

ADDING A THIRD NORMAL TO CLUBB

by

Sven Bergmann

A Dissertation Submitted in
Partial Fulfillment of the
Requirements for the Degree of

Master of Science
in Mathematics

at

The University of Wisconsin–Milwaukee

May 2024

ABSTRACT

ADDING A THIRD NORMAL TO CLUBB

by

Sven Bergmann

The University of Wisconsin–Milwaukee, 2024
Under the Supervision of Professor Vince Larson

The Cloud Layers Unified By Binormals (CLUBB) model plays a crucial role in simulating atmospheric phenomena. It uses the sum of two normal probability density functions (pdfs) to represent subgrid variability within a single grid layer. This binormal approach, while computationally efficient, restricts the model’s ability to capture the full spectrum of potential shapes encountered in real-world atmospheric data.

This thesis proposes an innovative extension to the CLUBB model. We introduce a third normal pdf strategically positioned between the existing two, significantly enhancing the model’s representational flexibility. This trinormal representation allows for a wider range of grid-layer shapes while permitting analytic solutions for certain higher order moments.

The core of this work lies in deriving the necessary mathematical transformations for incorporating the third normal pdf seamlessly into the CLUBB framework. This thesis lists all formulas, inputs, and outputs associated with the extended model. Additionally, it tries to describe certain asymptotic behavior of the trinormal pdf under various parameter settings.

Type dedication here.

TABLE OF CONTENTS

LIST OF FIGURES	vi
LIST OF TABLES	vii
LIST OF LISTINGS	viii
LIST OF ACRONYMS	x
LIST OF SYMBOLS	xi
ACKNOWLEDGEMENTS	xii
1 Introduction	1
2 Problem	3
2.1 Motivation to add a third normal component	3
2.2 Closing turbulence pdes by integration over a pdf	4
2.3 Derivation of trinormal closures by transformation of binormal closures . . .	6
2.4 Goal of this thesis	10
2.5 Inputs and outputs of the verification procedure	10
2.5.1 Inputs and outputs of a forward run	10
2.5.2 Inputs and outputs of a backward run (verification direction)	11
2.6 Steps for checking the formulas	11
3 Definitions	13
3.1 Normal distribution	13
3.1.1 Multivariate normal distribution	13
3.1.2 Moments	14
3.2 Variates of the pdf	15
4 Formulas	16
4.1 Definition of the trinormal distribution, P_{tmg}	16
4.2 Normalized variables	18
4.3 Deduced lower order moments	18
4.3.1 Moments for w	19
4.3.2 Moments for θ_l	20
4.3.3 Moments for r_t	21
4.3.4 Mixed moments	22
4.4 Solving for pdf parameters by using the moment terms	23
4.5 Leveraging the pdf for higher-order moment calculations	25
4.6 Addressing skewness and achieving closure	26
4.7 Formulating closure relationships for higher-order moments	29

5	Integration using SymPy	35
5.1	Analytic integration	35
5.2	Numeric integration	38
6	Asymptotics	43
7	Summary	45
8	Outlook	46
	Bibliography	48
A	Code	49
A.1	User Guide	49
A.1.1	Accessing the code	49
A.1.2	Key files and their purposes	49
A.1.3	Working with functions	49
A.1.4	Displaying equations effectively	49
A.1.5	Handling integrals	49
A.2	symbols.py	50
A.3	checked_functions.py	52
A.4	w_prime_4_bar.py	63
A.5	w_prime_2_theta_1_prime_bar.py	65
A.6	w_prime_theta_1_prime_2_bar.py	67
A.7	w_prime_r_t_prime_theta_1_prime_bar.py	72

LIST OF FIGURES

2.1	Binormal plot for two strong up-/downdrafts	3
2.2	Binormal plot for two strong up-/downdrafts with increased standard deviations	4
2.3	Trinormal plot for two strong up-/downdrafts with varying δ	5
2.4	Trinormal plot for two strong up-/downdrafts with a third peak in the middle	6
5.1	Output of listing 5.4	37
5.2	Output of listing 5.5	37
5.3	Output of listing 5.7	38
5.4	Output of listing 5.13	40
5.5	Output of listing 5.18	42

LIST OF TABLES

2.1 1D Plots for different δ and σ_{w3} 9

LIST OF LISTINGS

5.1	Import statements	36
5.2	Defining symbols	36
5.3	Defining the marginals	36
5.4	Defining and displaying the needed integral	37
5.5	Calculating and printing the integral	37
5.6	Python function for the second order moment	37
5.7	Printing the symbolic equation	38
5.8	Check if the integral and the given formula are the same	38
5.9	Import statements	38
5.10	Defining symbols	39
5.11	Defining the marginals	39
5.12	Defining and displaying the needed integral	40
5.13	Python function for $\overline{w'^2\theta_l}$	40
5.14	Create a dataframe and putting in arbitrary numbers	41
5.15	Attaching the “checkval” column to the dataframe	41
5.16	Attaching the “numint” column to the dataframe	42
5.17	Attaching the “diffnum” column to the dataframe	42
5.18	Calculating the mean difference	42
A.1	symbols.py	52
A.2	checked_functions.py	63
A.3	w_prime_4_bar.py	65
A.4	w_prime_2_theta_l_prime_bar.py	67

A.5	w_prime_theta_l_prime_2_bar.py	72
A.6	w_prime_r_t_prime_theta_l_prime_bar.py	77

LIST OF ACRONYMS

cas computer algebra system. 16, 46

CLUBB Cloud Layers Unified By Binormals. ii, 1, 2, 4, 5, 11, 18, 26, 29, 30, 45–47

lhs left hand side. 11

pde partial differential equation. 1, 4, 5, 25

pdf probability density function. ii, 1, 2, 6, 8, 10–13, 15, 17, 18, 20, 23–25, 27, 29–31, 43,
44, 46

rhs right hand side. 5, 11, 44

LIST OF SYMBOLS

θ'_l Standardized liquid water potential temperature ($\theta'_l = \theta_l - \overline{\theta_l}$). 15

r'_t Standardized total water mixing ratio ($r'_t = r_t - \overline{r_t}$). 15

w' Standardized upward wind ($w' = w - \overline{w}$). 15, 19

ACKNOWLEDGEMENTS

The following work is based on a the papers “Using probability density functions to derive consistent closure relationships among higher-order moments” by Larson and Golaz and *CLUBB-SILHS: A parameterization of subgrid variability in the atmosphere* by Larson.

1 Introduction

The CLUBB model is a powerful tool used to simulate atmospheric behavior within climate models. This document explores an extension to the current CLUBB framework. Currently, CLUBB utilizes the sum of two normal pdfs to represent a single atmospheric grid layer. While effective, this approach limits the model’s ability to capture the full spectrum of potential cloud layer shapes. This work proposes an innovative solution: incorporating a third normal pdf into the CLUBB framework. This addition aims to enhance the model’s representational capabilities while maintaining computational efficiency and numerical stability. To achieve this, the document dives into the details of the proposed method.

We begin by outlining the core problem we aim to address (chapter 2) by starting with the motivation, proceeding with a short explanation on how to close turbulence partial differential equations (pdes) (section 2.2) and explaining how we derive the transformation from the formulas given by the paper “Using probability density functions to derive consistent closure relationships among higher-order moments”¹ (section 2.3). After that, we define the goal of this thesis (section 2.4), talk about the inputs and outputs (section 2.5) and provide steps for checking those formulas (section 2.6).

Following this motivational chapter, we establish a foundation with clear definitions of the relevant concepts, including normal distributions and the thermodynamic scalars crucial for atmospheric modeling (chapter 3).

Chapter 4 forms the heart of this work, presenting the actual formulas associated with the

¹Vincent E Larson and Jean-Christophe Golaz. “Using probability density functions to derive consistent closure relationships among higher-order moments”. In: *Monthly Weather Review* 133.4 (2005), pp. 1023–1042.

extended CLUBB model. This chapter details the introduction of the third normal pdf (section 4.1) and the derivation of key moments within the model (section 4.3 - section 4.4). Additionally, section 4.6 proposes a diagnostic approach to account for the skewness of heat and moisture, while section 4.7 introduces analytic closure relations for higher-order moments based on the newly formed mixture of three normal distributions.

To handle the mathematical integrations required by the model, chapter 5 explores both exact parametric and numerical integration techniques of verifying the integrals, utilizing the SymPy library (section 5.1 & section 5.2).

Finally, chapter 6 investigates the asymptotic behavior of the extended model, providing valuable insights into its performance under various conditions.

Having talked about the trinormal representation within the CLUBB model, chapter 7 provides a concise recap of the key findings. This summary serves as a comprehensive overview of the essential concepts explored throughout this thesis.

The document concludes with an outlook in chapter 8, outlining potential future directions for research and exploration based on the findings presented here.

