

# Mitigating Eddington and Malmquist Biases in Latent-Inclination Inference of the Tully-Fisher Relation

HAI FU<sup>1</sup>

<sup>1</sup>*Department of Physics & Astronomy, University of Iowa, Iowa City, IA 52242, USA*

## ABSTRACT

The Tully-Fisher relation is a vital distance indicator, but its precise inference is challenged by selection bias, statistical bias, and uncertain inclination corrections. This study presents a Bayesian framework that simultaneously addresses these issues. To eliminate the need for individual inclination corrections, inclination is treated as a latent variable with a known probability distribution. To correct for the distance-dependent Malmquist bias arising from sample selection, the model incorporates Gaussian scatter in the dependent variable, the distribution of the independent variable, and the observational selection function into the data likelihood. To mitigate the statistical bias—termed the “general Eddington bias”—caused by Gaussian scatter and the non-uniform distribution of the independent variable, two methods are introduced: (1) analytical bias corrections applied to the dependent variable before likelihood computation, and (2) a dual-scatter model that accounts for Gaussian scatter in the independent variable within the likelihood function. The effectiveness of these methods is demonstrated using simulated datasets. By rigorously addressing selection and statistical biases in a latent-variable regression analysis, this work provides a robust approach for unbiased distance estimates from standardizable candles, which is critical for improving the accuracy of Hubble constant determinations.

**Keywords:** Linear regression (1945) — Bayesian statistics (1900) — Maximum likelihood estimation (1901) — Scaling relations (2031) — Disk galaxies (391)

## 1. INTRODUCTION

The Tully-Fisher relation (TFR) is an empirical correlation between luminosity (or mass) and maximum rotation velocity of spiral galaxies (Tully & Fisher 1977). The tightness of the correlation makes it a remarkable distance indicator in the category of standardizable candles, so it was immediately utilized to measure the Hubble constant  $H_0$  (Sandage & Tammann 1976; Tully & Fisher 1977). In fact, the discovery of the correlation between peak luminosity and rate of decline for Type Ia supernovae, the most popular standardizable candles today, was made possible by distances from the TFR and surface brightness fluctuations (Phillips 1993). But it was soon realized that the TF distances were affected by local peculiar velocities (Tully 1988) and the Malmquist bias (Giraud 1987). The classic Malmquist bias (Eddington 1914; Malmquist 1922) is caused by observational selection effects and the intrinsic scatter in the luminosity of standard candles. The intrinsic scatter of the TFR is expected, as the physical properties of the galaxies are unlikely to be fully captured by rotation velocity alone. When a sample is subdivided in redshift, the Malmquist bias starts to show distance-dependent behaviors, causing the artifact that the mean luminosity at fixed rotation velocity (and  $H_0$  as a consequence) increases outward (Sandage 1994a,b).

Multiple methods have been proposed to correct for the distance-dependent Malmquist bias: (1) the method of replacing the linewidth-predicted mean luminosity with the analytical mean luminosity that depends on both flux limit and redshift (Sandage 1994a; Butkevich et al. 2005), (2) the method of normalized distances, which enhances the S/N of the luminosity-redshift diagram by shifting galaxies of different luminosities diagonally along the luminosity limit and uses the plateau at lower normalized distances to obtain an unbiased luminosity estimate (Bottinelli et al. 1986, 1988), and (3) the method of maximum likelihood estimation (MLE) that includes the distance-dependent sample selection function in the data likelihood function (Willick 1994; Willick et al. 1997). Among the three, the MLE is the most powerful and versatile method, because it utilizes the full data set, naturally accounts for heteroscedastic measurement errors, and fits the slope, the intercept, and the intrinsic scatter simultaneously. Willick et al. (1997) derived the likelihood functions for the two commonly used “unidirectional” models<sup>2</sup>: the forward TFR and the inverse TFR, where the independent variable is velocity width and luminosity, re-

<sup>2</sup> These unidirectional models include the error and the intrinsic scatter of the dependent variable but neglect the scatter of the independent variable.

spectively. Recent applications of the MLE method on the Cosmicflows-4 (CF4) data include Kourkchi et al. (2022) for the inverse TFR and Boubel et al. (2024) for the forward TFR. However, even with identical data set, the inferred TF parameters from the two unidirectional models disagree well beyond the inferred statistical uncertainties. This happens because there is a separate bias in linear regression that stems from the independent variable.

It has long been recognized that linear regression coefficients are biased when the independent variable is measured with error, but the model ignores this error (e.g., Fuller 1987). In § 3, I will derive the analytical formula of this additional bias, which originates from the scatter and the distribution function of the independent variable. Because the result resembles that of the classic Eddington bias of the mean true value, it will be referred to as the “general Eddington bias”. The popular methods to mitigate this bias in general linear regression include (1) the bivariate correlated errors and intrinsic scatter (BCES) estimator (Akritas & Bershady 1996) and (2) the Bayesian dual-scatter methods (LINMIX<sup>3</sup> by Kelly 2007 and ROXY<sup>4</sup> by Bartlett & Desmond 2023)<sup>5</sup>. The former uses the moments of the data to estimate regression coefficients by correcting the bias in the ordinary least squares (OLS) estimator, while the latter derives the data likelihood function that includes not only the intrinsic scatter and selection function of the dependent variable, but also the covariance matrixes of measurement errors and the intrinsic distribution of the independent variable (approximated there as a mixture of Gaussians for flexibility and mathematical convenience). The dual-scatter methods can reduce the general Eddington bias but cannot address the distance-dependent Malmquist bias, as these linear regression methods do not treat distance as a separate variable within the dependent variable. For example, in Kelly (2007) the sample selection function  $p(I = 1|x, y)$  depends only on  $x$  and  $y$ ; while for a flux-limited survey, the selection function of  $y$  should depend on distance (which is part of  $y$  if it represents luminosity).

Besides the distance-dependent Malmquist bias and the general Eddington bias, a major data-related issue limits the precision of our statistical inference of the TFR. Before modeling, the observational data must be corrected for galaxy inclination. Inclination correction is necessary because (1) the observed velocity width is projected along the line of sight and the observed luminosity is attenuated by dust in the plane of the disk and (2) random inclinations cause non-Gaussian scatters in both axes. But inclination estimation with axial

ratio is highly uncertain even for well resolved galaxies, because (1) the axial ratio depends the method (isophote fitting vs. surface brightness modeling), the depth of the photometric data, and the bulge-to-disk ratio of the galaxy, and (2) the inclination angle also depends on the unknown edge-on thickness of the disk. As a result, galaxies with lower inclinations (e.g.,  $< 45^\circ - 60^\circ$  from face-on) are often discarded in TF studies because of their larger inclination corrections. Inclination correction also causes correlated errors because the same inclination is used for both luminosity and velocity width.

To avoid the problems caused by individual inclination correction, one can model the inclination statistically by treating it as a latent variable with a known probability density function (pdf). It is a prime time to develop such methods, because wide-area HI surveys like ALFALFA (Haynes et al. 2018) and WALLABY (Westmeier et al. 2022) continue to produce large samples of HI-selected galaxies at all inclination angles. In Fu (2024), I introduced an iterative latent-inclination method to restore the TFR from the full ALFALFA sample using measurements *not* corrected for inclination. The method is highly efficient and produces a tighter correlation than that from inclination-corrected data. But the method requires binning and the biases in the data are carried directly into the restored TFR. For unbiased inference of the TFR, one could incorporate the latent variable in the data likelihood function. Obreschkow & Meyer (2013) proposed a latent-inclination MLE method for the inverse TFR, where luminosity is used to predict velocity width. But they missed an important term in their derivation: the intrinsic distribution function of the independent variable. Even though this term drops out in the likelihood function of the inverse TFR (see § 5.2), its omission led the authors to incorrectly conclude that for a dual-scatter model, the scatter of the independent variable can be added in quadrature to the scatter of the dependent variable. As a result, their latent-variable MLE model is unidirectional, hence does not account for the general Eddington bias.

To summarize, precise and unbiased inference of the TFR is limited by the distance-dependent Malmquist bias, the general Eddington bias, and elusive measurement errors in the inclination angle. While the literature addresses each of these problems individually, no solution currently tackles all three simultaneously. For example, the unidirectional MLE method (Willick et al. 1997) accounts for the distance-dependent Malmquist bias, but it neglects the general Eddington bias from the scatter of the independent variable. The Gaussian-mixture dual-scatter Bayesian method (Kelly 2007; Bartlett & Desmond 2023) can account for the general Eddington bias, but it does not correct for the distance-dependent Malmquist bias because distance is not separated from luminosity in these general regression methods. In ad-

<sup>3</sup> <https://github.com/jmeyers314/linmix>

<sup>4</sup> <https://github.com/DeaglanBartlett/roxy>

<sup>5</sup> While LINMIX requires the user to set (1) the Gaussian scatter of the independent variable and (2) a reasonable number of Gaussians to model the distribution function, ROXY can infer these parameters from the data.

dition, neither methods address the inclination-angle problem: they both require the input line widths to be individually corrected by inclination, so low-inclination galaxies are excluded to avoid the most problematic corrections. Here I develop latent-variable, dual-scatter, Bayesian methods that incorporate (1) the Sine pdf of inclination angle, (2) the distance-dependent selection function of luminosity, (3) the Gaussian measurement errors and intrinsic scatters in both the dependent and the independent variables, and (4) the distribution function of the independent variable. The precision and efficacy of the methods in mitigating the biases will be tested with synthetic data that realistically simulate (1) the luminosity function of galaxies, (2) the distribution in inclination angle and redshift, (3) the data censorship due to detection limit, and (4) the measurement errors and intrinsic scatters in both luminosity and rotation velocity.

This paper is organized as follows. In §2, I describe the observational data used in TFR studies and formulate the TFR problem with shorthand notations. Next in §3, I start by discussing how biases of the regression coefficient impact measurements of the Hubble constant. Then I describe how observational selection effects and luminosity scatter bias the mean luminosity via the distance-dependent Malmquist bias. Finally, I show how the scatter in the independent variable distorts the distribution of the dependent variable through a formula analogous to the classic Eddington bias of the true mean value. Both biases shift the first moment of the dependent variable away from the true correlation, causing biases in the regression coefficients. The next four sections detail the likelihood-based framework that tackles the three primary challenges of Tully-Fisher relation inference: distance-dependent Malmquist bias, general Eddington bias, and inclination correction. First in §4, I outline the methodology. Next in §5, I derive the data likelihood functions for two latent-inclination unidirectional models: the forward model and the inverse model that use rotation velocity and luminosity as the independent variable, respectively. In §6, I implement the likelihood functions in a Bayesian framework and employ a Markov-Chain Monte Carlo method to sample the posterior pdf of the inferred parameters. Simulated data of disk galaxies with random sky orientations are then generated to test the numerical methods, to quantify the statistical uncertainties, and to expose the biases of the inferred parameters. As expected, results are biased when the independent variable is measured with error, due to the general Eddington bias described in §3. Two extensions of the unidirectional models are then introduced to mitigate the general Eddington bias. First, in §7.1, I reverse the expected Eddington bias in the velocity width data and use the moment-shifted data to constrain the inverse model; and in §7.2, I introduce the latent-inclination bidirectional dual-scatter model by incorporating scatter of the independent variable in the likelihood

function. Tests on simulated datasets show that both methods yield nearly unbiased estimates of model parameters, confirming the theory in §3. Finally, I summarize the work and discuss future extensions and applications in §8.

To keep the main text focused on key concepts and results, I defer extended equations and important numerical considerations to the Appendices. Appendix A presents the complete analytical expressions for the conditional pdfs that constitute the likelihood functions for flux-limited samples. Appendix B gives the alternative derivation of the dual-scatter model, starting from the inverse model, showing that it is bidirectional (symmetric). Appendix C outlines the key numerical methods that enhanced the evaluation speed of the dual-scatter model’s likelihood function by three orders of magnitude, rendering it computationally practical.

In general, Roman letters are used to denote observable quantities and Greek letters are reserved for model parameters. I will use  $\tilde{x}$  to indicate measured value (with errors) and  $x$  to indicate the true value. Similarly,  $\tilde{\theta}$  and  $\theta$  represent the estimated value and the true value of parameter  $\theta$ , respectively. And the bias of the parameter estimation is defined as  $B_\theta \equiv \tilde{\theta} - \theta$ .

Python functions and a Jupyter notebook implementing the methods described in this paper are publicly available on GitHub under the MIT license: [https://github.com/fuhaiaastro/TFR\\_biases](https://github.com/fuhaiaastro/TFR_biases) and on Zenodo under an open-source Creative Commons Attribution license: [doi:10.5281/zenodo.16378199](https://doi.org/10.5281/zenodo.16378199).

## 2. FORMULATION OF THE PROBLEM

The TFR has various empirical forms. The luminosity axis could be absolute magnitude, stellar mass, or baryonic mass; and the velocity axis could be line widths from neutral hydrogen (H I), molecular gas (CO), or ionized gas ([O II]). I will formulate the problem using the TFR between baryonic mass and H I line width, because (1) the addition of gas mass to stellar mass tends to reduce the curvature of the correlation (e.g., McGaugh et al. 2000; Kourkchi et al. 2022) and (2) H I is the best tracer of the maximum rotation velocity thanks to its extended spatial distribution. This choice does not restrict the application of the methods to only baryonic TFR, since the derived expressions can be easily adjusted to suit magnitude-based TFRs.

The baryonic TFR is a power-law correlation between the baryonic mass (stellar + gas) and the edge-on rotation velocity. In logarithmic, it is a linear relation with two coefficients ( $\beta, \gamma$ ):

$$\log M_b = \beta[\log W - \log \sin(\text{inc}) - 2.5] + \gamma \quad (1)$$

where  $\log M_b$  is the logarithmic baryonic mass in solar mass ( $M_\odot$ ) and  $\log W - \log \sin(\text{inc})$  is the observed velocity width ( $W$  in  $\text{km s}^{-1}$ ) corrected for the inclination angle ( $\text{inc}$ ). The

relation is anchored at  $\log V_0 = 2.5$  dex (or  $V_0 = 316 \text{ km s}^{-1}$ ), so that the intercept  $\gamma$  is the baryonic mass of spirals with edge-on velocity width of  $316 \text{ km s}^{-1}$ .

