.économie politique.

Saez, E., 2001. Using Elasticities to Derive Optimal Income Tax Rates. Review of Economic Studies 68, 205–229.

Simon, H.A., 1955. On a Class of Skew Distribution Functions. Biometrika 42, 425. doi:10.2307/2333389.

Sornette, D., 2006. Critical Phenomena in Natural Sciences: Chaos, Fractals, Selforganization and Disorder: Concepts and tools. Springer Science & Business Media.

Toda, A.A., Walsh, K., 2015. The double power law in consumption and implications for testing Euler equations. Journal of Political Economy 123, 1177–1200. doi:10.1086/682729.

# A Appendix

Let $\mathbf{N}$, the space of niches, be an orientable manifold equipped with a volume form $\omega$. The distribution of workers over niches is given by the density function $f$, which is a Morse function such that no critical points share the same value (see Matsumoto (2002) for a proof that almost all functions satisfy this property).

**Proposition A.1.** *There exists an almost everywhere smooth strictly monotonic function* $\tilde{f} : \mathbb{R}_+ \to \mathbb{R}_+$ *such that for* $\tilde{x} = \int_{\mathbf{N}} [f(x) \geqslant y] d\omega$*, we have* $\tilde{f}(\tilde{x}) = y$*.*

*Proof.* Define $G(y) = \int_{\mathbf{N}} [f(x) \geqslant y] d\omega$. We first want to show that $G : \mathbb{R} \to \mathbb{R}$ is continuous and almost everywhere smooth.

Let $y$ be a regular value of $f$. Then $M_y = \{x : f(x) = y\}$ is a manifold and $M_y^\epsilon = \{x : |f(x) - y| \leqslant \epsilon\}$ is diffeomorphic to $M_y \times [-\epsilon, \epsilon]$ for $\epsilon$ sufficiently small. The $G(y)$ is therefore smooth in a neighborhood of a regular value. Since the critical points of $f$ are of measure zero, $G$ does not jump at critical values and $G$ is therefore continuous and almost everywhere smooth and strictly monotonic. $G$ is clearly invertible around a regular value. Since $G$ is continuous and strictly monotonic outside critical values, it is also invertible at the critical values. Define $\tilde{f} = G^{-1}$. It is clear that $\tilde{f}$ satisfies $\tilde{f}(\tilde{x}) = y$. □