2 Problem

2.1 Motivation to add a third normal component

As is said in chapter 1, we try to describe more possible shapes by adding a third normal component. To illustrate that, we plot some of the shapes which are now possible with three normals but were not possible with only two. To be able to draw those plots, we are just using two variables, w , the upward wind, and θ_l , the liquid water potential temperature. To show how the binormal model handles strong winds, let us consider a scenario with a strong updraft at w_1 , as well as a strong downdraft at w_2 . The way the current binormal model would handle this could look like figure 2.1. However, this binormal distribution (figure 2.1)

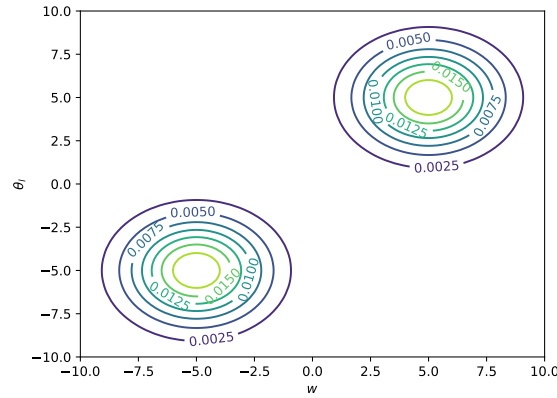


Figure 2.1: Binormal plot for two strong up-/downdrafts
 $w_1 = 5$, $w_2 = -5$, $\theta_{l1} = 5$, $\theta_{l2} = -5$, $\alpha = 0.5$, $\sigma_w = 2$, $\sigma_{\theta_{l1}} = 2$, $\sigma_{\theta_{l2}} = 2$.

does not accurately reflect reality. In nature, we would not expect such strong bimodality between the strong up- and downdrafts at w_1 and w_2 . There would most likely be some weaker drafts present in-between. The current binormal model can attempt to capture this smoother transition by simply increasing the standard deviations of both wind, and liquid

water potential temperature distributions. This results in a broader distribution with a connection between the two peaks, as shown in figure 2.2. Seeing figure 2.2, the issue with

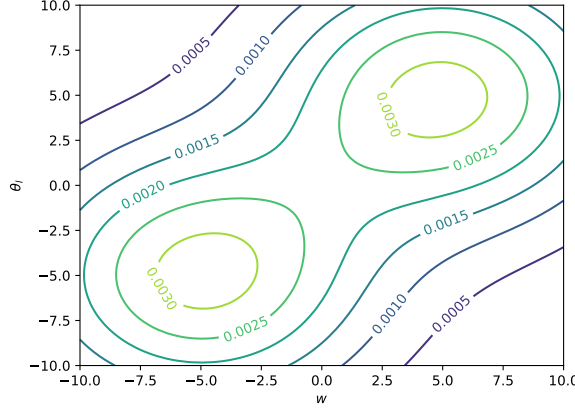


Figure 2.2: Binormal plot for two strong up-/downdrafts with increased standard deviations $w_1 = 5$, $w_2 = -5$, $\theta_{l1} = 5$, $\theta_{l2} = -5$, $\alpha = 0.5$, $\sigma_w = 5$, $\sigma_{\theta_{l1}} = 5$, $\sigma_{\theta_{l2}} = 5$.

having some values in the middle is mitigated but the general width of the normals was increased, too. Since CLUBB also has the simplification that there is no correlation between w and θ_l , and w and r_t – obviously – one cannot just increase it. Therefore, the idea is to add this third normal distribution, which actually has correlation between all three variables and especially in the bivariate case, between w and θ_l . Figure 2.2 would then change to figure 2.3. Now, one can easily model something like the described shape, as illustrated in figure 2.3. Also, some other (maybe weird) shapes are now possible, just like the one in figure 2.4.

2.2 Closing turbulence pdes by integration over a pdf

The CLUBB model relies on a set of pdes to represent atmospheric processes. These equations require closure, implying the expression of all terms solely in terms of known quantities. This closure process often involves integrals, and verifying their analytical solutions ensures

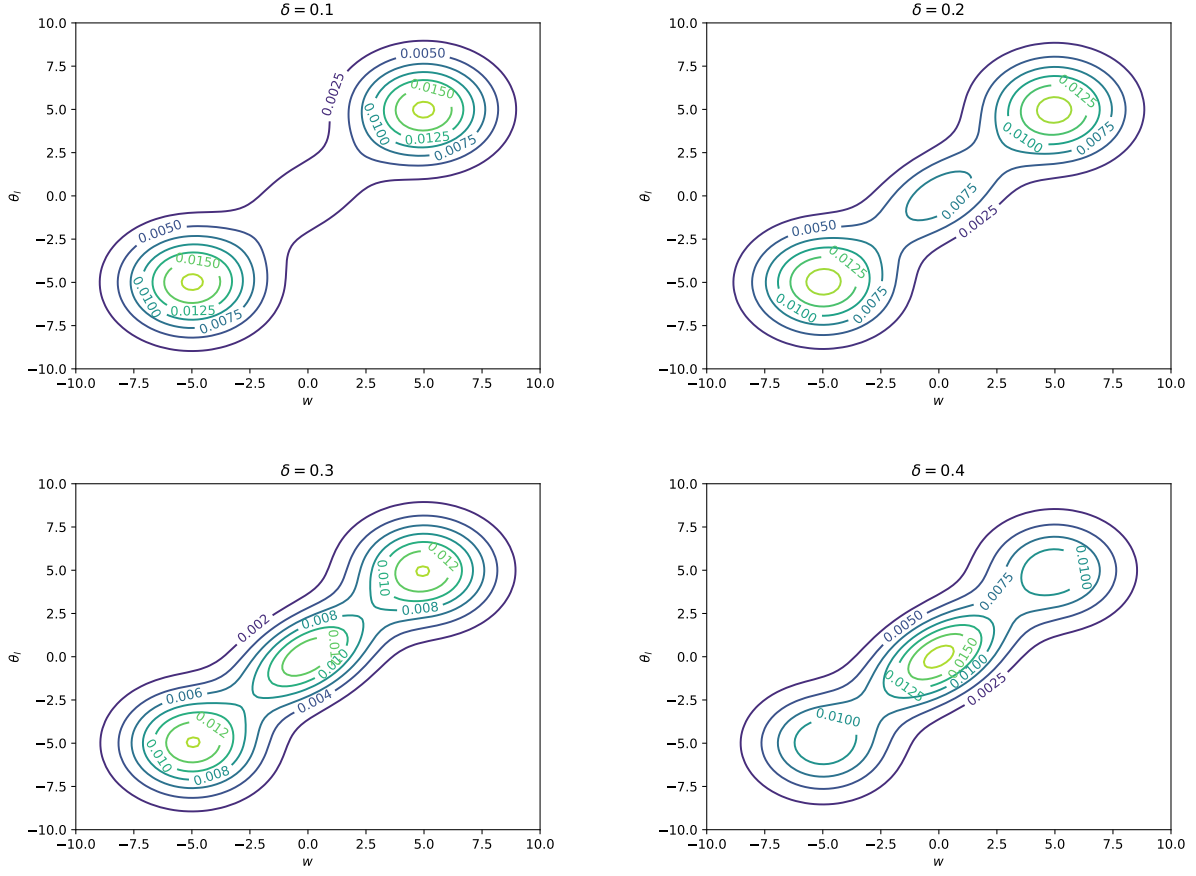


Figure 2.3: Trinormal plot for two strong up-/downdrafts with varying δ
 $w_1 = 5$, $w_2 = -5$, $\theta_{l1} = 5$, $\theta_{l2} = -5$, $\alpha = 0.5$, $\sigma_w = 2$, $\sigma_{\theta_{l1}} = 2$, $\sigma_{\theta_{l2}} = 2$, $\sigma_{w3} = 2$, $\sigma_{3\theta_l} = 2$,
 $\rho_{w\theta_l} = 0.5$.

the model's mathematical integrity. For instance, consider the following prognostic pde¹:

$$\frac{\partial \overline{w'\theta'_l}}{\partial t} = -\overline{w} \frac{\partial \overline{w'\theta'_l}}{\partial z} - \frac{1}{\rho_s} \frac{\partial \rho_s \overline{w'^2 \theta'_l}}{\partial z} - \overline{w'^2} \frac{\partial \overline{\theta'_l}}{\partial z} - \overline{w'\theta'_l} \frac{\partial \overline{w}}{\partial z} + \dots$$

While the details of initial and boundary conditions are essential for formally closing the prognostic equations in CLUBB (omitted for brevity), this section emphasizes the importance of efficiently calculating moments on the right hand side (rhs) of these equations. That is because the model steps forward in time and therefore needs the repeated calculation of moments for each unclosed prognostic equation at every time step. Those already expensive computational steps need to use mathematically simpler moment representations

¹Vincent E. Larson. *CLUBB-SILHS: A parameterization of subgrid variability in the atmosphere*. 2022. arXiv: 1711.03675 [physics.ao-ph], p. 21.

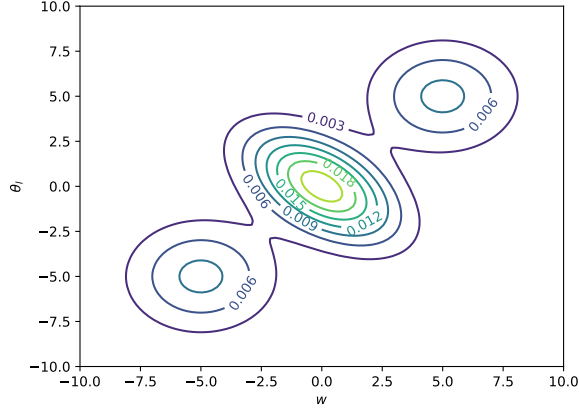


Figure 2.4: Trinormal plot for two strong up-/downdrafts with a third peak in the middle
 $w_1 = 5$, $w_2 = -5$, $\theta_{l1} = 5$, $\theta_{l2} = -5$, $\alpha = 0.5$, $\delta = 0.5$, $\sigma_w = 2$, $\sigma_{\theta_{l1}} = 2$, $\sigma_{\theta_{l2}} = 2$, $\sigma_{w3} = 2$,
 $\sigma_{\theta_{l3}} = 2$, $\rho_{w\theta_l} = 0.5$.

of e.g. $\overline{w'^2\theta'_l}$, even if they introduce slight limitations in capturing the full variability of the underlying atmospheric state.