The baryonic mass is calculated as the *apparent* baryonic mass multiplied by the square of the luminosity distance, which in logarithmic is:

$$\log M_b = \log m_b + 2 \log D_L \quad (2)$$

The apparent mass ( $\log m_b$ ) includes the  $4\pi$  term and is calculated using the H I 21 cm flux ( $S_{21}$  in  $\text{Jy km s}^{-1}$ ) and apparent magnitude in an optical/near-IR filter ( $m_\lambda$ ) using the following relation:

$$m_b = 2.356 \times 10^5 K_g S_{21} + 10^{-0.4(m_\lambda - M_{\lambda,\odot} - 25)} (M/L)_\lambda \quad (3)$$

where  $K_g$  is the H I-to-gas conversion factor (typically assumed to be 1.33 to account for Helium),  $M_{\lambda,\odot}$  is the absolute magnitude of the Sun (Willmer 2018), and  $(M/L)_\lambda$  is the color-estimated mass-to-light ratio in solar units ( $M_\odot/L_\odot$ ). The latter two must be in the same filter as the galaxy magnitude, as indicated by the subscript  $\lambda$ .

For most galaxies, distance is not a direct observable, but rather an inferred quantity from the observed redshift and a cosmological model. At low redshift ( $z < 0.1$ ), the redshift-derived luminosity distance can be accurately calculated using the Taylor expansion form of Caldwell & Kamionkowski (2004):

$$D_L \approx \frac{cz}{H_0} \left[ 1 + \frac{1}{2}(1-q_0)z - \frac{1}{6}(1-q_0-3q_0^2+j_0)z^2 \right] \quad (4)$$

where  $cz$  is the cosmological redshift in  $\text{km s}^{-1}$ ,  $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$  is the assumed Hubble constant,  $q_0 = -\frac{\ddot{a}a}{\dot{a}^2} = \Omega_{m,0}/2 - \Omega_{\Lambda,0}$  is the deceleration parameter, and  $j_0 = -\frac{\ddot{a} \cdot \dot{a}^2}{\dot{a}^3} = \Omega_{m,0} + \Omega_{\Lambda,0}$  is the cosmic jerk. When adopting  $\Omega_{m,0} = 0.315$ ,  $\Omega_{\Lambda,0} = 0.685$  from Planck Collaboration et al. (2020),  $q_0 = -0.53$  and  $j_0 = 1$ .

In the following, I will use shorthand symbols for the logarithmic of the apparent mass, the distance squared, the velocity width ratio, and the sine of the inclination:

$$\begin{aligned} m &\equiv \log m_b \\ d &\equiv 2 \log D_L \\ w &\equiv \log W - 2.5 \\ i &\equiv \log \sin(\text{inc}) \end{aligned} \quad (5)$$

With these shorthands, the TFR in Eq. 1 can be rewritten as:

$$m + d = \beta(w - i) + \gamma \quad (6)$$

The data available to the astronomer are  $\{\tilde{m}_i, \tilde{w}_i, \tilde{d}_i\}_{k=1}^N$  for a sample of  $N$  galaxies that survived the observational selection function  $S(\tilde{m}, \tilde{w})$ . The goal of parameter inference is to optimally utilize these data to constrain the TF parameters  $(\beta, \gamma)$ , ideally free of statistical and selection biases.

### 3. SELECTION BIAS AND STATISTICAL BIAS

It is important to mitigate biases in the inferred regression coefficients of the TFR because a biased  $\gamma$  directly leads to a biased Hubble constant ( $H_0$ ). Two separate samples are needed to estimate  $H_0$ : a larger and more distant redshift sample and a smaller and closer zero-point sample. The distances of the former are calculated using an assumed value of  $H_0$ , and the distances of the latter are from standard candles on the lower rung of the distance ladder (e.g., Cepheids). The two separate samples are needed for  $H_0$  measurement because (1) the redshifts of the zero-point sample are strongly influenced by peculiar velocities, and (2) the distances of the redshift sample are inferred from an assumed  $H_0$ . The TF coefficients  $(\beta, \gamma)$  are estimated using the redshift sample, while the zero-point sample determines the intercept ( $\gamma_{ZP}$ ) with Cepheid distances assuming the same slope  $\beta$ . Since it is fair to assume that both samples follow the same TFR, any difference between the intercepts from the two samples ( $\gamma$  and  $\gamma_{ZP}$ ) implies a deviation between the actual value of  $H_0$  and the assumed Hubble constant ( $70 \text{ km s}^{-1} \text{ Mpc}^{-1}$ ):

$$\log(H_0/70) = 0.5(\gamma - \gamma_{ZP}) \quad (7)$$

As a result, the bias in  $\gamma$  will directly transfer to the inferred Hubble constant:

$$B_{H_0} \approx 35 \ln 10 B_\gamma = 4.84(B_\gamma/0.06) \text{ km s}^{-1} \text{ Mpc}^{-1} \quad (8)$$

A bias of 0.06 dex (or 0.15 mag) in the intercept would lead to a bias of  $\sim 5 \text{ km s}^{-1} \text{ Mpc}^{-1}$  in the Hubble constant, comparable to the current ‘‘Hubble tension’’ (e.g. Riess et al. 2024).

Note that  $B_\gamma$  is only one of the two biases involved in the Hubble constant; the other is the bias of  $\gamma_{ZP}$  from the zero-point sample. It is thus critical to use the same statistical model for the redshift and the zero-point samples, in addition to consistent measurements of rotation velocities and masses (or luminosities) across the two samples (Bradford et al. 2016). If the intercept biases from the two samples are equal, they will cancel out, resulting in an unbiased  $H_0$  estimate. However, this is unlikely, as the biases depend on sample characteristics that typically differ between the two samples. Therefore, to obtain an unbiased Hubble constant determination, both samples must be analyzed using statistical models that account for their specific characteristics to correct for their distinct biases. In the following, I describe how random Gaussian scatter in both the dependent and independent variables biases regression coefficients through mechanisms akin to the classic Malmquist and Eddington biases.

The classic Malmquist bias (Eddington 1914; Malmquist 1922) is caused by data censorship and the Gaussian luminosity scatter of a standard candle. The classic Eddington bias of the mean true value (Dyson 1926; Eddington 1940) is due to measurement error and gradient in the distribution



function of the measurements. The two biases are orthogonal to each other: Malmquist takes the true mean luminosity as known and compare it to the mean of the measured luminosities that survived the selection ( $BM \equiv \langle \tilde{x} \rangle - x$ ), while Eddington takes the measurement as known and compare it to the mean of the unknown true values ( $BE \equiv \tilde{x} - \langle x \rangle$ ). However, the two biases are often conflated because, in the simplest cases, their mathematical outcomes appear similar. The classic Malmquist bias of absolute magnitude for uniformly distributed standard candles with a Gaussian luminosity function of a mean of  $M$  and a standard deviation of  $\sigma_M$  is  $\langle \tilde{M} \rangle - M = -1.382\sigma_M^2$ , and the Eddington bias of apparent magnitude for uniformly distributed sources with constant luminosity and a measurement error of  $\sigma_m$  is  $\langle \tilde{m} \rangle - \langle m \rangle = -1.382\sigma_m^2$ . In both cases, the constant 1.382 is  $0.6 \ln 10$ .

To formulate the problem, consider a family of standardizable candles on an idealized linear correlation:  $y(x) = \beta x + \gamma$ . For galaxies on the TFR,  $y$  would be the logarithmic luminosity and  $x$  the inclination-corrected logarithmic velocity width. And the inverse function of  $y(x)$  is  $x(y) = (y - \gamma)/\beta$ .

First, consider random Gaussian scatter only in the luminosity axis,  $\tilde{y} = y(x) + \epsilon_y$ , where the random variable  $\epsilon_y$  is drawn from a normal distribution with zero mean and a standard deviation of  $\sigma_y$ . So the conditional pdf of  $\tilde{y}$  at a given  $x$  is a Gaussian:

$$p(\tilde{y}|x) = \frac{1}{\sqrt{2\pi}\sigma_y} \exp\left(-\frac{[\tilde{y} - y(x)]^2}{2\sigma_y^2}\right) \quad (9)$$

When the sample is selected above a logarithmic luminosity limit of  $y_l$  that depends on distance ( $D_L$ ;  $y_l = 2 \log D_L + \text{const.}$  for a flux-limited sample), the selection function and the luminosity dispersion  $\sigma_y$  shifts the mean luminosity of the observed sample relative to the correlation-predicted mean luminosity according to:

$$\begin{aligned} y(x) - \langle \tilde{y} \rangle_x &= \frac{\int_{y_l}^{\infty} [y(x) - \tilde{y}] p(\tilde{y}|x) d\tilde{y}}{\int_{y_l}^{\infty} p(\tilde{y}|x) d\tilde{y}} \\ &= -\sigma_y \sqrt{\frac{2}{\pi}} \frac{\exp[-(y_l - y(x))^2 / (2\sigma_y^2)]}{\text{erfc}[(y_l - y(x)) / (\sqrt{2}\sigma_y)]} \end{aligned} \quad (10)$$

For the TFR,  $y(x)$  is the correlation-predicted luminosity for a given rotation velocity and  $\langle \tilde{y} \rangle_x$  is the mean of measured luminosities at the same rotation velocity. Because the difference between the two depends on distance, this is termed the “distance-dependent Malmquist bias” (e.g., Willick 1994; Butkevich et al. 2005).

Next, consider the scatter is instead only in the  $x$ -axis,  $\tilde{x} = x(y) + \epsilon_x$ , where the random variable  $\epsilon_x$  is drawn from a normal distribution with zero mean and a standard deviation of  $\sigma_x$ . In this case, there is no selection function, but the mean  $y$  will still be shifted from the correlation-predicted value, because the distribution of  $y$  at a given  $\tilde{x}$  is altered from

a Gaussian according to Bayes’ rule:

$$p(y|\tilde{x}) = \frac{p(\tilde{x}|y)p(y)}{p(\tilde{x})} \quad (11)$$

where the conditional pdf of  $\tilde{x}$  at a given  $y$  is a Gaussian:

$$\begin{aligned} p(\tilde{x}|y) &= \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left(-\frac{[\tilde{x} - x(y)]^2}{2\sigma_x^2}\right) \\ &= \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left(-\frac{[y(\tilde{x}) - y]^2}{2(\beta\sigma_x)^2}\right) \end{aligned} \quad (12)$$

For the TFR, the difference between the correlation-predicted luminosity at a measured rotation velocity  $[y(\tilde{x})]$  and the mean luminosity at the same rotation velocity ( $\langle y \rangle_{\tilde{x}}$ ) can be calculated as:

$$\begin{aligned} y(\tilde{x}) - \langle y \rangle_{\tilde{x}} &= \int [y(\tilde{x}) - y] \frac{p(\tilde{x}|y)p(y)}{p(\tilde{x})} dy \\ &= \int [y(\tilde{x}) - y] \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left(-\frac{[y(\tilde{x}) - y]^2}{2(\beta\sigma_x)^2}\right) \frac{p(y)}{p(\tilde{x})} dy \\ &= \frac{\beta^2\sigma_x}{\sqrt{2\pi}p(\tilde{x})} \int \frac{\partial}{\partial y} \left[ \exp\left(-\frac{[y(\tilde{x}) - y]^2}{2(\beta\sigma_x)^2}\right) \right] p(y) dy \end{aligned} \quad (13)$$

The symmetry between  $y$  and  $y(\tilde{x})$  in the exponential function allows the chain rule:

$$\begin{aligned} y(\tilde{x}) - \langle y \rangle_{\tilde{x}} &= -\frac{\beta^2\sigma_x}{\sqrt{2\pi}p(\tilde{x})} \int \frac{\partial}{\partial y(\tilde{x})} \left[ \exp\left(-\frac{[y(\tilde{x}) - y]^2}{2(\beta\sigma_x)^2}\right) \right] p(y) dy \\ &= -\frac{\beta\sigma_x^2}{p(\tilde{x})} \int \frac{\partial}{\partial \tilde{x}} [p(\tilde{x}|y)p(y)] dy \end{aligned} \quad (14)$$

The final result is obtained by taking the derivative out of the integral (Leibniz’s integral rule):

$$\begin{aligned} y(\tilde{x}) - \langle y \rangle_{\tilde{x}} &= -\frac{\beta\sigma_x^2}{p(\tilde{x})} \frac{d}{d\tilde{x}} \left( \int p(\tilde{x}|y)p(y) dy \right) \\ &= -\beta\sigma_x^2 \frac{d \ln p(\tilde{x})}{d\tilde{x}} \end{aligned} \quad (15)$$

The result is analogous to the classic Eddington bias of the mean true value (Dyson 1926; Eddington 1940):

$$\tilde{x} - \langle x \rangle_{\tilde{x}} = -\sigma^2 \frac{d \ln p(\tilde{x})}{d\tilde{x}} \quad (16)$$

where  $\tilde{x}$  is the measured value,  $\langle x \rangle_{\tilde{x}} = \int x p(x|\tilde{x}) dx$  the mean of the true values that could have produced the measured value,  $\sigma$  the measurement error, and  $p(\tilde{x})$  the distribution function of the measured values. In fact, the classic Eddington bias is a special case of Eq. 15 for  $\beta = 1$  and  $\gamma = 0$ . Thus, this bias is termed the “general Eddington bias”.

Both biases affect the estimates of regression coefficients by distorting the underlying data distribution. It is easy to

imagine how the general Eddington bias leads to biased regression coefficients. Suppose  $p(\tilde{x})$  is a simple exponential distribution function,  $p(\tilde{x}) \propto \exp(\tilde{x}/x_0)$ , the Eddington bias is a constant, implying a vertical shift of the best-fit relation from the truth, causing the intercept to change while keeping the slope intact. For other distributions (e.g., power-law or Schechter function), the Eddington bias varies with  $\tilde{x}$ , leading both the slope and the intercept to drift away from the truth. The distance-dependent Malmquist bias affects the regression coefficients in a more complicated way because the bias depends not only on luminosity scatter and flux limit, but also on redshift and the unknown true correlation.

The biases operate on the data simultaneously but are driven by different sample characteristics: Malmquist by luminosity dispersion and luminosity limit  $[\sigma_y, y_l(cz)]$ , while Eddington by dispersion of the independent variable and its distribution function  $[\sigma_x, p(\tilde{x})]$ . From a modeling perspective, correcting the biases in parameter estimation thus requires statistical models that build in all of those sample characteristics  $[\sigma_y, y_l(cz), \sigma_x, p(\tilde{x})]$ .

#### 4. OVERVIEW OF THE METHODOLOGY

Although the TFR is often described as a simple linear correlation in logarithmic scales, the inference of its slope and intercept is not a simple linear regression problem, because of several important characteristics of the TFR problem:

- The dependent variable (e.g., baryonic mass) is a combination of two observables: the apparent mass and the luminosity distance from redshift. For flux-limited samples, the selection function of the dependent variable depends on distance (not mass).
- The independent variable (maximum rotation velocity) requires de-projecting the apparent H I line width to the edge-on perspective. So it requires estimating the inclination angle of the disk relative to the line-of-sight.
- There are measurement errors in all four observables, and there are intrinsic dispersions in both axes<sup>6</sup>.

Because of these characteristics, precise and unbiased inference of the TFR is limited by three major factors:

- The uncertain estimate of the inclination angle.

- The distance-dependent Malmquist bias (Eq. 10) due to (1) the separation of flux limit and luminosity limit in a sample covering a range of distances, and (2) the measurement error and the intrinsic dispersion in mass.
- The general Eddington bias (Eq. 15) due to (1) non-uniform distributions of galaxies in rotation velocity, and (2) the measurement error and the intrinsic scatter in rotation velocity.

In this work, I will tackle all three problems simultaneously in a self-consistent Bayesian framework:

- To avoid errors associated with inclination measurements, galaxy inclination is treated as a latent variable with a known pdf, which is then marginalized when computing the data likelihood function.
- To mitigate the Malmquist bias, apparent masses and redshift-inferred distances are kept separate in all expressions, and the conditional probability of the apparent mass is normalized to account for the distance-dependent sample truncation.
- To mitigate the Eddington bias, two approaches are introduced: (1) the analytical formula of the bias from § 3 is utilized to correct for the moment of the data in an iterative fashion (§ 7.1), and (2) both the Schechter distribution function and the scatter of the independent variable are included in the data likelihood function of a “bidirectional dual-scatter model” (§ 7.2). The model is bidirectional (symmetric), producing equivalent likelihood functions regardless of whether rotation velocity or baryonic mass is selected as the independent variable.

#### 5. DERIVATION OF THE LIKELIHOOD FUNCTIONS

To begin the Bayesian inference process, I derive the data likelihood functions for the two unidirectional models: the forward model with rotation velocity as the independent variable and the inverse model with mass as the independent variable. These models are useful because they isolate the Eddington bias (§ 6.3) and lay the foundation for the more complicated bidirectional dual-scatter model (§ 7.2).

##### 5.1. The Forward Model

The forward model uses velocity width ( $w-i$ ) to predict the mass ( $m+d$ ). It also assumes Gaussian dispersion along the mass axis. By using  $w$  as an independent variable, its measurement error is inherently ignored in the forward model. To make this explicit in the notation and to distinguish the forward model with the dual-scatter model in § 7.2, I use  $w$  in place of  $\tilde{w}$  in this subsection. In addition, the residual redshift noise propagated into the distance parameter  $\tilde{d}$  is ignored and

<sup>6</sup> The dual intrinsic dispersion model posits that galaxies with identical luminosity may exhibit a range of maximum rotation velocities, and similarly, galaxies with identical rotation velocities may show a range of luminosities. The extent of these dispersions, both absolute and relative, should be derived from the data rather than arbitrarily set by the observer. Like measurement errors, the intrinsic Gaussian dispersions along the two axes cannot be merged into a single Gaussian when distribution functions are non-uniform.

$d$  is used in place of  $\tilde{d}$  to make this assumption explicit. But I will comment on how to properly include redshift noise in the likelihood function.

Given the TFR in Eq. 6, the forward model can be expressed as:

$$\tilde{m} = \beta(w - i) - d + \gamma + \epsilon_{im} + \epsilon_{em} \quad (17)$$

where the random variables represent the intrinsic dispersion ( $\epsilon_{im}$ ) and the measurement error ( $\epsilon_{em}$ ), and they are drawn from Gaussian distributions with zero means and standard deviations of  $\sigma_{im}$  and  $\sigma_{em}$ , respectively.

In addition, the forward model also requires parameterizing the distribution function of edge-on rotation velocity ( $w - i$ ) of the galaxy population (hereafter “velocity function”). Given the TFR that links velocity with mass, the velocity function is closely related to the mass function, which is parameterized as a Schechter function:

$$\frac{dn}{d(m+d)} = \phi_* 10^{(\alpha+1)(m+d-M_*)} \exp(-10^{m+d-M_*}) \quad (18)$$

where  $\alpha$  is the faint-end slope,  $M_*$  the characteristic mass (above which the exponential drop-off commences). Typically, the normalization factor  $\phi_*$  is in units of  $\text{Mpc}^{-3} \text{dex}^{-1}$ ; for the purpose here, I redefine  $\phi_*$  as the factor that normalizes the distribution function to unity so that it becomes a properly normalized pdf  $p(m+d)$ .

Substituting  $m+d$  with  $\beta(w-i) + \gamma$ , one obtains the corresponding velocity function:

$$\frac{dn}{d(w-i)} = \beta \phi_* 10^{(\alpha+1)[\beta(w-i)-v_*]} \exp[-10^{\beta(w-i)-v_*}] \quad (19)$$

where  $v_* = M_* - \gamma$ , which equals the characteristic logarithmic velocity multiplied by  $\beta$ . Notice the extra  $\beta$  in front of  $\phi_*$  because  $d(m+d) = \beta d(w-i)$ .

In the following, I will start by expressing the joint probability of  $(\tilde{m}, d, w, i)$ , then marginalize it over  $i$  to obtain the joint probability of the observables  $(\tilde{m}, d, w)$ , and finally using the joint probability and the selection function to calculate the conditional probability of  $\tilde{m}$ , which makes up the likelihood function.

The joint probability that a galaxy has the three observables  $(\tilde{m}, \tilde{d}, w)$  and an inclination  $i$  can be factorized using the multiplication rule:

$$p(\tilde{m}, d, w, i) = p(\tilde{m}|w, i, d) p(i|w) p(w) p(d) \quad (20)$$

When choosing these factors, I have considered relations between the variables and have made the following assumptions:

- the velocity function parameters does not evolve over the redshift range covered by the sample, so that  $p(w|i)$  does not depend on distance.

- inclination and distance are independent variables, so that  $p(i, d) = p(i)p(d)$ .

The particular set of conditional pdfs are straightforward to express. The first term is obtained by marginalizing the joint pdf  $p(\tilde{m}, m|w, i, d)$  over  $m$ , which is a convolution of two Gaussians:

$$p(\tilde{m}|w, i, d) = \int_{-\infty}^{\infty} p(\tilde{m}|m) p(m|w, i, d) dm \\ = \frac{1}{\sigma_m \sqrt{2\pi}} \exp \left[ -\frac{[\tilde{m} + d - \beta(w-i) - \gamma]^2}{2\sigma_m^2} \right] \quad (21)$$

Per the convolution theorem of Gaussians, the variance of the convolved Gaussian,  $\sigma_m^2$ , is the sum of the variances from intrinsic dispersion and measurement error:

$$\sigma_m^2 = \sigma_{im}^2 + \sigma_{em}^2 \quad (22)$$

Note that the likelihood-based method naturally handles heteroscedastic measurement errors, because a different  $\sigma_{em}$  can be supplied for each source while  $\sigma_{im}$  is kept as a constant for the whole sample.

The second term is the probability of projected velocity given the inclination and it can be expressed by the normalized Schechter velocity function in Eq. 19:

$$p(w|i) = \beta \phi_* 10^{(\alpha+1)[\beta(w-i)-v_*]} \exp[-10^{\beta(w-i)-v_*}] \quad (23)$$

The velocity function is a critical component of the forward model because the term depends on  $i$  so it does not drop out in the conditional pdf in Eq. 30 below.

The third term is the prior pdf of the inclination parameter. Assuming isotropic random orientation on the sky, the pdf of inclination angle  $\text{inc}$  is a sine function:  $p(\text{inc}) = \sin(\text{inc})$  for  $0 < \text{inc} < \pi/2$ . The pdf of  $i \equiv \log \sin(\text{inc})$  can then be derived using the pdf identity  $p(i)di = p(\text{inc})d\text{inc}$ , and the result is:

$$p(i) = \ln 10 \frac{10^{2i}}{\sqrt{1-10^{2i}}} \quad (24)$$

Note that  $i \leq 0$  given its definition.

The last term is the probability of observing a galaxy at a given distance. It is proportional to the integrated galaxy volume density at the distance  $n(d)$  (i.e., the redshift distribution) multiplied by the survey volume  $\Omega D_L^2 dD_L$ . Expressed in distance parameter  $d$ , the probability is:

$$p(d) \propto 10^{3d/2} n(d) \quad (25)$$

When one ignores the difference between the redshift-inferred distance  $\tilde{d}$  and the true distance  $d$  (as assumed here), this term drops out in the conditional pdf in Eq. 30 below. But if one opts to account for this difference (e.g., Willick et al. 1997), the following term needs to be multiplied to the joint pdf in Eq. 28 below to obtain  $p(\tilde{m}, w, \tilde{d}, d, i)$ :

$$p(\tilde{d}|d) = \frac{1}{\sigma_d \sqrt{2\pi}} \exp \left[ -\frac{(\tilde{d}-d)^2}{2\sigma_d^2} \right] \quad (26)$$

where the standard deviation is proportional to the ratio between the velocity noise ( $\sigma_{cz}$ ) and the redshift:

$$\sigma_d \approx \frac{2}{\ln 10} \frac{\sigma_{cz}}{cz} \quad (27)$$

and an additional marginalization over  $d$  is required when calculating the joint pdf of the three observables in Eq. 29. Note that because of the additional  $d$ -dependent terms in  $p(d)$  (Eq. 25), the marginalization over  $d$  is not as simple as the marginalization over  $m$  in Eq. 21. For this reason, it is incorrect to absorb this integral over  $d$  by simply adding  $\sigma_d^2$  to  $\sigma_m^2$  in Eq. 22.

Combining the terms defined above, the joint probability can be written is:

$$\begin{aligned} p(\tilde{m}, w, d, i) &= p(d) \times \frac{\ln 10 \cdot 10^{2i}}{\sqrt{1 - 10^{2i}}} \\ &\times \frac{1}{\sigma_m \sqrt{2\pi}} \exp \left[ -\frac{[\tilde{m} + d - \beta(w - i) - \gamma]^2}{2\sigma_m^2} \right] \\ &\times \beta \phi_* 10^{(\alpha+1)[\beta(w-i)-\gamma]} \exp[-10^{\beta(w-i)-\gamma}] \end{aligned} \quad (28)$$

which can then be marginalized over  $i$  to obtain the joint probability of the three observables:

$$p(\tilde{m}, w, d) = \int_{-\infty}^0 p(\tilde{m}, w, d, i) di \quad (29)$$

The observational selection function,  $S(\tilde{m}, w)$ , truncates the joint pdf,  $p(\tilde{m}, w, d)$ , so that the resulting pdf is no longer normalized. To account for  $S(\tilde{m}, w)$  and the associated Malmquist bias, one shall compute the conditional pdf of  $\tilde{m}$  by normalizing the truncated joint pdf:

$$\begin{aligned} p(\tilde{m}|w, d) &= \frac{S(\tilde{m}, w)p(\tilde{m}, w, d)}{\int_{-\infty}^0 S(\tilde{m}, w)p(\tilde{m}, w, d) d\tilde{m}} \\ &= \frac{S(\tilde{m}, w) \int_{-\infty}^0 p(\tilde{m}|w, i, d)p(w|i)p(i) di}{\int_{-\infty}^0 S(\tilde{m}, w) \int_{-\infty}^0 p(\tilde{m}|w, i, d)p(w|i)p(i) di d\tilde{m}} \end{aligned} \quad (30)$$

Notice that the distance distribution function,  $p(d)$ , has cancelled out in the division. When the sample is simply flux-limited by a step function, the integral over  $\tilde{m}$  on the denominator becomes an error function, making it much faster to evaluate. I provide the full expression of the conditional pdf for flux-limited samples in Eq. A3 in Appendix A.

The data likelihood function ( $\mathcal{L}$ ) is defined as the probability of the data given the model. For the forward model, the data are  $\{\tilde{m}_k\}_{k=1}^N$  and the model is the combination of the model parameters  $\theta \equiv (\beta, \gamma, \sigma_m, v_*, \alpha)$  and the independent variables  $\{w_k, d_k\}_{k=1}^N$ . It is thus the product of the conditional probabilities in Eq. 30 of all valid data points. In practice, its logarithmic is preferred:

$$\ln \mathcal{L} = \sum_{k=1}^N \ln p(\tilde{m}_k | w_k, d_k) \quad (31)$$

## 5.2. The Inverse Model

The inverse model uses mass ( $m + d$ ) to predict velocity width ( $w - i$ ) and assumes Gaussian dispersion along the velocity axis:

$$\tilde{w} - i = (m + d - \gamma) / \beta + \epsilon_{iw} + \epsilon_{ew} \quad (32)$$

where the random variables represent the intrinsic dispersion in velocity ( $\epsilon_{iw}$ ) and measurement error ( $\epsilon_{ew}$ ). By using  $m + d$  as the independent variable, its measurement error is inherently ignored in the inverse model, so I use  $m$  and  $d$  in place of  $\tilde{m}$  and  $\tilde{d}$  in this subsection.