2.3 Derivation of trinormal closures by transformation of binormal closures

Analytic closures between higher and lower order moments for the binormal case are available (see CLUBB-SILHS²). We wish to derive similar analytic closures for the proposed trinormal pdf. Deriving an analytic closure for a general trinormal pdf is difficult. However, doing so is tractable in the special case that the third normal component is located at the mean of the binormal pdf. In fact, the trinormal closures can be derived by making a simple transformation of the binormal closures³, e.g.

$$\overline{w'^2} = \alpha[(w_1 - \bar{w})^2 + \sigma_w^2] + (1 - \alpha)[(w_2 - \bar{w})^2 + \sigma_w^2]. \quad (2.3.1)$$

²Larson, *CLUBB-SILHS: A parameterization of subgrid variability in the atmosphere*.

³Larson and Golaz, “Using probability density functions to derive consistent closure relationships among higher-order moments”.

This section will demonstrate that the following transformations, denoted by the subscript “dGn” for the binormal case, successfully achieve this conversion.

$$\overline{w'^2} \frac{1 - \delta \lambda_w}{1 - \delta} = \overline{w'^2}_{dGn}, \quad (2.3.2)$$

$$\overline{w'^3} \frac{1}{1 - \delta} = \overline{w'^3}_{dGn}, \quad (2.3.3)$$

$$\frac{\overline{w'^3}}{\overline{w'^2}^{3/2}} \frac{(1 - \delta)^{1/2}}{(1 - \lambda_w \delta)^{3/2}} = \frac{\overline{w'^3}_{dGn}}{\overline{w'^2}_{dGn}^{3/2}}, \quad (2.3.4)$$

$$\overline{\theta_l'^2} \frac{1 - \delta \lambda_\theta}{1 - \delta} = \overline{\theta_l'^2}_{dGn}, \quad (2.3.5)$$

$$\overline{w' \theta_l'} \frac{1 - \delta \lambda_{w\theta}}{1 - \delta} = \overline{w' \theta_l'}_{dGn}, \quad (2.3.6)$$

$$\left(\overline{w'^4} - 3\delta \lambda_w^2 (\overline{w'^2})^2 \right) \frac{1}{1 - \delta} = \overline{w'^4}_{dGn} \quad (2.3.7)$$

$$\left(\frac{\overline{w'^4}}{(\overline{w'^2})^2} - 3\delta \lambda_w^2 \right) \frac{1 - \delta}{(1 - \lambda_w \delta)^2} = \frac{\overline{w'^4}_{dGn}}{(\overline{w'^2}_{dGn})^2} \quad (2.3.8)$$

To get a sense of what those transformations mean and why they should work, we pick e.g. equation (2.3.2). If we substitute in the already defined formula for λ_w (equation (4.1.4)), we get

$$\begin{aligned} \overline{w'^2} \left(1 - \delta \frac{\sigma_{w3}^2}{\overline{w'^2}} \right) &= (1 - \delta) \overline{w'^2}_{dGn} \\ \overline{w'^2} - \delta \sigma_{w3}^2 &= (1 - \delta) \overline{w'^2}_{dGn} \\ \overline{w'^2} &= \overline{w'^2}_{dGn} - \delta \overline{w'^2}_{dGn} + \delta \sigma_{w3}^2 \\ \overline{w'^2} &= \overline{w'^2}_{dGn} - \delta \left(\overline{w'^2}_{dGn} - \sigma_{w3}^2 \right). \end{aligned} \quad (2.3.9)$$

Our analysis reveals a key relationship between the parameter δ and the overall variance (often referred to as “width”) of the trinormal distribution. As the value of δ approaches 1

(but strictly remains less than 1), the standard deviation of the third normal distribution has a progressively stronger influence on the overall variance of the combined distribution. This intuitively makes sense because a larger weight assigned to the third normal distribution through δ will contribute more significantly to the spread of the combined pdf.

Also, if we look at equation (2.3.3), we see that there is no more λ_w present. It makes sense graphically, that as δ grows, which means that the normal pdf in the middle is growing, the overall skewness of all three normals has to change also, depending on the value of σ_{w3} . We can see this, as well as the relationship between the variance in table 2.1. Table 2.1 offers a visual representation of how the parameter δ influences the shape of the trinormal distribution. Each plot illustrates pdfs for different combinations of σ_{w3} (standard deviation of the third normal distribution) and δ . The row values in the table correspond to σ_{w3} , while the column values represent δ . We can observe two key trends within these plots:

1. *Influence of σ_{w3} :* As expected, varying σ_{w3} primarily affects the “width” or overall variance of the combined distribution. When σ_{w3} is larger than the width of the original binormal sum (orange/red line), choosing a larger δ allows the overall variance to increase significantly, as predicted by equation (2.3.2).
2. *Decreasing skewness with increasing δ :* The plots also reveal a distinct relationship between δ and the skewness of the resulting distribution. As δ increases, the skewness of the combined trinormal distribution (green line) progressively reduces. This phenomenon can be attributed to the placement of the third normal distribution. Placed directly between the two original normal distributions, the third normal distribution acts as a centralizing force. As the weight of the third normal distribution (controlled by δ) grows, its symmetric nature counteracts the potential skewness of the initial binormal sum. This effect is particularly strong in the bottom three plots, where a larger value of $\sigma_{w3} = 10$ is used. We observe a clear reduction in the skewness of the

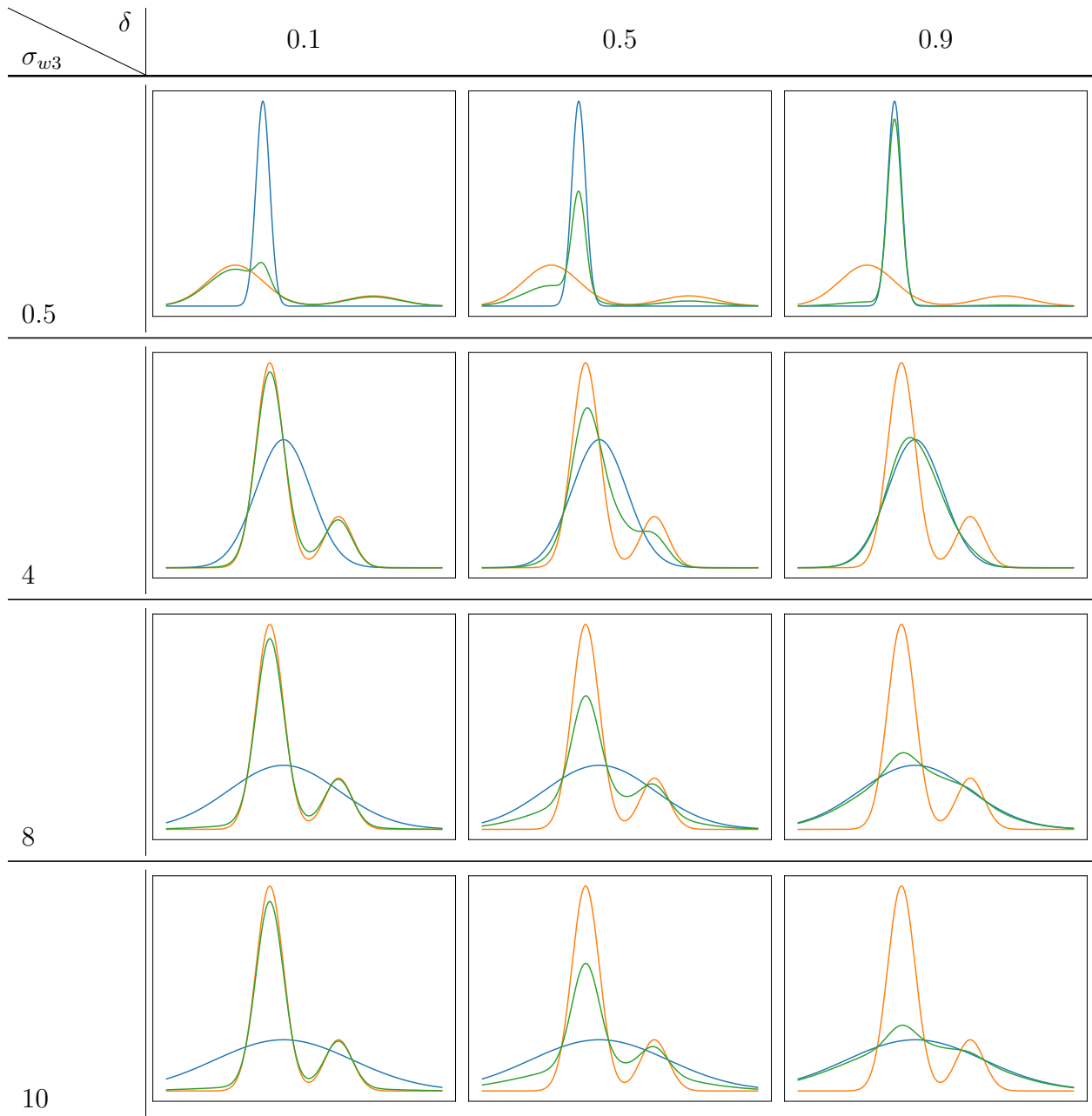


Table 2.1: 1D Plots for different δ and σ_{w3}
 $w_1 = 5$, $w_2 = -5$, $\alpha = 0.2$, $\sigma_w = 2$. The blue plot represents the third normal, the orange/red one represents the binormal, and the green one represents the mixture. The x and y labels and ticks are omitted for clarity.

green plot (mixture) as δ approaches 1.