Starting by factorizing the joint pdf:

$$p(\tilde{w}, m, d, i) = p(\tilde{w} | m, d, i) p(m | d) p(d) p(i) \quad (33)$$

Similar to the forward model, the first term is obtained by marginalizing the joint pdf  $p(\tilde{w}, w | m, d, i)$  over  $w$ :

$$\begin{aligned} p(\tilde{w} | m, d, i) &= \int_{-\infty}^{\infty} p(\tilde{w} | w) p(w | m, d, i) dw \\ &= \frac{1}{\sigma_w \sqrt{2\pi}} \exp \left[ -\frac{[\tilde{w} - i - (m + d - \gamma) / \beta]^2}{2\sigma_w^2} \right] \end{aligned} \quad (34)$$

where the variance  $\sigma_w^2$  is the sum of the variances from intrinsic dispersion and measurement error:

$$\sigma_w^2 = \sigma_{iw}^2 + \sigma_{ew}^2 \quad (35)$$

The second term is the normalized mass function in Eq 18:

$$p(m | d) = \phi_* 10^{(\alpha+1)(m+d-M_*)} \exp(-10^{m+d-M_*}) \quad (36)$$

For simplicity, I assume that the baryonic mass function is independent of inclination angle, implying that intrinsic dust extinction has been corrected for when calculating the stellar-mass-to-light ratio using color. The last two terms of the joint pdf are the same as in Eqs. 24 and 25. Therefore, the joint pdf can be written as:

$$\begin{aligned} p(\tilde{w}, m, d, i) &= p(d) \times \frac{\ln 10 \cdot 10^{2i}}{\sqrt{1 - 10^{2i}}} \\ &\times \frac{1}{\sigma_w \sqrt{2\pi}} \exp \left[ -\frac{[\tilde{w} - i - (m + d - \gamma) / \beta]^2}{2\sigma_w^2} \right] \\ &\times \phi_* 10^{(\alpha+1)(m+d-M_*)} \exp(-10^{m+d-M_*}) \end{aligned} \quad (37)$$

which can then be marginalized over  $i$  to obtain the joint probability of the three observables:

$$p(\tilde{w}, m, d) = \int_{-\infty}^0 p(\tilde{w}, m, d, i) di \quad (38)$$

The conditional probability of  $\tilde{w}$  is then obtained by normalization:

$$\begin{aligned} p(\tilde{w} | m, d) &= \frac{S(m, \tilde{w}) p(\tilde{w}, m, d)}{\int_{-\infty}^0 S(m, \tilde{w}) p(\tilde{w}, m, d) d\tilde{w}} \\ &= \frac{S(m, \tilde{w}) \int_{-\infty}^0 p(\tilde{w} | m, i, d) p(i) di}{\int_{-\infty}^0 S(m, \tilde{w}) \int_{-\infty}^0 p(\tilde{w} | m, i, d) p(i) di d\tilde{w}} \end{aligned} \quad (39)$$



where both the mass distribution function,  $p(m|d)$ , and the distance distribution,  $p(d)$ , have dropped out. Hence, only the three TFR model parameters are constrained by the inverse model:  $\theta \equiv (\beta, \gamma, \sigma_w)$ . When the selection function depends only on  $m$  [i.e.,  $S(m, \tilde{w}) = S(m)$ ], it also drops out from the expression, further simplifying the result:

$$\begin{aligned} p(\tilde{w}|m, d) &= \int_{-\infty}^0 p(\tilde{w}|m, d, i) p(i) di \\ &= \frac{\ln 10}{\sigma_w \sqrt{2\pi}} \int_{-\infty}^0 \frac{10^{2i}}{\sqrt{1 - 10^{2i}}} \exp \left[ -\frac{[\tilde{w} - i - (m + d - \gamma)/\beta]^2}{2\sigma_w^2} \right] di \end{aligned} \quad (40)$$

The data likelihood function is obtained from the conditional pdf in the same manner as in the forward model:

$$\ln \mathcal{L} = \sum_{k=1}^N \ln p(\tilde{w}_k | m_k, d_k) \quad (41)$$

## 6. IMPLEMENTATION AND TESTING

This section describes how I implement the latent-inclination unidirectional models in a Bayesian framework (§ 6.1), simulate TF galaxy samples with random sky orientations (§ 6.2), and test the models with simulated data sets (§ 6.3). The testing reveals the biases of inferred parameters when the independent variable randomly scatters, as expected from the general Eddington bias described in § 3. For measurement errors and sample sizes typical for the ALFALFA sample, the biases are much greater than the inferred statistical uncertainties and they scale with the scatter of the independent variable.

### 6.1. Parameter Inference

The Bayesian theorem uses the data likelihood function  $\ln \mathcal{L}$  to update the prior knowledge of the model to provide the posterior distribution of the model parameters given the data ( $D$ ):

$$\ln p(\theta|D) = \ln \mathcal{L} + \ln p(\theta) - \ln p(D) \quad (42)$$

where  $p(\theta)$  is the prior pdf of the model parameters and  $p(D)$  is the Bayesian evidence of the model as a whole.

When prior knowledge is insufficient or ignored, the priors are assumed to be flat within specified bounds:

$$p(\theta) = \prod_j \frac{1}{\Delta \theta_j} \quad (43)$$

When such bounded flat priors are assumed, the posterior peaks at the parameters that maximizing the data likelihood, making the Bayesian inference essentially the same as MLE.

The evidence is defined as the marginalized probability of the data ( $D$ ) over all model parameters ( $\theta$ ) of a given model:

$$p(D) = \int p(D|\theta) p(\theta) d\theta \quad (44)$$

It is often challenging to evaluate the evidence accurately because it is a multi-dimensional integral that requires calculations of the data likelihood function over the entire parameter space. But it can be safely ignored for the purpose of parameter inference, because it is a constant for a given model and a given data set. In addition, Markov-Chain Monte Carlo (MCMC) algorithms do not require *normalized* pdf to sample the parameter space at frequencies proportional to the pdf, making them the favored method for Bayesian inference.

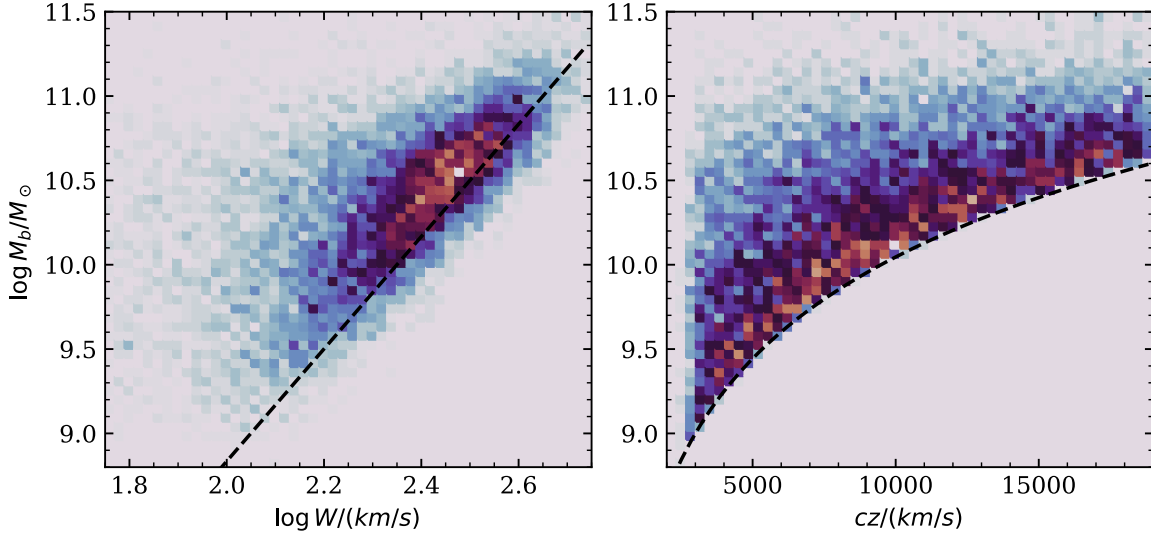
I employ the affine-invariant MCMC ensemble sampler implemented in the Python code `emcee`<sup>7</sup> (Foreman-Mackey et al. 2013) to sample the posterior pdfs. The ensemble approach is naturally parallelizable, inherently handles correlated parameters, and minimizes the need for manual tuning of step sizes for poorly scaled parameters. Standard procedure to set up the sampler is followed. The number of walkers is set to an integer number of CPU cores and is set between two and three times the number of free parameters. Starting from random initial positions within the bounds, the walkers proceed until the length of the chains exceeds 50 times the estimated autocorrelation length ( $l$ ; typically between 40 and 200 steps). After discarding the initial  $2l$  steps (“burn-in”) and thinning the chains by keeping one step for every  $l/2$  steps, the chains from all walkers are combined to form the final parameter array, whose distribution functions should trace the posterior pdfs of the parameters given the data. For each model parameter, I quote the best-fit value and the  $1\sigma$  (68%) credible interval given by the 50th, 16th, and 84th percentiles of the marginalized cumulative distribution function.

### 6.2. Simulated Data for Testing

To test whether the models in § 5 can recover unbiased parameter estimation, I simulate three samples of galaxies where the ground truth of the model parameters are known. I start the process by random sampling of the velocity function in narrow redshift intervals. This is necessary because other than the Schechter function parameters, the survey volume and the redshift distribution also change the probability of sampling a galaxy with a particular rotation velocity, and both of these factors could vary with redshift. So the first step is to divide the specified redshift range into a fine grid of equal intervals ( $\delta cz$ ), then for each redshift, a subsample is generated following the procedure below:

1. The redshift grid center,  $cz_0$ , represents the true cosmological redshift of the simulated subset, i.e., all galaxies in this subset are at the same true distance  $d(cz_0)$ , calculated using Eq. 4.
2. The initial subsample size for each  $cz_0$  is calculated as  $N_0 = a(cz_0)^{2+n} \delta cz$ , where  $a$  is a scale factor and  $n$  is

<sup>7</sup> <https://emcee.readthedocs.io>



**Figure 1.** Distributions of simulated sample C in baryonic mass vs. projected line width (left) and baryonic mass vs. redshift (right). The dashed line in the left panel shows the prescribed TFR and the dashed curve in the right panel shows the mass limit as a function of redshift.

**Table 1.** Input Parameters vs. Inferred Parameters

Parameter	Simulated Data Set		
	A	B	C
Input Parameters (Truth)			
$\sigma_m$	0.15	0.0	0.15
$\sigma_w$	0.0	0.045	0.045
$\gamma$	10.50	10.50	10.50
$\beta$	3.33	3.33	3.33
$v_*$	0.3	0.3	0.3
$\alpha$	-1.27	-1.27	-1.27
$m_l$	5.736	5.736	5.736
$N$	10,132	10,111	10,147
Inferred Parameters from Forward Model			
$\gamma$	$10.502^{+0.004}_{-0.004}$	$10.444^{+0.004}_{-0.003}$	$10.450^{+0.005}_{-0.005}$
$\beta$	$3.355^{+0.031}_{-0.031}$	$3.094^{+0.024}_{-0.025}$	$3.164^{+0.039}_{-0.037}$
$\sigma_m$	$0.162^{+0.003}_{-0.003}$	$0.145^{+0.003}_{-0.003}$	$0.210^{+0.004}_{-0.004}$
$v_*$	$0.264^{+0.033}_{-0.033}$	$0.273^{+0.033}_{-0.030}$	$0.248^{+0.040}_{-0.037}$
$\alpha$	$-1.206^{+0.039}_{-0.040}$	$-1.225^{+0.042}_{-0.043}$	$-1.181^{+0.046}_{-0.043}$
Inferred Parameters from Inverse Model			
$\gamma$	$10.560^{+0.004}_{-0.003}$	$10.500^{+0.003}_{-0.003}$	$10.554^{+0.005}_{-0.004}$
$\beta$	$3.587^{+0.025}_{-0.024}$	$3.337^{+0.023}_{-0.024}$	$3.593^{+0.029}_{-0.029}$
$\sigma_w$	$0.046^{+0.001}_{-0.001}$	$0.046^{+0.001}_{-0.001}$	$0.062^{+0.001}_{-0.001}$

NOTE—The fitted parameters are assumed to have flat priors within the following bounds:  $10 < \gamma < 11$ ,  $2.5 < \beta < 4.5$ ,  $0.001 < \sigma_m < 0.3$ ,  $0.001 < \sigma_w < 0.1$ ,  $-1 < v_* < 1$ , and  $-2 < \alpha < 0$ . For both the forward and the dual-scatter models, the standard mass limit  $m_l$  is fixed to the input value.

the power-law index of the integrated galaxy volume density. Both parameters can be adjusted to change

the redshift distribution and the size of the final sample after the selection function is applied.

3. *Optional:* Random redshifts perturbed by velocity noise,  $\{cz_i\}_{i=1}^{N_0}$ , are drawn around  $cz_0$  following a normal distribution with a mean of  $cz_0$  and a standard deviation of  $\sigma_{cz}$ :  $\mathcal{N}(cz_0, \sigma_{cz}^2)$ . Distances,  $\{\tilde{d}_i\}$ , are then calculated from  $\{cz_i\}$  using Eq. 4. The velocity noise is thus propagated into  $\{\tilde{d}_i\}$ .
4. Intrinsic edge-on velocities,  $\{v_i\}$ , are drawn from the Schechter function in Eq. 19 with specified parameters ( $v_*, \alpha$ ) using the inverse transform sampling method. The default sampling range is  $-3.5 < \beta v - v_* < 1.5$ .
5. Intrinsic masses,  $\{M_i\}$ , are calculated from the edge-on velocities using an idealized TFR:  $M_i = \beta v_i + \gamma$ . The exact linear relation between  $\{v_i\}$  and  $\{M_i\}$  makes sampling the intrinsic velocity function equivalent to sampling the intrinsic mass function in Eq. 18, as long as  $M_*$  is set to equal  $v_* + \gamma$ .
6. Apparent masses,  $\{\tilde{m}_i\}$ , are calculated from the idealized TFR mass by subtracting the true distance and adding Gaussian scatters:  $\tilde{m}_i = M_i - d(cz_0) + \epsilon_m$ . The random scatter  $\epsilon_m$  is drawn from a normal distribution with a standard deviation of  $\sigma_m$ :  $\mathcal{N}(0, \sigma_m^2)$ .
7. Projected velocities,  $\{\tilde{w}_i\}$ , are calculated from the edge-on velocities by adding random inclinations and Gaussian scatters:  $\tilde{w}_i = v_i + \log \sin \epsilon_i + \epsilon_w$ . To simulate samples with isotropic random orientations, the cosine of the inclination angle,  $\cos \epsilon_i$ , is drawn from a uniform distribution between 0 and 1, which then allow  $\sin \epsilon_i$  to be computed as  $\sqrt{1 - \cos^2 \epsilon_i}$ . The random scatter in

$w$  is drawn from a normal distribution with a standard deviation of  $\sigma_w$ , i.e.,  $\epsilon_w \sim \mathcal{N}(0, \sigma_w^2)$ .

8. Finally, the observational selection function,  $S(\tilde{m}, \tilde{w})$ , is applied to the simulated data set, and the survived subsample is kept.

For simplicity, I use homoscedastic scatters in the simulated data (i.e.,  $\sigma_m$  and  $\sigma_w$  are both constants for the entire sample), although the Bayesian methods can handle heteroscedastic measurement errors. It is worth noting that the frequently reported increase in mass/magnitude dispersion among slow rotators in a TF sample is mainly due to the smearing of the luminosity limit across the redshift range of the sample (see Fig. 1 for example). Dispersion measurements in mass/magnitude near these luminosity limits should be deemed as upper limits, thus they do not offer evidence for heteroscedastic intrinsic dispersion in the TFR.

Three simulated samples were generated for testing. They assume modest levels of scattering that are typically expected in galaxy samples like the ALFALFA or CF4:

- A.  $\sigma_m = 0.15, \sigma_w = 0$  (no scatter in  $w$ )
- B.  $\sigma_m = 0, \sigma_w = 0.045$  (no scatter in  $m$ )
- C.  $\sigma_m = 0.15, \sigma_w = 0.045$  (scatters in both axes)

Samples A and B match the assumptions of the forward model and the inverse model, respectively, and Sample C represents the more realistic case where scatters are present in both axes. The ratio of the dispersions,  $\sigma_m/\sigma_w$ , matches the assumed slope of the TFR ( $\beta = 3.33$ ) to ensure equal contributions.