2.4 Goal of this thesis

The goal of this thesis is to verify closure for higher-order moments, such as the third order moment $\overline{w'^2\theta'_l}$ and others like it. This closure is achieved by expressing these higher-order moments – i.e. equation (4.7.3), equation (4.7.5), and equation (4.7.7) – analytically in terms of readily calculable lower-order moments. This analytic approach uses the relationships between moments within a normal distribution, enabling efficient model updates during the time-stepping process.

2.5 Inputs and outputs of the verification procedure

While defining inputs and outputs can seem challenging at first glance, it is a crucial step towards understanding a system.

2.5.1 Inputs and outputs of a forward run

When forecasting the weather (“forward run”), the code provides us with a set of moment terms: \overline{w} , $\overline{w'^2}$, $\overline{w'^3}$, $\overline{\theta_l}$, $\overline{w'\theta'_l}$, $\overline{r_t}$, $\overline{w'r'_t}$, $\overline{\theta'^2_l}$, $\overline{r'^2_t}$, $\overline{r'_t\theta'_l}$. These are the inputs. From these inputs, we want to determine certain parameters which describe the shape of the underlying pdf. Those pdf parameters are standardized and some also normalized. So we try to solve these pdf parameters (13), namely α , \widehat{w}_1 , \widehat{w}_2 , $\tilde{\theta}_{l1}$, $\tilde{\theta}_{l2}$, \tilde{r}_{t1} , \tilde{r}_{t2} , $\tilde{\sigma}_w$, $\tilde{\sigma}_{\theta_{l1}}$, $\tilde{\sigma}_{\theta_{l2}}$, $\tilde{\sigma}_{r_{t1}}$, $\tilde{\sigma}_{r_{t2}}$, and $r_{r_t\theta_l}$. All the formulas are listed in chapter 4. Ultimately, the code needs to express even higher order moments such as $\overline{w'^2\theta'_l}$ in terms of the lower order moments. These higher order moments are the outputs in the “forward run”.

2.5.2 Inputs and outputs of a backward run (verification direction)

Although a “forward run” models the higher order moments in terms of the lower order moments, we want to verify these formulas, namely equation (4.7.3), equation (4.7.5), and equation (4.7.7). To achieve this, we will take a more traditional approach, working in the “backward” direction. This means we will:

1. *Specify the pdf parameters:* Start by explicitly defining the parameters that characterize the underlying pdf.
2. *Calculate the moments:* Once the pdf is defined, we can then calculate the desired moments, such as \bar{w} , through integration.

This can be done, e.g. by calculating the integral:

$$\bar{w} = \int_{\mathbb{R}} \int_{\mathbb{R}} \int_{\mathbb{R}} w \cdot P_{tmg} dw dr_t d\theta_l, \quad (2.5.1)$$

where P_{tmg} (**T**ri-variate **M**ixture of **G**aussians) is the pdf of the sum of all three normal distributions. Since some integrals are challenging to verify symbolically with SymPy, we are using the quadrature method of SymPy to calculate the integrals and choose arbitrary values for the inputs. All of this can be seen in section 5.2.

2.6 Steps for checking the formulas

This section outlines a general approach for verifying all of those integral expressions employed within the CLUBB model. This approach ensures the accuracy of the computed moment relationships. Chapter 5 discusses some actual examples. We verify the expressions using the following method where the order is crucial. We always want to check if left hand side (lhs) equals rhs:

1. Choose *dimensional* parameters (parameters without any tilde or hat) that determine the pdf, i.e. choose dimensional pdf parameters, e.g. σ_{w3} . Then the pdf is known and any moments of it can be calculated by integration.
2. Calculate the means, e.g. $\bar{w} = \mathbb{E}[w]$ by integration over the pdf. The formula for \bar{w} (equation (4.3.1)) in terms of the pdf parameters can be checked.
3. Once the means are known, we calculate the central variances, e.g. $\overline{w'^2} = \overline{(w - \bar{w})^2}$ by integration over the pdf. The formula for $\overline{w'^2}$ (equation (4.3.3)) in terms of the pdf parameters can be checked.
4. Once the variances, e.g. $\overline{w'^2}$, are known, then the *non-dimensional* pdf parameters such as λ_w (equation (4.1.4)) can be calculated by their definitions.
5. We can also calculate the covariances by 2D integration over a 2D pdf. Again, our formulas in terms of pdf parameters can be checked.
6. Finally, we can calculate the higher order moments, i.e. $\overline{w'^4}$ (equation (4.7.3)) or $\overline{w'^2\theta'_l}$ (equation (4.7.5)) or $\overline{w'\theta'^2_l}$ (equation (4.7.7)), by integration over the pdf.

3 Definitions

For better understanding of the topics covered in this thesis, it follows a brief introduction of all formulas and terms used.

3.1 Normal distribution

We say that a random variable X is distributed according to a normal distribution ($X \sim \mathcal{N}(\mu, \sigma^2)$) when it has the following pdf:

Definition 1 (pdf of a normal distribution)

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right). \quad (3.1.1)$$

3.1.1 Multivariate normal distribution

We say that a random vector \mathbf{X} ($r \times r$) is distributed according to a multivariate normal distribution when it has the following joint density function:¹

Definition 2 (pdf of a multivariate normal distribution)

$$f(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-\frac{r}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right), \mathbf{x} \in \mathbb{R}^r, \quad (3.1.2)$$

¹Alan Julian Izenman. *Modern multivariate statistical techniques: regression, classification, and manifold learning*. Springer texts in statistics. OCLC: ocn225427579. New York ; London: Springer, 2008. ISBN: 9780387781884, p. 59.

where

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_r \end{pmatrix} \in \mathbb{R}^r \quad (3.1.3)$$

is the mean vector, and

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \rho_{13}\sigma_1\sigma_3 & \dots & \rho_{1r}\sigma_1\sigma_r \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 & \rho_{23}\sigma_2\sigma_3 & \dots & \vdots \\ \rho_{13}\sigma_1\sigma_3 & \rho_{23}\sigma_2\sigma_3 & \sigma_3^2 & \dots & \vdots \\ \vdots & \dots & \dots & \ddots & \vdots \\ \rho_{1r}\sigma_1\sigma_r & \dots & \dots & \dots & \sigma_r^2 \end{pmatrix} \in \mathbb{R}^{r \times r} \quad (3.1.4)$$

is the (symmetric, positive definite) covariance matrix. This is also often expressed as $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, meaning that \mathbf{X} ($r \times r$ random vector) is distributed according to a multivariate normal distribution with the given parameters.

3.1.2 Moments

Especially for this thesis, we are interested in the moments of the given multivariate normal distribution. We can express the first order moment as the mean, denoted as $\bar{X} = \mathbb{E}[X]$, where X is a random variable. The second order moment is $\mathbb{E}[X^2]$, also denoted as the variance if it is a central moment. The standardized third and fourth order moments have special names, so called skewness and kurtosis respectively. We denote this by the following:

$$\mathbb{E}[X^3] = \mathbb{E} \left[\left(\frac{X - \mu}{\sigma} \right)^3 \right] = \frac{\mu_3}{\sigma^3} = \frac{\mathbb{E}[(X - \mu)^3]}{(\mathbb{E}[(X - \mu)^2])^{3/2}}, \quad (3.1.5)$$

$$\mathbb{E}[X^4] = \mathbb{E} \left[\left(\frac{X - \mu}{\sigma} \right)^4 \right] = \frac{\mathbb{E}[(X - \mu)^4]}{(\mathbb{E}[(X - \mu)^2])^2} = \frac{\mu_4}{\sigma^4}. \quad (3.1.6)$$

3.2 Variates of the pdf

We denote the variates of the pdf by w , r_t , and θ_l , where w is the upward wind, r_t is the liquid water potential temperature and θ_l is the liquid water potential temperature.² Variables denoted by w' are defined by $w - \bar{w}$ where \bar{w} is the mean of w over the whole pdf. We define r'_t and θ'_l in the same way.

²Larson, *CLUBB-SILHS: A parameterization of subgrid variability in the atmosphere*, p. 10.

4 Formulas

This chapter lists all formulas which are derived from the binormal model to this model with an additional normal. All formulas listed are either tested by using a computer algebra system (cas) and calculating the integrals analytically with ranges $-\infty$ to ∞ or using the quadrature procedure with large enough ranges such that the error is (numerically) zero. Those two procedures are explained in chapter 5.