The other model parameters are shared across the samples:

- The velocity noise due to residual peculiar motion is ignored ( $\sigma_{cz} = 0$ ) to avoid the additional marginalization over the true distance, as explained in § 5.1.
- The selection function depends only on the apparent logarithmic mass ( $\tilde{m}$ ) and is a step function with a detection limit at  $m_l = 5.736$ , appropriate for the ALFALFA sample.
- The power-law index of the integrated galaxy volume density is set to  $n = -1$  to produce a flatter redshift distribution, and the sampling scale factor is set to produce  $\sim 10,000$  galaxies over the redshift range between  $4,000 < cz < 18,000 \text{ km s}^{-1}$  above the detection limit. The sample size and its redshift distribution are comparable to those of H I-selected galaxy samples like the ALFALFA and CF4.
- The TF parameters are set to  $\beta = 3.33$  and  $\gamma = 10.5$ , similar to the values found in Kourkchi et al. (2022).

- The Schechter function parameters for the intrinsic velocity and mass distributions are set to  $\alpha = -1.27$ ,  $v_* = 0.3$ , and  $M_* = v_* + \gamma = 10.8$ . These values are consistent with the tabulated baryonic mass function of Papastergis et al. (2012).

The input parameters and the final sample sizes ( $N$ ) are listed in Table 1. As an example, Fig. 1 shows the distributions of simulated sample C in the plane of mass vs. projected velocity width and mass vs. redshift. The former illustrates the combination of Sine scatter and Gaussian scatters, while the latter illustrates the observational selection function.

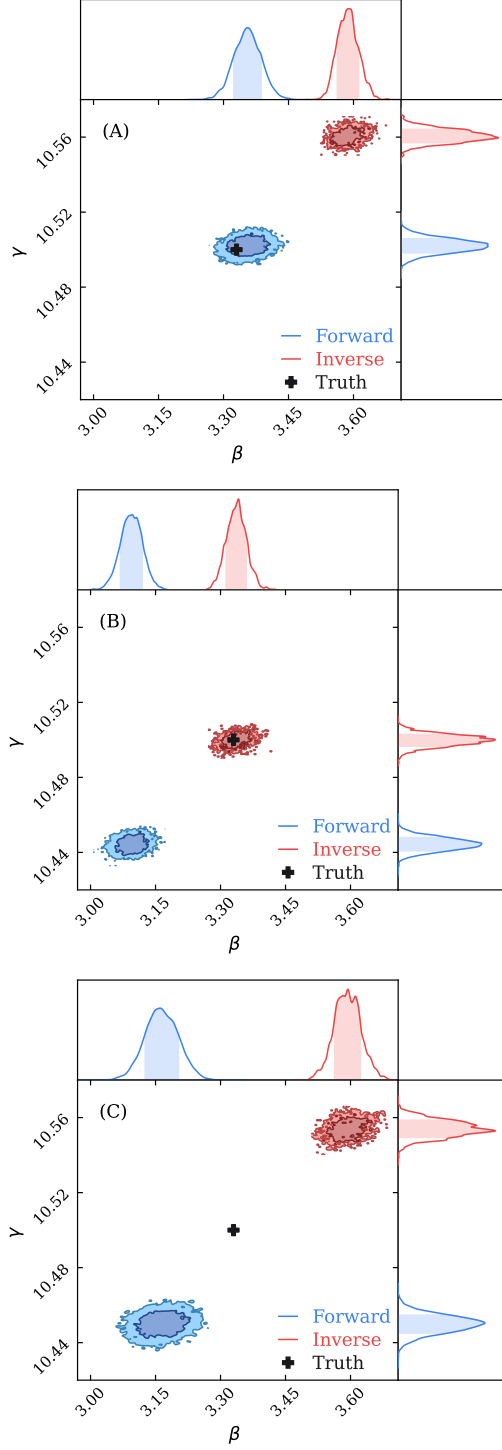
### 6.3. Code Validation and the General Eddington Bias

According to the theory of the general Eddington bias in § 3, parameter estimation from unidirectional models would be unbiased only when there is no scatter in the independent variable; otherwise, the estimated parameters will be biased. The testing in this subsection thus serves two purposes: (1) to validate the implementation of the likelihood functions in the code, and more important, (2) to quantify the level of Eddington biases when there is scatter in the independent variable.

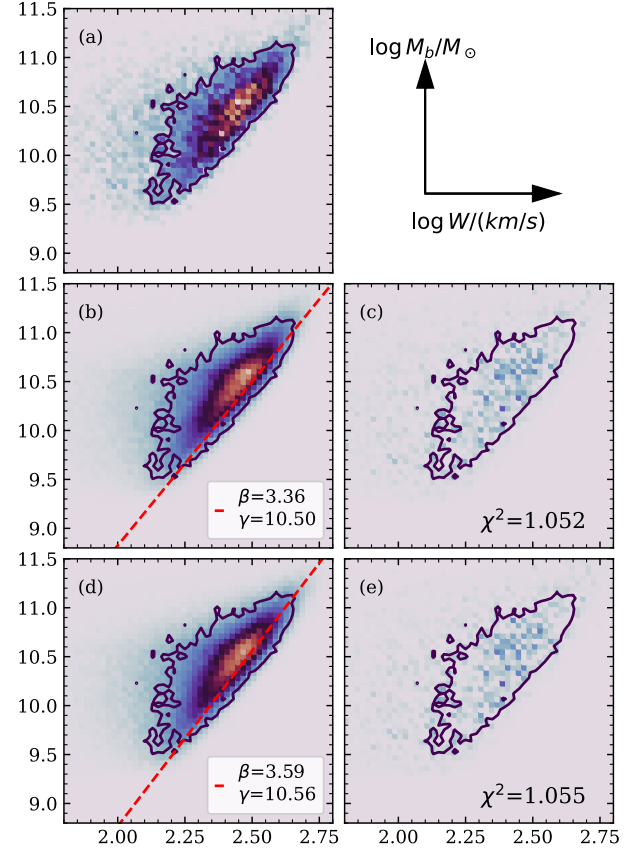
Assuming bounded flat priors for all parameters, I run the forward and the inverse models on the three simulated samples and the inferred parameters are compared to the input parameters in Table 1. The MCMC-sampled posterior pdfs of the two key parameters (TF slope  $\beta$  and intercept  $\gamma$ ) are presented in Fig. 2. As expected, the forward model recovers the truth using Sample A, and the inverse model recovers the truth using Sample B. These findings provide confidence in the implemented likelihood functions of the two models. On the other hand, significant biases are observed when the forward model is used on Samples B and C (where  $\sigma_w \neq 0$ ), and when the inverse model is used on Samples A and C (where  $\sigma_m \neq 0$ ).

The directions of the biases are that the forward model underestimate both parameters while the inverse model overestimate them. The amount of biases ( $B_\gamma \simeq \pm 0.06, B_\beta \simeq \pm 0.24$ ) are significantly greater than the statistical uncertainties shown in the marginalized pdfs ( $\sigma_\gamma \simeq 0.004, \sigma_\beta \simeq 0.03$ ). Further experiments show that both biases increase almost linearly with the variance of the independent variable (see § 7.3).

It is worth noting that the same biases are present when inclination-corrected data are used; i.e., the bias is not introduced by treating inclination as a latent variable. Instead, it is a feature of unidirectional regression models. As mentioned earlier, it is well known that the OLS estimate of the regression slope is biased to zero when the independent variable is measured with error (e.g., Fuller 1987; Akritas & Bershadsky 1996). Since least-squares is a maximum likelihood estimator, the same biases are expected in unidirectional MLE methods that neglect the error of the independent variable,



**Figure 2.** MCMC-sampled posterior pdfs from the forward (blue) and the inverse (red) models for samples A, B, and C, respectively. For simplicity, only the two TF parameters are shown: slope ( $\beta$ ) and intercept ( $\gamma$ ). In all posterior plots, the contours in the joint pdfs enclose 68% and 95% of the volume, the highlighted regions in the marginalized pdfs show the credible intervals defined by the 16th and the 84th percentiles, and the black crosses indicate the truth. Biases appear whenever the data contains scattering in the model-assumed independent variable.



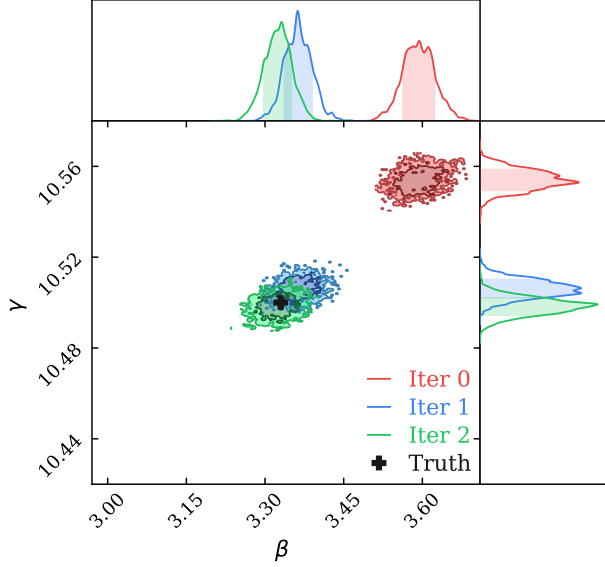
**Figure 3.** The distribution of mass vs. projected line width of the input data (a) is compared with the distributions of larger samples simulated using the best-fit parameters of the forward model (b) and the inverse model (d). The contours enclose bins containing a minimum of 9 objects in panel a. In this example, the input data are from sample A, so only the forward model inferred the correct TF parameters. However, both models fit the distribution of the data equally well according to the residuals in panels c and e. This result shows that it is difficult to use a goodness-of-fit parameter such as the reduced  $\chi^2$  to select the correct model.

regardless whether the data are corrected for inclination or not.

Despite of the biased parameter estimation, both models fit the input data equally well for all three input samples. I demonstrate this with sample A in Fig. 3, where simulated samples  $20\times$  larger than the input sample are generated using the best-fit parameters of the two models<sup>8</sup>, and their distributions are compared with that of the input sample. The model residuals show statistical fluctuations comparable to the Poisson noise, as indicated by the reduced  $\chi^2$  values around unity. The slight difference in the  $\chi^2$  values between the two models is actually due to the stochastic nature of the simulated

<sup>8</sup> Unconstrained  $\sigma$  is set to zero and unconstrained Schechter function parameters are set to the truth values.





**Figure 4.** MCMC-sampled posterior pdfs from the inverse model after iterative corrections of the general Eddington bias. Sample C is used as the input data. Iteration 0 is before any bias correction, so it is the same as the posterior of the inverse model in Fig. 2c. In subsequent iterations,  $\{\tilde{w}_k\}_{k=1}^N$  are bias corrected using Eq. 51. The process converges after just a few iterations, but unbiased parameter inference is achieved only when the correct  $\sigma_m$  is specified.

samples. This finding shows that the goodness-of-fit cannot be used to tell which model is better supported by the data. Instead of model selection, one should focus on methods that can mitigate the general Eddington bias, which are the topics of the next section.

## 7. UNBIASED PARAMETER INFERENCE

This section introduces two latent-inclination methods that can mitigate both Malmquist bias and Eddington bias in the inference of the TFR. The first method shifts the dependent variable using the predicted Eddington bias from its analytical expression (§ 7.1), and the second method incorporates the scatter of the independent variable in the likelihood function of a bidirectional dual-scatter model (§ 7.2). Technically, there is a third method—choosing an empirical unbiased anchor point for the correlation (§ 7.3)—but it only reduces the bias of the intercept parameter, limiting its application.

### 7.1. Shifting the Moment of the Dependent Variable

In § 3, I have explained that the general Eddington bias is caused by the scatter in the independent variable and the gradient of its distribution function. It is thus expected that, when the predicted Eddington bias of the dependent variable is corrected for using its analytical expression, the inferred regression coefficients will be no longer biased.

For calibrating the TFR using inclination-corrected data, one could choose to debias the dependent variable in either

the forward model or the inverse model. But when the inclination angle is considered a latent variable, one can only correct  $\tilde{w}$  for the inverse model. This is because to debias  $\tilde{m}$  for the forward model using Eq. 15 requires knowing the inclination-corrected velocity width ( $x \equiv \tilde{w} - i$ ).

For the forward model, the general Eddington bias of the mean  $y$  at a fixed  $\tilde{x}$  is given by Eq. 15 in § 3:

$$y(\tilde{x}) - \langle y \rangle_{\tilde{x}} = -\beta \sigma_x^2 \frac{d \ln p(\tilde{x})}{d \tilde{x}} \quad (45)$$

For the inverse model, the general Eddington bias of the mean  $x$  at a fixed  $\tilde{y}$  is:

$$x(\tilde{y}) - \langle x \rangle_{\tilde{y}} = -\frac{\sigma_y^2}{\beta} \frac{d \ln p(\tilde{y})}{d \tilde{y}} \quad (46)$$

To remove the bias, one can shift the value of individual  $x$ :

$$x_c = x - \frac{\sigma_y^2}{\beta} \frac{d \ln p(\tilde{y})}{d \tilde{y}} \quad (47)$$

so that the mean of  $x_c$  matches the  $\tilde{y}$ -predicted  $x$ :

$$x(\tilde{y}) - \langle x_c \rangle_{\tilde{y}} = x(\tilde{y}) - \langle x \rangle_{\tilde{y}} + \frac{\sigma_y^2}{\beta} \frac{d \ln p(\tilde{y})}{d \tilde{y}} = 0 \quad (48)$$

If  $\tilde{y}$  follows a Schechter function:

$$p(\tilde{y}) \propto 10^{(\alpha+1)(\tilde{y}-y_*)} \exp(-10^{\tilde{y}-y_*}) \quad (49)$$

then the Eddington bias correction in  $x$  is:

$$x_c = x - \frac{\ln 10}{\beta} \sigma_y^2 [(\alpha+1) - 10^{\tilde{y}-y_*}] \quad (50)$$

For the case of the inverse TFR, the bias correction in  $\tilde{w}$  is:

$$\tilde{w}_c = \tilde{w} - \frac{\ln 10}{\beta} \sigma_m^2 [(\alpha+1) - 10^{(\tilde{m}+d)-M_*}] \quad (51)$$

which shows that the correction depends on four parameters ( $\sigma_m, \beta, M_*, \alpha$ ) and the observed mass of the galaxy ( $\tilde{m} + d$ ). The success of the bias correction depends on the accurate specification of the four parameters.