4.1 Definition of the trinormal distribution, P_{tmg}

We would like to add a third normal to the already existing two trivariate normals, which is placed right in the middle between those two. For our proposed mixture of normals we then have

$$P_{tmg}(w, \theta_l, r_t) = \alpha(1 - \delta)\mathcal{N}(\mu_1, \Sigma_1) + (1 - \alpha)(1 - \delta)\mathcal{N}(\mu_2, \Sigma_2) + \delta\mathcal{N}(\mu_3, \Sigma_3), \quad (4.1.1)$$

where \mathcal{N} denotes the multivariate normal distribution, $\alpha \in (0, 1)$ is the mixture fraction of the binormal, and $\delta \in [0, 1)$ is the weight of the third normal. The mean vectors and the covariance matrices are defined in the following.

We define the mean vectors of the first and second normal distributions as $\mu_1 = (w_1, \theta_{l1}, r_{t1})^\top$, and $\mu_2 = (w_2, \theta_{l2}, r_{t2})^\top$, where $w_1 > w_2$ (due to a convention in the code) and the covariance

matrices as

$$\Sigma_1 = \begin{pmatrix} \sigma_w^2 & 0 & 0 \\ 0 & \sigma_{\theta_l}^2 & \rho_{\theta_l r_t} \sigma_{\theta_l 3} \sigma_{r_t 3} \\ 0 & \rho_{\theta_l r_t} \sigma_{\theta_l 3} \sigma_{r_t 3} & \sigma_{r_t}^2 \end{pmatrix}, \text{ and } \Sigma_2 = \begin{pmatrix} \sigma_w^2 & 0 & 0 \\ 0 & \sigma_{\theta_{l2}}^2 & \rho_{\theta_{l2} r_t} \sigma_{\theta_{l2} 3} \sigma_{r_t 3} \\ 0 & \rho_{\theta_{l2} r_t} \sigma_{\theta_{l2} 3} \sigma_{r_t 3} & \sigma_{r_t}^2 \end{pmatrix}. \quad (4.1.2)$$

It might be of interest that there is no correlation between w and θ_l or w and r_t . That is to make the pdfs mathematically more tractable by therefore also taking away some variability. This is not the case for the third normal, though.

It has already been said, that we would like to place the third normal right at the mean, therefore μ_3 and Σ_3 are defined as

$$\mu_3 = \begin{pmatrix} \overline{w} \\ \overline{\theta_l} \\ \overline{r_t} \end{pmatrix}, \text{ and } \Sigma_3 = \begin{pmatrix} \sigma_{w3}^2 & \rho_{w\theta_l} \sigma_{w3} \sigma_{\theta_l 3} & \rho_{wr_t} \sigma_{w3} \sigma_{r_t 3} \\ \rho_{w\theta_l} \sigma_{w3} \sigma_{\theta_l 3} & \sigma_{\theta_l 3}^2 & \rho_{\theta_l r_t} \sigma_{\theta_l 3} \sigma_{r_t 3} \\ \rho_{wr_t} \sigma_{w3} \sigma_{r_t 3} & \rho_{\theta_l r_t} \sigma_{\theta_l 3} \sigma_{r_t 3} & \sigma_{r_t 3}^2 \end{pmatrix}. \quad (4.1.3)$$

The advantage over just two normal pdfs is that we can now express a greater variety of shapes. We also define some additional relationships for this third normal distribution.

$$\lambda_w \equiv \frac{\sigma_{w3}^2}{w'^2}, \quad \lambda_\theta \equiv \frac{\sigma_{\theta_l 3}^2}{\theta_l'^2}, \quad \lambda_r \equiv \frac{\sigma_{r_t 3}^2}{r_t'^2}, \quad (4.1.4)$$

$$\lambda_{\theta r} \equiv \frac{\rho_{\theta_l r_t} \sigma_{\theta_l 3} \sigma_{r_t 3}}{r_t' \theta_l'}, \quad \lambda_{w\theta} \equiv \frac{\rho_{w\theta_l} \sigma_{w3} \sigma_{\theta_l 3}}{w' \theta_l'}, \quad \lambda_{wr} \equiv \frac{\rho_{wr_t} \sigma_{w3} \sigma_{r_t 3}}{w' r_t'}. \quad (4.1.5)$$

Hence, we can rewrite Σ_3 as

$$\Sigma_3 = \begin{pmatrix} \sigma_{w3}^2 & \overline{w' \theta_l'} \cdot \lambda_{w\theta} & \overline{w' r_t'} \cdot \lambda_{wr} \\ \overline{w' \theta_l'} \cdot \lambda_{w\theta} & \sigma_{\theta_l 3}^2 & \overline{r_t' \theta_l'} \cdot \lambda_{\theta r} \\ \overline{w' r_t'} \cdot \lambda_{wr} & \overline{r_t' \theta_l'} \cdot \lambda_{\theta r} & \sigma_{r_t 3}^2 \end{pmatrix}. \quad (4.1.6)$$

4.2 Normalized variables

Since CLUBB is mostly using “normalized variables”, we are going to list those, which are given in standard form. We are also doing that for making the transformations easier.

$$\tilde{\theta}'_l \equiv \frac{\theta_l - \bar{\theta}_l}{\sqrt{\theta_l'^2}} \frac{1}{\sqrt{\frac{1-\delta\lambda_\theta}{1-\delta}}}, \quad (4.2.1)$$

$$\tilde{r}'_t \equiv \frac{r_t - \bar{r}_t}{\sqrt{r_t'^2}} \frac{1}{\sqrt{\frac{1-\delta\lambda_r}{1-\delta}}}, \quad (4.2.2)$$

where $\bar{\theta}_l$ and \bar{r}_t are the means for the full summed up pdf and $\bar{\theta}_l'^2$ as well as $\bar{r}_t'^2$ are the variances for θ_l and r_t .

For the standard deviations, we define

$$\tilde{\sigma}_w \equiv \frac{\sigma_w}{\sqrt{w'^2}} \frac{1}{\sqrt{\frac{1-\delta\lambda_w}{1-\delta}}}, \quad (4.2.3)$$

where σ_w denotes the standard deviation of the w -component, $\bar{w'^2}$ is the variance for w ,

$$\tilde{\sigma}_{\theta_i} \equiv \frac{\sigma_{\theta_i}}{\sqrt{\theta_l'^2}} \frac{1}{\sqrt{\frac{1-\delta\lambda_\theta}{1-\delta}}}, \quad (4.2.4)$$

where σ_{θ_i} denotes the standard deviation of the i^{th} θ_l -component, and

$$\tilde{\sigma}_{r_i} \equiv \frac{\sigma_{r_i}}{\sqrt{r_t'^2}} \frac{1}{\sqrt{\frac{1-\delta\lambda_r}{1-\delta}}}. \quad (4.2.5)$$

where σ_{r_i} denotes the standard deviation of the i^{th} σ_{r_t} -component,

4.3 Deduced lower order moments

We start by outlining the equations capturing lower-order moments, expressed in terms of pdf parameters. These equations can be presented in either *dimensional* or *non-dimensional*

form. While both representations are mathematically valid, the *non-dimensional* form offers a distinct advantage: it highlights the underlying connection to the bivariate case.

4.3.1 Moments for w

The relationship for \bar{w} is given as follows:

$$\bar{w} = (1 - \delta)\alpha w_1 + (1 - \delta)(1 - \alpha)w_2 + \delta w_3, \quad (4.3.1)$$

where $w_3 \equiv \alpha w_1 + (1 - \alpha)w_2$. Therefore the mean of w stays the same as in the bivariate case. The relationship for the *non-dimensional* form is:

$$0 = \alpha \hat{w}_1 + (1 - \alpha) \hat{w}_2 \quad (4.3.2)$$

For all other moments – except for the mean – we are using the standardized versions of the variables, written as w' .

The second order moment is given as:

$$\begin{aligned} \overline{w'^2} &= (1 - \delta)\alpha[(w_1 - \bar{w})^2 + \sigma_w^2] \\ &\quad + (1 - \delta)(1 - \alpha)[(w_2 - \bar{w})^2 + \sigma_w^2] \\ &\quad + \delta\sigma_{w3}^2, \end{aligned} \quad (4.3.3)$$

where σ_{w3} is defined as $\lambda_w \overline{w'^2}$. This moment is also the variance of w at the same time, since

$$\begin{aligned} \overline{w'^2} &= \mathbb{E}[w'^2] = \mathbb{E}[(w - \bar{w})^2] = \mathbb{E}[(w - \mathbb{E}[w])^2] = \mathbb{E}[w^2 - 2w\mathbb{E}[w] + \mathbb{E}[w]^2] \\ &= \mathbb{E}[w^2] - 2\mathbb{E}[w\mathbb{E}[w]] + \mathbb{E}[\mathbb{E}[w]^2] = \mathbb{E}[w^2] - 2\mathbb{E}[w]\mathbb{E}[w] + \mathbb{E}[w]^2 \\ &= \mathbb{E}[w^2] - \mathbb{E}[w]^2 = \text{Var}[w]. \end{aligned}$$

The *non-dimensional* relationship would then be:

$$\frac{1}{(1 - \tilde{\sigma}_w^2)} = \alpha \left(\hat{w}_1^2 + \frac{\tilde{\sigma}_w^2}{(1 - \tilde{\sigma}_w^2)^2} \right) + (1 - \alpha) \left(\hat{w}_2^2 + \frac{\tilde{\sigma}_w^2}{(1 - \tilde{\sigma}_w^2)^2} \right). \quad (4.3.4)$$

The third order moment is given as:

$$\begin{aligned}\overline{w'^3} &= (1 - \delta)\alpha[(w_1 - \bar{w})^3 + 3\sigma_w^2(w_1 - \bar{w})] \\ &+ (1 - \delta)(1 - \alpha)[(w_2 - \bar{w})^3 + 3\sigma_w^2(w_2 - \bar{w})]\end{aligned}\quad (4.3.5)$$

Since we want to make use of the specific shape of the pdf, we also have a relationship for $\overline{w'^3}$, which is called \widehat{Sk}_w , meaning the skewness of the variable w :

$$\begin{aligned}\widehat{Sk}_w &\equiv \frac{1}{(1 - \tilde{\sigma}_w^2)^{3/2}} \frac{\overline{w'^3}}{\left(\overline{w'^2}\right)^{3/2}} \frac{1}{\left(\frac{1-\delta\lambda_w}{1-\delta}\right)^{3/2}} \frac{1}{1 - \delta} \\ &= \alpha \left(\widehat{w}_1^3 + 3\widehat{w}_1 \frac{\tilde{\sigma}_w^2}{(1 - \tilde{\sigma}_w^2)} \right) + (1 - \alpha) \left(\widehat{w}_2^3 + 3\widehat{w}_2 \frac{\tilde{\sigma}_w^2}{(1 - \tilde{\sigma}_w^2)} \right)\end{aligned}\quad (4.3.6)$$

4.3.2 Moments for θ_l

For θ_l we have a similar *non-dimensional* relationship:

$$0 = \alpha\tilde{\theta}_{l1} + (1 - \alpha)\tilde{\theta}_{l2}\quad (4.3.7)$$

Similarly but with a different standard deviation, $\overline{\theta_l'^2}$ is given as:

$$\begin{aligned}\overline{\theta_l'^2} &= (1 - \delta)\alpha[(\theta_{l1} - \bar{\theta}_l)^2 + \sigma_{\theta_{l1}}^2] \\ &+ (1 - \delta)(1 - \alpha)[(\theta_{l2} - \bar{\theta}_l)^2 + \sigma_{\theta_{l2}}^2] \\ &+ \delta\sigma_{\theta_{l3}}^2,\end{aligned}\quad (4.3.8)$$

where $\sigma_{\theta_{l3}}$ is defined as $\lambda_{\theta_l}\overline{\theta_l'^2}$. This can also be expressed as the variance following the same approach as the one for $\overline{w'^2}$.

The third order moment is given as:

$$\begin{aligned}\overline{\theta_l'^3} &= (1 - \delta)\alpha[(\theta_{l1} - \bar{\theta}_l)^3 + 3\sigma_{\theta_{l1}}^2(\theta_{l1} - \bar{\theta}_l)] \\ &+ (1 - \delta)(1 - \alpha)[(\theta_{l2} - \bar{\theta}_l)^3 + 3\sigma_{\theta_{l2}}^2(\theta_{l2} - \bar{\theta}_l)]\end{aligned}\quad (4.3.9)$$

Similarly to equation (4.3.6), we also list a moment which is more diagnosed than prognosed:

$$\begin{aligned}\widehat{Sk_{\theta_l}} &\equiv \frac{\overline{\theta_l^3}}{\left(\overline{\theta_l'^2}\right)^{3/2}} \left(\frac{1}{\frac{1-\delta\lambda_\theta}{1-\delta}}\right)^{3/2} \frac{1}{1-\delta} \\ &= \alpha \left(\tilde{\theta}_{l1}^3 + 3\tilde{\theta}_{l1}\tilde{\sigma}_{\theta_{l1}}^2\right) + (1-\alpha) \left(\tilde{\theta}_{l2}^3 + 3\tilde{\theta}_{l2}\tilde{\sigma}_{\theta_{l2}}^2\right).\end{aligned}\quad (4.3.10)$$

4.3.3 Moments for r_t

The relationships for r_t and $\overline{r_t'^2}$ are given as follows

$$0 = \alpha\tilde{r}_{t1} + (1-\alpha)\tilde{r}_{t2}, \quad (4.3.11)$$

and

$$1 = \alpha \left(\tilde{r}_{t1}^2 + \tilde{\sigma}_{r_{t1}}^2\right) + (1-\alpha) \left(\tilde{r}_{t2}^2 + \tilde{\sigma}_{r_{t2}}^2\right). \quad (4.3.12)$$

Since this relationship is similar to the relationships of θ_l and $\theta_l'^2$, we are using nearly the same formulas:

$$\begin{aligned}\overline{r_t'^2} &= (1-\delta)\alpha[(r_{t1} - \overline{r_t})^2 + \sigma_{r_{t1}}^2] \\ &\quad + (1-\delta)(1-\alpha)[(r_{t2} - \overline{r_t})^2 + \sigma_{\theta_{l2}}^2] \\ &\quad + \delta\sigma_{r_{t3}}^2,\end{aligned}\quad (4.3.13)$$

and

$$\begin{aligned}\overline{r_t'^3} &= (1-\delta)\alpha[(r_{t1} - \overline{r_t})^3 + 3\sigma_{r_{t1}}^2(r_{t1} - \overline{r_t})] \\ &\quad + (1-\delta)(1-\alpha)[(r_{t2} - \overline{r_t})^3 + 3\sigma_{r_{t2}}^2(r_{t2} - \overline{r_t})]\end{aligned}\quad (4.3.14)$$

4.3.4 Mixed moments

There are also equations for two or even three variables, which are listed in the following.

$$\begin{aligned}\overline{w'\theta'_l} &= (1 - \delta)\alpha[(w_1 - \bar{w})(\theta_{l1} - \bar{\theta}_l)] \\ &\quad + (1 - \delta)(1 - \alpha)[(w_2 - \bar{w})(\theta_{l2} - \bar{\theta}_l)] \\ &\quad + \delta\lambda_{w\theta}\overline{w'\theta'_l},\end{aligned}\tag{4.3.15}$$

$$\begin{aligned}\overline{w'r'_t} &= (1 - \delta)\alpha[(w_1 - \bar{w})(r_{t1} - \bar{r}_t)] \\ &\quad + (1 - \delta)(1 - \alpha)[(w_2 - \bar{w})(r_{t2} - \bar{r}_t)] \\ &\quad + \delta\lambda_{wr}\overline{w'r'_t},\end{aligned}\tag{4.3.16}$$

and

$$\begin{aligned}\overline{r'_t\theta'_l} &= (1 - \delta)\alpha[(r_{t1} - \bar{r}_t)(\theta_{l1} - \bar{\theta}_l) + r_{r_t\theta_l}\sigma_{r_{t1}}\sigma_{\theta_{l1}}] \\ &\quad + (1 - \delta)(1 - \alpha)[(r_{t2} - \bar{r}_t)(\theta_{l2} - \bar{\theta}_l) + r_{r_t\theta_l}\sigma_{r_{t2}}\sigma_{\theta_{l2}}] \\ &\quad + \delta\lambda_{r\theta}\overline{r'_t\theta'_l}.\end{aligned}\tag{4.3.17}$$

We have the *non-dimensional* relationship for those moments given as

$$\begin{aligned}\hat{c}_{w\theta_l} &\equiv \frac{1}{(1 - \tilde{\sigma}_w^2)^{1/2}} \frac{\overline{w'\theta'_l}}{\sqrt{w'^2}\sqrt{\theta_l'^2}} \frac{1}{\sqrt{\frac{1-\delta\lambda_w}{1-\delta}}} \frac{1}{\sqrt{\frac{1-\delta\lambda_\theta}{1-\delta}}} \frac{1 - \delta\lambda_{w\theta}}{1 - \delta} \\ &= \alpha\hat{w}_1\tilde{\theta}_{l1} + (1 - \alpha)\hat{w}_2\tilde{\theta}_{l2},\end{aligned}\tag{4.3.18}$$

$$\begin{aligned}\hat{c}_{wr_t} &\equiv \frac{1}{(1 - \tilde{\sigma}_w^2)^{1/2}} \frac{\overline{w'r'_t}}{\sqrt{w'^2}\sqrt{r_t'^2}} \frac{1}{\sqrt{\frac{1-\delta\lambda_w}{1-\delta}}} \frac{1}{\sqrt{\frac{1-\delta\lambda_r}{1-\delta}}} \frac{1 - \delta\lambda_{wr}}{1 - \delta} \\ &= \alpha\hat{w}_1\tilde{r}_{t1} + (1 - \alpha)\hat{w}_2\tilde{r}_{t2},\end{aligned}\tag{4.3.19}$$

and

$$\begin{aligned}\hat{c}_{r_t\theta_l} &\equiv \frac{\overline{r'_t\theta'_l}}{\sqrt{r_t'^2}\sqrt{\theta_l'^2}} \frac{1}{\sqrt{\frac{1-\delta\lambda_r}{1-\delta}}} \frac{1}{\sqrt{\frac{1-\delta\lambda_\theta}{1-\delta}}} \frac{1 - \delta\lambda_{r\theta}}{1 - \delta} \\ &= \alpha\left(\tilde{r}_{t1}\tilde{\theta}_{l1} + r_{r_t\theta_l}\tilde{\sigma}_{r_{t1}}\tilde{\sigma}_{\theta_{l1}}\right) + (1 - \alpha)\left(\tilde{r}_{t2}\tilde{\theta}_{l2} + r_{r_t\theta_l}\tilde{\sigma}_{r_{t2}}\tilde{\sigma}_{\theta_{l2}}\right),\end{aligned}\tag{4.3.20}$$

where we can think about \hat{c} as the correlation.