First,  $M_*$  and  $\alpha$  characterize the observed mass function, which can be estimated through various methods. For example, it can be measured using luminosity-function estimators that correct for the Malmquist bias, e.g., the  $C^-$  method (Lynden-Bell 1971), followed by a Schechter function fit. Here I opt to infer  $v_*$  and  $\alpha$  using the latent-inclination forward model introduced in § 5.1, and calculate  $M_*$  as  $v_* + \gamma$ . It is safe to use  $v_*$  and  $\alpha$  from the forward model because (1) they are mildly biased (unlike  $\beta, \gamma$ ), i.e., their deviations from the truth are within twice the statistical error (see the blue curves in the third and fourth columns of Fig. 5), and (2) the Eddington bias correction is relatively insensitive to these parameters.

Next,  $\beta$  is the slope of the TFR, which is a key parameter to be determined by the model. Using it as an input parameter thus requires an iterative process. It starts from the results of the forward model and the inverse model without Eddington bias correction. The former provides  $v_*$ ,  $\alpha$ , and the upper limit on  $\sigma_m$ , the latter provides  $\beta_j$  and  $\gamma_j$  for  $j = 0$ , where the subscripts indicate that they are the initial guess values and will be updated in subsequent iterations. For a specified value of  $\sigma_m$ , the line widths are then corrected for the Eddington bias using Eq. 51 with parameters  $(\sigma_m, \beta_j, v_* + \gamma_j, \alpha)$ . The subsequent runs of the inverse model uses Eddington-bias-corrected line widths ( $\tilde{w}_c$ ) as the input data and produce updated values of  $\beta_j$  and  $\gamma_j$ . Fig. 4 illustrates the iterative process and shows that results converge in just a few iterations.

Finally,  $\sigma_m$  is the mass scatter that includes both measurement error and intrinsic dispersion. This is the only parameter that needs to be specified by the user based on prior knowledge of the data. In the worse case, one can use the forward model to estimate the total dispersion of the data in the  $m$ -direction ( $\sigma_m^t$ ), which gives the upper limit of  $\sigma_m$ :  $0 < \sigma_m < \sigma_m^t$ . The bias-corrected results should lie somewhere between the two extremes defined by the forward and the inverse models without corrections (e.g., see Fig. 2 for Sample C). Underestimating  $\sigma_m$  causes under-correction of the Eddington bias, so the inferred  $\beta$  and  $\gamma$  will be biased high (although their biases are reduced compared to those from the uncorrected model). On the other hand, overestimating  $\sigma_m$  causes over-corrections of the Eddington bias, so the TF parameters overshoot the truth and will be biased low as they approaches those inferred from the forward model. Therefore, the primary limitation of this method is that the inferred parameters will be unbiased only when the correct value of  $\sigma_m$  is prescribed by the user.

## 7.2. The Bidirectional Dual-Scatter Model

In this subsection, I introduce the dual-scatter model where I expand the likelihood function of the latent-inclination forward model by including scatters of the independent variable  $w$ . In Appendix B, I expand the likelihood function of the inverse model by including scatters in  $m$ . It is shown there that the results are mathematically equivalent, making the model bidirectional (symmetric).

When  $\tilde{w}$  and  $\tilde{m}$  are independent<sup>9</sup>, the joint probability of the dual-scatter model can be decomposed as:

$$\begin{aligned} p(\tilde{m}, \tilde{w}, d, w, i) &= p(\tilde{m}, \tilde{w}, d|w, i)p(w, i) \\ &= p(\tilde{m}|d, w, i)p(\tilde{w}|w)p(d) \times p(w|i)p(i) \\ &= p(\tilde{w}|w)p(\tilde{m}, w, d, i) \end{aligned} \quad (52)$$

<sup>9</sup> This is a valid assumption because correlated measurement errors in TFR data are introduced by inclination correction.

where  $p(\tilde{m}, w, d, i)$  is given by Eq. 28, and the new term,  $p(\tilde{w}|w)$ , is a Gaussian:

$$p(\tilde{w}|w) = \frac{1}{\sigma_w \sqrt{2\pi}} \exp \left[ -\frac{(\tilde{w} - w)^2}{2\sigma_w^2} \right] \quad (53)$$

The joint pdf of the observables  $(\tilde{m}, \tilde{w}, d)$  is then calculated by marginalizing Eq. 52 over both  $w$  and  $i$ :

$$\begin{aligned} p(\tilde{m}, \tilde{w}, d) &= \int_{-\infty}^{\infty} \int_{-\infty}^0 p(\tilde{m}, \tilde{w}, d, w, i) di dw \\ &= \frac{\beta \phi_* p(d)}{2\pi \sigma_m \sigma_w} \int_{-\infty}^{\infty} \int_{-\infty}^0 \frac{\ln 10 \cdot 10^{2i}}{\sqrt{1 - 10^{2i}}} \\ &\quad \times \exp \left[ -\frac{(\tilde{w} - w)^2}{2\sigma_w^2} \right] \exp \left[ -\frac{(\tilde{m} + d - \beta(w - i) - \gamma)^2}{2\sigma_m^2} \right] \\ &\quad \times 10^{(\alpha+1)[\beta(w-i)-v_*]} \exp(-10^{\beta(w-i)-v_*}) di dw \end{aligned} \quad (54)$$

Accounting for the selection function  $S(\tilde{m}, \tilde{w})$ , the conditional pdf of  $\tilde{m}$  is:

$$p(\tilde{m}|\tilde{w}, d) = \frac{S(\tilde{m}, \tilde{w})p(\tilde{m}, \tilde{w}, d)}{\int_{-\infty}^{\infty} S(\tilde{m}, \tilde{w})p(\tilde{m}, \tilde{w}, d) d\tilde{m}} \quad (55)$$

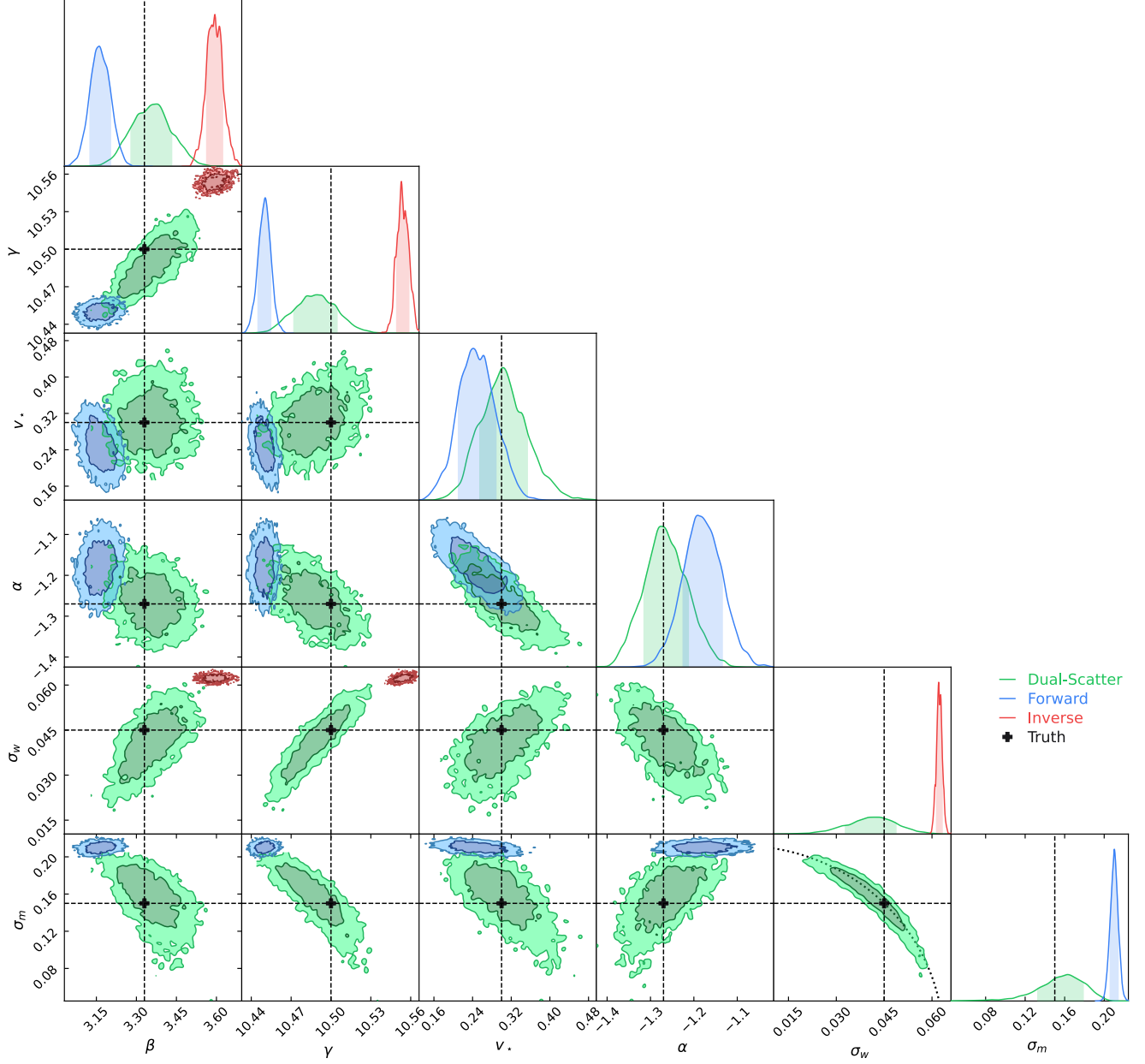
And the data likelihood function is composed of the above conditional pdfs:

$$\ln \mathcal{L} = \sum_{k=1}^N \ln p(\tilde{m}_k|\tilde{w}_k, d_k) \quad (56)$$

Like the forward model, when the selection function is a step function of  $\tilde{m}$ , the integral over  $\tilde{m}$  on the denominator of Eq. 55 becomes an error function. This replacement reduces the triple integral to a double integral. I provide the full expression of the conditional pdf in Eq. A4 in Appendix A.

Note that even after this simplification, evaluating the likelihood function remains computationally expensive because each data point still requires calculating two double integrals. In Appendix C, I describe a Fast Fourier Transform (FFT) method that drastically accelerated the computation of the double integrals. Additional acceleration is achieved by vectorizing the conditional pdf and leveraging the integrated Graphics Processing Unit (GPU). Together, the computation time is reduced by three orders of magnitude when compared to direct integration, making the dual-scatter model even faster than the forward model, when the latter is run on the CPU.

The dual-scatter model is applied to the simulated Sample C, which includes significant amounts of scatter in both axes. I show the MCMC-sampled posterior pdfs in Fig. 5 and compare them with those from the forward and the inverse models. The results from the dual-scatter model can be summarized as follows:



**Figure 5.** MCMC-sampled posterior pdfs from the dual-scatter model (*green*) compared to those from the forward model (*blue*) and the inverse model (*red*). The data are from sample C, whose input parameters are indicated by black dashed lines and black crosses. The dual-scatter model produces relatively unbiased inferences for all parameters, gives more realistic statistical uncertainties, and reveals parameter degeneracies. For example, the two  $\sigma$ 's are tightly correlated on an ellipse:  $\sigma_m^2 + \beta^2 \sigma_w^2 = \sigma_t^2$ , as indicated by the dotted curve in the panel.

- The marginalized posteriors show that the best-fit (median) parameters are relatively unbiased for all three samples, i.e., they lie within  $1\sigma$  from the truth.
- The statistical uncertainties of the inferred parameters are larger compared to those from the forward and the inverse model: the TF parameters ( $\beta, \gamma$ ) show  $2\text{--}3\times$  larger uncertainties, the dispersion parameters ( $\sigma_m, \sigma_w$ ) show  $2\text{--}8\times$  greater uncertainties, while the Schechter parameters ( $v_*, \alpha$ ) show comparable uncertainties.
- The joint posteriors show significant correlations among parameters. In particular, the two dispersion parameters are tightly correlated on an ellipse:  $\sigma_m^2 + \beta^2 \sigma_w^2 = \sigma_t^2$ . These covariances explain the larger uncertainties of the inferred parameters.

Compared to the forward model, including the scatter of the independent variable  $w$  requires multiplication of another Gaussian function of  $\tilde{w} - w$  and an additional marginalization over  $w$  in both the numerator and the denominator of the

conditional pdf that constitutes the likelihood function. The added complexity makes the MCMC sampling significantly more computationally expensive, motivating innovative applications of numerical methods and GPU acceleration. The resulting posteriors show nearly unbiased parameter estimation, more realistic estimates of the uncertainties, and strong degeneracy between the two dispersion parameters. The degeneracy is a feature of the mathematical problem, so it cannot be eliminated, but its effects decrease as the sample size increases.

### 7.3. The Unbiased Anchor Point

The TFR is typically anchored at  $\log V_0 = 2.5$ , which is an arbitrary choice. In this subsection, I describe the existence of a preferred anchor point at which the inferred intercepts of both the forward and the inverse models are unbiased.

By simulating a number of data sets that share the same TFR ( $\beta = 3.33, \gamma = 10.5$ ) and observational selection function but have different amount of scatters in the independent variable ( $w$ ), I find the following power-law relations between the biases of the regression coefficients and the amount of scatter ( $\sigma_w$ ):

$$\begin{aligned} B_\gamma &= \begin{cases} -0.058(\sigma_w/0.05)^{1.8} & \text{forward model} \\ +0.060(\sigma_m/0.15)^{1.8} & \text{inverse model} \end{cases} \\ B_\beta &= \begin{cases} -0.246(\sigma_w/0.05)^{1.8} & \text{forward model} \\ +0.237(\sigma_m/0.15)^{1.8} & \text{inverse model} \end{cases} \end{aligned} \quad (57)$$

Notice that because the biases of both coefficients scale with  $\sigma_w$  following power laws of the same index, their ratios are independent of the scatter. A consequence of the constant bias ratio  $B_\gamma/B_\beta$  is that the true TFR and the inferred TFRs from the two models all converge at a single point at:

$$\begin{aligned} \log V_0 &= 2.5 - \frac{B_\gamma}{B_\beta} \\ \gamma_0 &= \gamma - \beta \frac{B_\gamma}{B_\beta} \end{aligned} \quad (58)$$

At this converging point, the intercepts of all the inferred TFRs are unbiased. This finding suggests that one can reduce the general Eddington bias by choosing the unbiased anchor location instead of the default anchor at  $\log V_0 = 2.5$ .