We also list a trivariate moment $(\overline{w'r'_t\theta'_l})$, given by:

$$\begin{aligned}\overline{w'r'_t\theta'_l} &= (1 - \delta)\alpha(w_1 - \bar{w}) \left[(r_{t1} - \bar{r}_t) (\theta_{l1} - \bar{\theta}_l) + r_{r_t\theta_l} \sigma_{r_{t1}} \sigma_{\theta_{l1}} \right] \\ &+ (1 - \delta)(1 - \alpha)(w_2 - \bar{w}) \left[(r_{t2} - \bar{r}_t) (\theta_{l2} - \bar{\theta}_l) + r_{r_t\theta_l} \sigma_{r_{t2}} \sigma_{\theta_{l2}} \right].\end{aligned}\quad (4.3.21)$$

4.4 Solving for pdf parameters by using the moment terms

Having established the prognosed moments for the desired pdf, we now try to retrieve the specific pdf that generates these moments. This process essentially involves inverting the relationship between the moments and the parameters that define the pdf.

In our case, we refer back to the normal mixture family of pdfs (equation (4.1.1)), which offers a representation for atmospheric grid layers. To select a particular member within this family that best aligns with the prognosed moments, we perform a parameter retrieval step.

This retrieval is achieved by inverting equations (4.3.2) to (4.3.20). These equations express the prognosed moments (mean, variance, covariances, etc.) as functions of the underlying pdf parameters (weights, means, and standard deviations). By inverting these relationships, we aim to find a set of pdf parameters that produces a distribution with moments matching the prognosed values.

However, it is important to mention that this inversion is not a straightforward process. That is because the equations are non-linear with respect to the pdf parameters. Despite this non-linearity, the relatively simple structure of the normal mixture pdf (equation (4.1.1)) allows for an analytical solution to the inversion problem. This analytical solution enables

us to efficiently map the prognosed moments back to the corresponding pdf parameters.

The proposed solution procedure¹ is as follows.

1. Solve for α , \widehat{w}_1 , and \widehat{w}_2 from the equations for \overline{w} (equation (4.3.2)), $\overline{w'^2}$ (equation (4.3.4)), $\overline{w'^3}$ (equation (4.3.6)):

$$\alpha = \frac{1}{2} \left[1 - \widehat{S}k_w \sqrt{\frac{1}{4 + \widehat{S}k_w^2}} \right], \quad (4.4.1)$$

$$\widehat{w}_1 = \sqrt{\frac{1 - \alpha}{\alpha}}, \quad (4.4.2)$$

$$\widehat{w}_2 = -\sqrt{\frac{\alpha}{1 - \alpha}}. \quad (4.4.3)$$

Without loss of generality, it has been chosen to set $\widehat{w}_1 > \widehat{w}_2$.

2. Looking at equation (4.4.1), we see that $\widehat{S}k_w$ is determined only by α :

$$\widehat{S}k_w = \frac{1 - 2\alpha}{\sqrt{\alpha(1 - \alpha)}}. \quad (4.4.4)$$

3. $\tilde{\theta}_{l1}$ and $\tilde{\theta}_{l2}$ are taken from solving equation (4.3.7) for $\overline{\theta}_l$, and equation (4.3.18) for $\overline{w'\theta'_l}$:

$$\tilde{\theta}_{l1} = -\frac{\widehat{c}_{w\theta_l}}{\widehat{w}_2}, \quad (4.4.5)$$

$$\tilde{\theta}_{l2} = -\frac{\widehat{c}_{w\theta_l}}{\widehat{w}_1}. \quad (4.4.6)$$

4. We can get $\tilde{\sigma}_{\theta_{l1}}$ and $\tilde{\sigma}_{\theta_{l2}}$ by fulfilling equation (4.3.8) for $\overline{\theta_l'^2}$, and equation (4.3.9) for $\overline{\theta_l'^3}$:

$$\tilde{\sigma}_{\theta_{l1}}^2 = (1 - \widehat{c}_{w\theta_l}^2) + \left(\sqrt{\frac{1 - \alpha}{\alpha}} \right) \frac{1}{3\widehat{c}_{w\theta_l}} \left(\widehat{S}k_{\theta_l} - \widehat{c}_{w\theta_l}^3 \widehat{S}k_w \right), \quad (4.4.7)$$

¹Larson and Golaz, “Using probability density functions to derive consistent closure relationships among higher-order moments”.

$$\tilde{\sigma}_{\theta_l 2}^2 = (1 - \hat{c}_{w\theta_l}^2) - \left(\sqrt{\frac{\alpha}{1 - \alpha}} \right) \frac{1}{3\hat{c}_{w\theta_l}} \left(\widehat{Sk}_{\theta_l} - \hat{c}_{w\theta_l}^3 \widehat{Sk}_w \right). \quad (4.4.8)$$

\widehat{Sk}_{θ_l} represents the skewness of θ_l , which has to be provided by an equation such as equation (4.3.10) below.

5. Finding formulas for \tilde{r}_{t1} , \tilde{r}_{t2} , $\tilde{\sigma}_{r_t 1}^2$, and $\tilde{\sigma}_{r_t 2}^2$ can be done by replacing θ_l by r_t everywhere in the equations (4.4.5), (4.4.6), (4.4.7), and (4.4.8).
6. The last step is to get a relationship between $r_{r_t \theta_l}$, the in-between normal correlation and $c_{r_t \theta_l}$, the total correlation. This can be done by using equation (4.3.17):

$$r_{r_t \theta_l} = \frac{\hat{c}_{r_t \theta_l} - \hat{c}_{wr_t} \hat{c}_{w\theta_l}}{\alpha \tilde{\sigma}_{r_t 1} \tilde{\sigma}_{\theta_l 1} + (1 - \alpha) \tilde{\sigma}_{r_t 2} \tilde{\sigma}_{\theta_l 2}}. \quad (4.4.9)$$

4.5 Leveraging the pdf for higher-order moment calculations

Upon determining the pdf parameters, we gain the ability to compute all higher-order moments associated with the distribution. These moments play a crucial role for closing the already described pdes. The symbolic calculation of higher-order moments can be achieved through integration over the specified pdf. Formulas for calculating various higher-order moments within the context of a binormal pdf are readily available in the literature².

We state the transformed formulas needed for closure in the following.

²Larson and Golaz, “Using probability density functions to derive consistent closure relationships among higher-order moments”.

$$\begin{aligned}
\frac{1}{(1 - \tilde{\sigma}_w^2)^2} \frac{(1 - \delta)}{(1 - \delta\lambda_w)^2} \frac{\overline{w'^4}}{(\overline{w'^2})^2} &= \alpha \left[\widehat{w}_1^4 + 6\widehat{w}_1^2 \frac{\tilde{\sigma}_w^2}{(1 - \tilde{\sigma}_w^2)} + 3 \frac{\tilde{\sigma}_w^4}{(1 - \tilde{\sigma}_w^2)^2} \right] \\
&+ (1 - \alpha) \left[\widehat{w}_2^4 + 6\widehat{w}_2^2 \frac{\tilde{\sigma}_w^2}{(1 - \tilde{\sigma}_w^2)} + 3 \frac{\tilde{\sigma}_w^4}{(1 - \tilde{\sigma}_w^2)^2} \right] \\
&+ \frac{1}{(1 - \tilde{\sigma}_w^2)^2} \frac{(1 - \delta)}{(1 - \delta\lambda_w)^2} \delta 3\lambda_w^2, \tag{4.5.1}
\end{aligned}$$

$$\begin{aligned}
\frac{1}{(1 - \tilde{\sigma}_w^2)} \frac{(1 - \delta)^{1/2}}{(1 - \delta\lambda_w)(1 - \delta\lambda_\theta)^{1/2}} \frac{\overline{w'^2\theta'_l}}{w'^2 (\overline{\theta'^2})^{1/2}} &= \alpha \left[\widehat{w}_1^2 + \frac{\tilde{\sigma}_w^2}{(1 - \tilde{\sigma}_w^2)} \right] \tilde{\theta}_{l1} \\
&+ (1 - \alpha) \left[\widehat{w}_2^2 + \frac{\tilde{\sigma}_w^2}{(1 - \tilde{\sigma}_w^2)} \right] \tilde{\theta}_{l2}, \tag{4.5.2}
\end{aligned}$$