This unbiased anchor location depends on sample characteristics and the coefficients of the TFR, but it can be determined empirically by running the forward model and the inverse model on the same data. Take Sample C as an example, I use the inferred TF parameters listed in the last column of Table 1. I first calculate the differences between the inferred parameters of the two models,  $\Delta B_\gamma = 10.554 - 10.450 = 0.104$ ,  $\Delta B_\beta = 3.593 - 3.164 = 0.429$ , then their ratio,  $\Delta B_\gamma/\Delta B_\beta = 0.24$ , and finally the unbiased anchor point

at  $\log V_0 = 2.5 - \Delta B_\gamma/\Delta B_\beta = 2.26$ . At this anchor point, the intercepts of both the forward model and the inverse model are at  $\gamma_0 = 9.69$ , which is essentially the same as the intercept of the true relation at  $\log V_0 = 2.26$  ( $\gamma_0 = 9.70$ ). Despite of their different slopes, adopting the converging point as the anchor has minimized the bias of the inferred intercepts for both the forward and the inverse models.

## 8. SUMMARY AND DISCUSSION

Linear regression of the TFR requires the observed data to be corrected for inclination. But the current method to estimate inclination angle from galaxy morphology is highly uncertain and requires well-resolved images. One solution is to treat the inclination as a latent variable with a known distribution function, so that one can infer the TFR using data uncorrected for inclinations. In this paper, I have developed Bayesian latent-inclination inference methods that can also mitigate the distance-dependent Malmquist bias due to sample selection and the general Eddington bias due to scatter in the independent variable. These biases are important because they directly impact measurements of the Hubble constant.

I start by deriving the likelihood functions for two latent-inclination unidirectional models that neglect the scatter of the independent variable: the forward model (velocity width as independent variable) and the inverse model (mass as independent variable). Both likelihood functions are implemented in Python with a MCMC-sampler and are tested on simulated galaxy samples with random orientations and with random Gaussian scatters in mass and velocity. I find that, when the independent variable has significant scatters, the inferred regression coefficients are biased (Fig. 2), although biased models demonstrate similarly good quality of fit to the data (Fig. 3). The biases of both coefficients increase with the scatter of the independent variable  $B_\theta \propto \sigma_x^{1.8}$ . For example, using the inverse TFR model, neglecting a scatter of 0.15 dex in luminosity would lead to an overestimate of  $H_0$  by  $\sim 5 \text{ km s}^{-1} \text{ Mpc}^{-1}$  (Eq. 8), comparable to the ‘‘Hubble tension’’.

Evidently, the scatter of the independent variable ( $x$ ) does not simply propagate into the dependent variable ( $y$ ) through the correlation as  $\sigma_y = \beta\sigma_x$ , because that would not have led to biases in the regression coefficients. Instead, the bias shifts the first moment of  $y$  at each  $x$  by altering the conditional pdf  $p(y|x)$  according to Bayes’ rule. I derive the analytical form of this bias (Eq. 15). Because the classic Eddington bias is a special case of this bias, it is termed the ‘‘general Eddington bias’’.

Methods aimed at unbiased parameter inference must include both (1) the scatter of the dependent variable and the selection function to account for the distance-dependent Malmquist bias, and (2) the scatter and the distribution function of the independent variable to account for the general



Eddington bias. I then introduce two approaches to mitigate the general Eddington bias:

- Shifting the dependent variable by reversing the expected amount of Eddington bias (*Moment Shifting*, § 7.1),
- Incorporating the scatter of the independent variable in the likelihood function (*Dual-scatter Model*, § 7.2).

Testing on simulated data sets shows that both methods can effectively reduce or eliminate both the Malmquist bias and the Eddington bias (Figs. 4 and 5). Each method has its own strengths and limitations, and it is a trade-off between efficiency and accuracy:

- The moment-shifting method can iteratively determine most of the required parameters for bias correction, but the dispersion parameter of the independent variable must be specified *a priori*.
- The dual-scatter model is the most versatile and is recommended for most scenarios. It fits both dispersion parameters simultaneously, produces more realistic estimates of parameter uncertainties, and reveals covariances among model parameters. But it is the most computationally expensive method, requiring FFT techniques and GPUs to accelerate (Appendix C).

Technically, there is one additional method to mitigate the general Eddington bias: defining the intercept of the TFR at an unbiased anchor point (§ 7.3). The “unbiased anchor” method takes advantage of the correlation between the biases of the regression coefficients and determines the unbiased anchor empirically. The inferred intercepts at this preferred anchor are unbiased for both unidirectional models. However, the inferred slopes remain biased, and the location of the unbiased anchor depends on sample characteristics. This variability complicates applying the method to Hubble constant measurements, as the unbiased anchors for the redshift and zero-point samples are likely distinct.

Naturally, the next step is to apply the latent-inclination likelihood-based methods on actual data. Currently there are two major data sets for TF studies in the nearby Universe ( $cz < 20,000 \text{ km s}^{-1}$ ): the ALFALFA sample of  $\sim 31,500$  galaxies<sup>10</sup> (Haynes et al. 2018; Durbala et al. 2020) and the CF4 BTF-Distances catalog of  $\sim 10,000$  galaxies<sup>11</sup> (Kourkchi et al. 2022). The ALFALFA sample is better suited for the latent-inclination methods because it is a  $7000 \text{ deg}^2$  H I survey from a single instrument (Arecibo

ALFA) and it includes galaxies at all inclinations. In contrast, the CF4 sample is a compilation of H I surveys from seven radio telescopes and excludes galaxies with estimated inclination less than  $45^\circ$ . Some of the compiled subsamples in CF4 were selected to be edge-on systems, causing a spike near  $90^\circ$  in the distribution of measured inclinations. The ALFALFA sample is also preferred because of selection bias: a single observational selection function that describes the full data set can be used in the likelihood function. For a heterogeneous sample like the CF4, the complex selection effects from various instruments are unlikely to be captured by a single post facto selection function. Instead, each constituent subsample should be separately analyzed using their own selection function, and the final result would be a weighted average of the results from the subsamples.

Future applications of the methods on the ALFALFA data also face challenges. First, the H I flux limit increases with line width. Like all spectroscopic line surveys with fixed integration time per area, ALFALFA is more sensitive to narrower lines because the same amount of signal is spread over fewer spectroscopic channels. The line flux limits set gas mass limits, making the latter also depending on the line width. Second, both the total H I flux and the optical photometry become unreliable for the nearest galaxies. The former has been addressed by the special catalog for extended sources (Hoffman et al. 2019). But to address the latter, one needs to replace the automatic photometry from pipelines with more involved asymptotic photometry (Courtois et al. 2011) for a large sample of galaxies. Because of the inclination cutoff at  $45^\circ$ , only 43% (13,617) of the ALFALFA sample are included in the CF4 Initial Candidates catalog, leaving 57% of the sample without asymptotic photometry.

Thus far, the discussion has centered on the linear baryonic TFR. Depending on the available data, researchers may prefer a different form of the TFR. Fortunately, the expressions derived for the baryonic TFR can be readily adapted to other empirical forms of the TFR:

- To apply the methods to magnitude-based TFR, only a simple modification is required: compute  $m$  as  $-0.4$  times the apparent magnitude at wavelength  $\lambda$  (i.e.,  $m \equiv -0.4m_\lambda$ ) to convert the decreasing magnitude scale defined by Pogson’s ratio to an increasing decade scale. With this conversion, the best-fit intercept parameter  $\gamma$  is related to the fiducial absolute magnitude  $M_\lambda^0$  as:  $M_\lambda^0 = -2.5\gamma - 25$ .
- To apply the methods to non-linear TFR (e.g., to capture the curvature of the  $i$ -band TFR at the luminous end), one can replace the linear relation in Eq. 6 with more complex functions of  $(w-i)$  (e.g., a polynomial).

I have made the following simplifications to keep the discussion focused on the Sine scatter of the velocity width

<sup>10</sup> The ALFALFA sample contains 25,432 high S/N H I detections (Code 1) and 6,068 low S/N detections with prior optical detection (Code 2).

<sup>11</sup> The CF4 catalogs are available in the Extragalactic Distance Database: <https://edd.ifa.hawaii.edu>.

and the Gaussian scatters of logarithmic mass and velocity width: (1) the redshifts are assumed to have been corrected for peculiar velocities, (2) the residual velocity noise is ignored, and (3) the apparent baryonic mass is assumed to be unaffected by galaxy inclination (i.e., either gas mass dominates or stellar mass extinction is accounted for by the color-derived mass-to-light ratio). Below I describe future extensions of the model to relax these assumptions.

Incorporating velocity noise in the likelihood function requires multiplying a Gaussian function of  $\tilde{d}-d$  to the joint pdf and an additional marginalization over the true distance  $d$  (§ 5.1). In addition, models that simultaneously fit the scale parameters of a peculiar velocity model and the velocity noise have been developed previously for TFR studies using inclination-corrected data (e.g., Willick et al. 1997; Boubel et al. 2024). The same methodology can be applied to the latent-inclination models.

For magnitude-based TFRs, to account for the inclination-dependent internal extinction in the forward and the dual-scatter model, one needs to replace  $\tilde{m}$  with its extinction-corrected counterpart  $\tilde{m}-0.4A(\text{inc})$  in  $p(\tilde{m}|w,i,d)$  (Eq. 21). A common way to parameterize the extinction is  $A(\text{inc}) = \tau \log[\sec(\text{inc})]$  (e.g., Shao et al. 2007). The extinction-corrected mass is then  $\tilde{m}-0.2\tau \log(1-10^{2i})$  (where  $i \equiv \log[\sin(\text{inc})]$  as defined in Eq. 5), and the conditional pdf becomes:

$$p(\tilde{m}|w,i,d) = \frac{1}{\sigma_m \sqrt{2\pi}} \times \exp \left[ -\frac{[\tilde{m}-0.2\tau \log(1-10^{2i})+d-\beta(w-i)-\gamma]^2}{2\sigma_m^2} \right] \quad (59)$$

The replacement adds  $\tau$  (the face-on extinction) as a new parameter and changes the subsequent calculation of the conditional pdf  $p(\tilde{m}|w,d)$  in Eq. 30 and  $p(\tilde{m}|\tilde{w},d)$  in Eq. 55.

To account for the inclination-dependent extinction in the inverse model, one needs to replace  $m$  with  $m-0.2\tau \log(1-10^{2i})$  in both  $p(\tilde{w}|m,d,i)$  and  $p(m|d)$  (Eqs. 34 and 36). This modification causes the mass function term,  $p(m|d)$ , to depend on inclination, preventing it from being cancelled out in the conditional pdf in Eq. 39.

Latent-variable Bayesian methods provide wide applicability in observational astronomy. Although this paper focuses on inferring the Tully-Fisher relation, the likelihood-based framework is highly generalizable and can be adapted to other linear regression problems involving Gaussian and non-Gaussian scatter. The primary requirement is knowledge of the latent variable's prior probability distribution, which can be expressed analytically or provided numerically as a tabulated function.

## ACKNOWLEDGMENTS

I appreciate helpful discussions with my colleagues Ken Gayley, Kevin Hall, and Steve Spangler. This work is supported by the University of Iowa through the Investment in Strategic Priorities (ISP) program and the National Science Foundation (NSF) grant AST-2103251.

*Software:* NumPy (Harris et al. 2020), SciPy (Virtanen et al. 2020), emcee (Foreman-Mackey et al. 2013), PyTorch (Paszke et al. 2019), ChainConsumer (Hinton 2016)

## REFERENCES

- Akritas, M. G., & Bershadsky, M. A. 1996, *ApJ*, 470, 706
- Bartlett, D. J., & Desmond, H. 2023, *The Open Journal of Astrophysics*, 6, 42
- Bottinelli, L., Gouguenheim, L., Paturel, G., & Teerikorpi, P. 1986, *A&A*, 166, 393
- . 1988, *ApJ*, 328, 4
- Boubel, P., Colless, M., Said, K., & Staveley-Smith, L. 2024, *MNRAS*, 531, 84
- Bradford, J. D., Geha, M. C., & van den Bosch, F. C. 2016, *ApJ*, 832, 11
- Butkevich, A. G., Berdyugin, A. V., & Teerikorpi, P. 2005, *MNRAS*, 362, 321
- Caldwell, R. R., & Kamionkowski, M. 2004, *JCAP*, 2004, 009
- Courtois, H. M., Tully, R. B., & Héraudeau, P. 2011, *MNRAS*, 415, 1935
- Durbala, A., Finn, R. A., Crone Odekon, M., et al. 2020, *AJ*, 160, 271
- Dyson, Frank, S. 1926, *MNRAS*, 86, 686
- Eddington, A. S. 1914, *Stellar movements and the structure of the universe*, Chapter 8, Page 172 (London, Macmillan and co., limited)
- Eddington, A. S., S. 1940, *MNRAS*, 100, 354
- Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, *PASP*, 125, 306
- Fu, H. 2024, *ApJL*, 963, L19
- Fuller, W. A. 1987, *Measurement error models* (Wiley Series in Probability and Mathematical Statistics, New York: Wiley)
- Giraud, E. 1987, *A&A*, 174, 23
- Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, *Nature*, 585, 357
- Haynes, M. P., Giovanelli, R., Kent, B. R., et al. 2018, *ApJ*, 861, 49
- Hinton, S. R. 2016, *The Journal of Open Source Software*, 1, 00045
- Hoffman, G. L., Dowell, J., Haynes, M. P., & Giovanelli, R. 2019, *AJ*, 157, 194

- Kelly, B. C. 2007, *ApJ*, 665, 1489
- Kourkchi, E., Tully, R. B., Courtois, H. M., Dupuy, A., & Guinet, D. 2022, *MNRAS*, 511, 6160
- Lynden-Bell, D. 1971, *MNRAS*, 155, 95
- Malmquist, K. G. 1922, *Meddelanden fran Lunds Astronomiska Observatorium Serie I*, 100, 1
- McGaugh, S. S., Schombert, J. M., Bothun, G. D., & de Blok, W. J. G. 2000, *ApJL*, 533, L99
- Obreschkow, D., & Meyer, M. 2013, *ApJ*, 777, 140
- Papastergis, E., Cattaneo, A., Huang, S., Giovanelli, R., & Haynes, M. P. 2012, *ApJ*, 759, 138
- Paszke, A., Gross, S., Massa, F., et al. 2019, *arXiv e-prints*, [arXiv:1912.01703](https://arxiv.org/abs/1912.01703)
- Phillips, M. M. 1993, *ApJL*, 413, L105
- Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020, *A&A*, 641, A6
- Riess, A. G., Anand, G. S., Yuan, W., et al. 2024, *ApJL*, 962, L17
- Sandage, A. 1994a, *ApJ*, 430, 1
- . 1994b, *ApJ*, 430, 13
- Sandage, A., & Tammann, G. A. 1976, *ApJ*, 210, 7
- Shao, Z., Xiao, Q., Shen, S., et al. 2007, *ApJ*, 659, 1159
- Tully, R. B. 1988, *Nature*, 334, 209
- Tully, R. B., & Fisher, J. R. 1977, *A&A*, 54, 661
- Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, *Nature Methods*, 17, 261
- Westmeier, T., Deg, N., Spekkens, K., et al. 2022, *PASA*, 39, e058
- Willick, J. A. 1994, *ApJS*, 92, 1
- Willick, J. A., Strauss, M. A., Dekel, A., & Kolatt, T. 1997, *ApJ*, 486, 629
- Willmer, C. N. A. 2018, *ApJS*, 236, 47

## APPENDIX

## A. THE CONDITIONAL PROBABILITIES OF FLUX-LIMITED SAMPLES

In this Appendix, I present the complete analytical expressions for the conditional PDFs of the forward, inverse, and dual-scatter models for a common scenario where the sample is flux-limited and the selection function is a step function:

$$S(\tilde{m}) = \begin{cases} 1, & \text{if } \tilde{m} \geq m_l \\ 0, & \text{otherwise} \end{cases} \quad (\text{A1})$$

This particular selection function significantly simplifies the marginalization of the joint pdf over  $\tilde{m}$ , because the only  $\tilde{m}$ -dependent term is a Gaussian, and its integral is the complementary error function (erfc):

$$\begin{aligned} \int_{-\infty}^{\infty} S(\tilde{m}) p(\tilde{m}|w, i, d) d\tilde{m} &= \int_{m_l}^{\infty} p(\tilde{m}|w, i, d) d\tilde{m} \\ &= \int_{m_l}^{\infty} \frac{1}{\sigma_m \sqrt{2\pi}} \exp \left[ -\frac{[\tilde{m} + d - \beta(w-i) - \gamma]^2}{2\sigma_m^2} \right] d\tilde{m} \\ &= \frac{1}{2} \text{erfc} \left[ \frac{m_l + d - \beta(w-i) - \gamma}{\sqrt{2}\sigma_m} \right] \end{aligned} \quad (\text{A2})$$

After this simplification, the conditional pdfs for the forward and the dual-scatter models become:

- The forward model:

$$p(\tilde{m}|w, d) = \frac{\sqrt{2}}{\sqrt{\pi}\sigma_m} \frac{\int_{-\infty}^0 \frac{10^{2i}}{\sqrt{1-10^{2i}}} \exp \left[ -\frac{[\tilde{m} + d - \beta(w-i) - \gamma]^2}{2\sigma_m^2} \right] 10^{(\alpha+1)[\beta(w-i)-v_*]} \exp(-10^{\beta(w-i)-v_*}) di}{\int_{-\infty}^0 \frac{10^{2i}}{\sqrt{1-10^{2i}}} \text{erfc} \left[ \frac{m_l + d - \beta(w-i) - \gamma}{\sqrt{2}\sigma_m} \right] 10^{(\alpha+1)[\beta(w-i)-v_*]} \exp(-10^{\beta(w-i)-v_*}) di} \quad (\text{A3})$$

- The dual-scatter model:

$$p(\tilde{m}|\tilde{w}, d) = \frac{\sqrt{2}}{\sqrt{\pi}\sigma_m} \frac{\int_{-\infty}^0 \frac{10^{2i}}{\sqrt{1-10^{2i}}} \left( \int_{-\infty}^{\infty} \exp \left[ -\frac{(w-\tilde{w})^2}{2\sigma_w^2} \right] \exp \left[ -\frac{[\tilde{m} + d - \beta(w-i) - \gamma]^2}{2\sigma_m^2} \right] 10^{(\alpha+1)[\beta(w-i)-v_*]} \exp(-10^{\beta(w-i)-v_*}) dw \right) di}{\int_{-\infty}^0 \frac{10^{2i}}{\sqrt{1-10^{2i}}} \left( \int_{-\infty}^{\infty} \exp \left[ -\frac{(w-\tilde{w})^2}{2\sigma_w^2} \right] \text{erfc} \left( \frac{m_l + d - \beta(w-i) - \gamma}{\sqrt{2}\sigma_m} \right) 10^{(\alpha+1)[\beta(w-i)-v_*]} \exp(-10^{\beta(w-i)-v_*}) dw \right) di} \quad (\text{A4})$$

For the inverse model, simplifying the conditional pdf does not require a step function, but any function that depends only on  $m$  [i.e.,  $S(m, \tilde{w}) = S(m)$ ]. When this is the case, the selection functions can be taken out of the integrals and cancel out. After the cancellation of  $S(m)$ , the denominator becomes unity because both  $p(\tilde{w}|m, i, d)$  and  $p(i)$  are properly normalized pdfs. The result has been given in Eq. 40 and is repeated here for completeness:

$$\begin{aligned} p(\tilde{w}|m, d) &= \frac{S(m) \int_{-\infty}^0 p(\tilde{w}|m, d, i) p(i) di}{\int_{-\infty}^0 \left( \int_{-\infty}^{\infty} S(m) p(\tilde{w}|m, d, i) d\tilde{w} \right) p(i) di} \\ &= \int_{-\infty}^0 p(\tilde{w}|m, d, i) p(i) di \\ &= \frac{\ln 10}{\sigma_w \sqrt{2\pi}} \int_{-\infty}^0 \frac{10^{2i}}{\sqrt{1-10^{2i}}} \exp \left[ -\frac{[\tilde{w} - i - (m + d - \gamma)/\beta]^2}{2\sigma_w^2} \right] di \end{aligned} \quad (\text{A5})$$



## B. DUAL-SCATTER MODEL STARTING FROM THE INVERSE MODEL

In this Appendix, I derive the joint PDF of the observables for the latent-inclination dual-scatter model, starting from the inverse model, complementing the derivation in § 7.2 that began with the forward model.

Starting from the joint pdf of  $p(\tilde{w}, m, d, i)$  in Eq. 37:

$$\begin{aligned} p(\tilde{w}, m, d, i) &= p(\tilde{w}|m, d, i)p(m|d)p(d)p(i) \\ &= p(d) \times \frac{\ln 10 \cdot 10^{2i}}{\sqrt{1-10^{2i}}} \frac{1}{\sigma_w \sqrt{2\pi}} \exp \left[ -\frac{[\tilde{w}-i-(m+d-\gamma)/\beta]^2}{2\sigma_w^2} \right] \phi_* 10^{(\alpha+1)(m+d-M_*)} \exp(-10^{m+d-M_*}) \end{aligned} \quad (\text{B6})$$

To account for measurement error in  $m$ , one needs to multiply an additional term because the joint probability including  $\tilde{m}$  is:

$$p(\tilde{w}, \tilde{m}, d, m, i) = p(\tilde{m}|m)p(\tilde{w}, m, d, i) \quad (\text{B7})$$

where the new factor is a Gaussian:

$$p(\tilde{m}|m) = \frac{1}{\sigma_m \sqrt{2\pi}} \exp \left[ -\frac{(\tilde{m}-m)^2}{2\sigma_m^2} \right] \quad (\text{B8})$$

The joint pdf of the observables  $(\tilde{w}, \tilde{m}, d)$  is then calculated by marginalizing the above joint pdf over both  $m$  and  $i$ :

$$\begin{aligned} p(\tilde{w}, \tilde{m}, d) &= \int_{-\infty}^0 \int_{-\infty}^{\infty} p(\tilde{w}, \tilde{m}, d, m, i) dm di \\ &= \frac{\phi_* p(d)}{2\pi \sigma_w \sigma_m} \int_{-\infty}^0 \int_{-\infty}^{\infty} \exp \left[ -\frac{[\tilde{w}-i-(m+d-\gamma)/\beta]^2}{2\sigma_w^2} \right] \exp \left[ -\frac{(\tilde{m}-m)^2}{2\sigma_m^2} \right] 10^{(\alpha+1)(m+d-M_*)} \exp(-10^{m+d-M_*}) p(i) dm di \end{aligned} \quad (\text{B9})$$

When substituting  $m$  for  $\beta(w-i)+\gamma-d$  using the idealized TFR, the above equation becomes:

$$\begin{aligned} p(\tilde{w}, \tilde{m}, d) &= \frac{\beta \phi_* p(d)}{2\pi \sigma_w \sigma_m} \int_{-\infty}^0 \int_{-\infty}^{\infty} \exp \left[ -\frac{(\tilde{w}-w)^2}{2\sigma_w^2} \right] \exp \left[ -\frac{[\tilde{m}+d-\beta(w-i)-\gamma]^2}{2\sigma_m^2} \right] 10^{(\alpha+1)[\beta(w-i)-v_*]} \exp[-10^{\beta(w-i)-v_*}] p(i) dw di \end{aligned} \quad (\text{B10})$$

which is equivalent to  $p(\tilde{m}, \tilde{w}, d)$  in Eq. 54. This exercise shows that the data likelihood function of the latent-inclination dual-scatter model remains the same whether one starts from the forward model or the inverse model. This proves that the dual-scatter model is bidirectional (i.e., symmetric), unlike the forward and the inverse models.

### C. NUMERICAL ACCELERATION BY FAST FOURIER TRANSFORM AND PYTORCH

For the latent-inclination dual-scatter model with step-like selection function, the conditional pdf that constitutes the likelihood function is (Eq. A4 in Appendix A):

$$p(\tilde{m}|\tilde{w}, d) = \frac{\sqrt{2}}{\sqrt{\pi}\sigma_m} \frac{\int_{-\infty}^0 \frac{10^{2i}}{\sqrt{1-10^{2i}}} \left( \int_{-\infty}^{\infty} \exp\left[-\frac{(w-\tilde{w})^2}{2\sigma_w^2}\right] \exp\left[-\frac{(\tilde{m}+d-\beta(w-i)-\gamma)^2}{2\sigma_m^2}\right] 10^{(\alpha+1)[\beta(w-i)-v_*]} \exp(-10^{\beta(w-i)-v_*}) dw \right) di}{\int_{-\infty}^0 \frac{10^{2i}}{\sqrt{1-10^{2i}}} \left( \int_{-\infty}^{\infty} \exp\left[-\frac{(w-\tilde{w})^2}{2\sigma_w^2}\right] \text{erfc}\left(\frac{m_l+d-\beta(w-i)-\gamma}{\sqrt{2}\sigma_m}\right) 10^{(\alpha+1)[\beta(w-i)-v_*]} \exp(-10^{\beta(w-i)-v_*}) dw \right) di} \quad (\text{C11})$$

Direct integration of the double integrals for each data point is computationally expensive. For instance, on a 2021 Macbook Pro with an M1 Pro chip (8 performance CPU cores and 2 efficiency cores), direct integrations of 1,024 nodes in each integration axis take  $\sim 17$  min per MCMC step for 16 walkers and 10,147 data points. This is painfully slow and forces the user to either reduce the number of nodes or to bin the data points, neither of which is desirable. In this Appendix, I describe numerical methods that expedited the computation by thousands of times.

The key is to realize that the integral over  $i$  on the numerator of Eq. C11,

$$\int_{-\infty}^0 \frac{10^{2i}}{\sqrt{1-10^{2i}}} \exp\left[-\frac{(\tilde{m}+d-\beta(w-i)-\gamma)^2}{2\sigma_m^2}\right] 10^{(\alpha+1)[\beta(w-i)-v_*]} \exp(-10^{\beta(w-i)-v_*}) di \quad (\text{C12})$$

is a convolution between:

$$f(i) = \begin{cases} \frac{10^{2i}}{\sqrt{1-10^{2i}}}, & \text{if } i < 0 \\ 0, & \text{otherwise} \end{cases} \quad (\text{C13})$$

and

$$g(w-i) = \exp\left[-\frac{(\tilde{m}+d-\beta(w-i)-\gamma)^2}{2\sigma_m^2}\right] 10^{(\alpha+1)[\beta(w-i)-v_*]} \exp(-10^{\beta(w-i)-v_*}) \quad (\text{C14})$$

The result is a new function of  $w$ :

$$\{f * g\}(w) = \int_{-\infty}^{\infty} f(i)g(w-i)di \quad (\text{C15})$$

which can be calculated with Fourier transform given the convolution theorem:

$$\{f * g\}(w) = \mathcal{F}^{-1}[\mathcal{F}\{f\}(w) \cdot \mathcal{F}\{g\}(w)] \quad (\text{C16})$$

Since the computation of the Fast Fourier Transform (FFT) of  $f$  and  $g$  and the inverse FFT of their product has a total complexity of  $\mathcal{O}(N \log N)$  and direct computation of the convolution integral has a complexity of  $\mathcal{O}(N^2)$ , substantial improvement in computing efficiency can be achieved with FFT-based convolution when the number of nodes in each integration axis ( $N$ ) is a large number.

The result of the convolution can then be supplied to the integral over  $w$  to complete the calculation of the numerator:

$$\int_{-\infty}^{\infty} \{f * g\}(w) \exp\left[-\frac{(w-\tilde{w})^2}{2\sigma_w^2}\right] dw \quad (\text{C17})$$

Notice that even though this second integral is also a convolution, direct evaluation is faster because the convolution result is needed only at a specific value of  $\tilde{w}$  instead of an array of  $\tilde{w}$ . In such cases, FFT methods have no computational advantage over direct evaluation.

The same method can be applied to compute the denominator, simply replace the function of  $g$  as:

$$g(w-i) = \text{erfc}\left(\frac{m_l+d-\beta(w-i)-\gamma}{\sqrt{2}\sigma_m}\right) 10^{(\alpha+1)[\beta(w-i)-v_*]} \exp(-10^{\beta(w-i)-v_*}) \quad (\text{C18})$$

On the same 2021 Macbook Pro with the same 1,024 nodes in each integral, the FFT-accelerated dual-scatter model takes 1.65 s per MCMC step for 16 walkers and 10,147 data points, which is  $\sim 600\times$  faster than direct integration. The FFT-accelerated dual-scatter model is only  $1.9\times$  and  $7.6\times$  slower than the much simpler forward and inverse models, respectively.

An additional  $\sim 4\times$  acceleration is achieved by vectorizing the function that computes the conditional pdf in Eq. C11 and leveraging the integrated 16-core Graphics Processing Unit (GPU) with PyTorch's Metal Performance Shaders (MPS) backend (Paszke et al. 2019). The GPU- and FFT-powered dual-scatter model takes 0.44 s per MCMC step, which is even  $2\times$  faster than the simpler forward model on the CPU. The MCMC chains sufficiently converged after  $\sim 6,000$  steps, so the total computing time is  $\sim 0.75$  hr. The PyTorch version of the code is made available along with the NumPy version.