$$\frac{1}{(1 - \tilde{\sigma}_w^2)^{1/2}} \frac{(1 - \delta)^{1/2}}{(1 - \delta\lambda_w)^{1/2}(1 - \delta\lambda_\theta)} \frac{\overline{w'\theta'_l{}^2}}{(\overline{w'^2})^{1/2} \overline{\theta'^2}} = \alpha \widehat{w}_1 \left(\tilde{\theta}_{l1}^2 + \tilde{\sigma}_{\theta_{l1}}^2 \right) + (1 - \alpha) \widehat{w}_2 \left(\tilde{\theta}_{l2}^2 + \tilde{\sigma}_{\theta_{l2}}^2 \right), \tag{4.5.3}$$

$$\begin{aligned}
\frac{1}{(1 - \tilde{\sigma}_w^2)^{1/2}} \frac{(1 - \delta)^{1/2}}{(1 - \delta\lambda_w)^{1/2}(1 - \delta\lambda_\theta)^{1/2}(1 - \delta\lambda_{q_t})^{1/2}} \frac{\overline{w'r'_t\theta'_l}}{(\overline{w'^2})^{1/2} (\overline{r'^2_t})^{1/2} (\overline{\theta'^2_l})^{1/2}} \\
= \alpha \widehat{w}_1 \left(\tilde{r}_{t1} \tilde{\theta}_{l1} + r_{rt\theta_l} \tilde{\sigma}_{r_{t1}} \tilde{\sigma}_{\theta_{l1}} \right) + (1 - \alpha) \widehat{w}_2 \left(\tilde{r}_{t2} \tilde{\theta}_{l2} + r_{rt\theta_l} \tilde{\sigma}_{r_{t2}} \tilde{\sigma}_{\theta_{l2}} \right) \tag{4.5.4}
\end{aligned}$$

Equations for $\overline{w'^2 r'_t}$ and $\overline{w' r'^2_t}$ are similar to equation (4.5.2) and equation (4.5.3) by replacing θ_l with r_t everywhere.

4.6 Addressing skewness and achieving closure

Closing the system of prognostic equations within CLUBB needs the specification of the skewness terms Sk_{θ_l} and Sk_{r_t} . These skewness values appear in the solutions for $\tilde{\sigma}_{\theta_{l1}}$ (equation (4.4.7)) and $\tilde{\sigma}_{\theta_{l2}}$ (equation (4.4.8)), respectively. Traditionally, these skewness terms

could be treated as prognostic variables, requiring their own prognostic equations and adding to the overall intense computation.

Therefore, the paper³ which this work is based on, proposes an alternative approach that uses a diagnostic formula for skewness. The formula provides a reasonable estimate of the skewness terms based on the readily available prognostic moments, avoiding the need for dedicated prognostic equations for skewness. This strategy results in closure of the system of equations while maintaining a computationally tractable model. The proposed formula is the following:

$$\widehat{Sk}_{\theta_l} = \widehat{Sk}_w \widehat{c}_{w\theta_l} [\beta + (1 - \beta) \widehat{c}_{w\theta_l}^2], \quad (4.6.1)$$

which is similar for \widehat{Sk}_{r_t} by again just replacing r_t with θ_l . They define a parameter β which is dimensionless. We also solve for β because we are going to need the equation later on to show that other equations are true.

$$\implies \beta = \frac{\frac{\widehat{Sk}_{\theta_l}}{\widehat{Sk}_w \widehat{c}_{w\theta_l}} - \widehat{c}_{w\theta_l}^2}{1 - \widehat{c}_{w\theta_l}^2} \quad (4.6.2)$$

Equation (4.6.1) presents a diagnostic formula for estimating the skewness of θ_l . This formula offers a physically intuitive relationship. It proposes a proportionality between Sk_{θ_l} and Sk_{r_t} . However, it's crucial to acknowledge the limitations in this diagnostic. The formula suggests that an increase in the parameter β leads to a larger magnitude of Sk_{θ_l} . This translates to a pdf with a more extended tail in the θ_l domain. Furthermore, the formula captures the behavior when w and θ_l are correlated. That is, positive skewness in w leads to positive skewness in θ_l (positive correlation), and vice versa (negative correlation). However, it is important to mention that real-world large eddy simulations may show deviations from this simplified relationship.

Another limitation appears when either Sk_w or the covariance between w and θ_l ($c_{w\theta_l}$)

³Larson and Golaz, "Using probability density functions to derive consistent closure relationships among higher-order moments".

approaches zero. The formula predicts a vanishing Sk_{θ_l} in these scenarios, which may not always be true.

Finally, the diagnostic approach allows for the magnitude of $|Sk_{\theta_l}|$ to be either smaller or larger than $|Sk_w|$. This behavior depends on the interplay between the variance of w ($\tilde{\sigma}_w^2$), the covariance ($c_{w\theta_l}$), and the parameter β . This highlights the potential for discrepancies between the estimated skewness and the actual skewness observed in real-world atmospheric data.

To summarize, equation (4.6.1) offers a computationally efficient method for skewness estimation, but it comes with limitations. While it captures some key aspects of the relationship between the skewness in w and the skewness in θ_l , one should be aware of deviations from its predictions.

We proceed with using equation (4.3.10) for \widehat{Sk}_{θ_l} and find the following relationships⁴ for $\tilde{\sigma}_{\theta_l 1}^2$ and $\tilde{\sigma}_{\theta_l 2}^2$:

$$\tilde{\sigma}_{\theta_l 1}^2 = \frac{(1 - \widehat{c}_{w\theta_l}^2)}{\alpha} \left[\frac{1}{3}\beta + \alpha \left(1 - \frac{2}{3}\beta \right) \right], \quad (4.6.3)$$

and

$$\tilde{\sigma}_{\theta_l 2}^2 = \frac{(1 - \widehat{c}_{w\theta_l}^2)}{1 - \alpha} \left\{ 1 - \left[\frac{1}{3}\beta + \alpha \left(1 - \frac{2}{3}\beta \right) \right] \right\}. \quad (4.6.4)$$

By using the previously stated expressions for the standard deviations (equations (4.6.3) and (4.6.4), with their r_t counterparts), we can substitute them into the formula for the correlation between r_t and θ_l (equation (4.4.9)). This substitution leads to a more concise representation.

$$r_{r_t \theta_l} = \frac{c_{r_t \theta_l} - \widehat{c}_{wr_t} \widehat{c}_{w\theta_l}}{(1 - \widehat{c}_{wr_t}^2)^{1/2} (1 - \widehat{c}_{w\theta_l}^2)^{1/2}}, \quad (4.6.5)$$

⁴Larson and Golaz, “Using probability density functions to derive consistent closure relationships among higher-order moments”.

where the correlation of r_t and θ_l within the individual normal distributions is $r_{r_t\theta_l}$, and $c_{r_t\theta_l}$ represents the total correlation across the entire trinormal pdf.

CLUBB chose a specific formula for the w -“width” of the individual normals, which is the following.

$$\tilde{\sigma}_w^2 = \gamma [1 - \max(c_{w\theta_l}^2, c_{wr_t}^2)] . \quad (4.6.6)$$

The stated formula incorporates the covariances – $c_{wr_t}^2$ and $c_{w\theta_l}^2$ – and ensures that $\tilde{\sigma}_w^2$ is bounded between 0 and 1. That is because a new dimensionless parameter γ was introduced whose domain is $[0, 1)$. This mathematical safeguard guarantees that even when the original covariances ($c_{wr_t}^2$ and $c_{w\theta_l}^2$) become large in magnitude, the derived variance remains within a realistic and mathematically tractable bound ($0 \leq \tilde{c}_{w\theta_l}^2, \tilde{c}_{wr_t}^2 \leq 1$). It also follows, that $\tilde{\sigma}_{r_{ti}}$, and $r_{r_t\theta_l}$ remain in a realistic bound.

4.7 Formulating closure relationships for higher-order moments

This section delves into the derivation of closure relationships for crucial higher-order moments employed within the CLUBB parameterization.

Our focus here lies on achieving closure for the following terms:

- $\overline{w'^4}$: The fourth-order moment of up-/downdrafts,
- $\overline{w'^2\theta'_l}$: the so called flux,
- $\overline{w'\theta_l'^2}$: the counterpart of the previous one, where we replace every θ_l by r_t ,
- $\overline{w'r'_t\theta'_l}$: and the mean over all three standardized variables.

Closure, in this context, refers to expressing these higher-order moments completely in terms of known quantities, typically lower-order moments that are directly prognosed by the model. The approach to derive those formulas is based on the previously established expressions for the pdf parameters (equations (4.4.1) to (4.4.9)). By substituting those derived pdf parameter expressions into the relevant equations for the higher-order moments (equations (4.5.1) to (4.5.4)), we find the desired closure relationships.

We first present the equation for the third moment of θ'_l , $\overline{\theta'^3_l}$. This expression is derived by dimensionalizing equation (4.6.1), which relates the skewness of θ_l to the skewness of w and other model parameters.

$$\overline{\theta'^3_l} = \frac{(1 - \delta\lambda_{w\theta})(1 - \delta\lambda_\theta)}{(1 - \tilde{\sigma}_w^2)^2 (1 - \delta\lambda_w)^2} \frac{\overline{w'^3}}{(\overline{w'^2})^2} \overline{\theta'^2_l} \overline{w'\theta'_l} \left(\beta + (1 - \beta) \frac{(1 - \delta\lambda_{w\theta})^2}{1 - \tilde{\sigma}_w^2 (1 - \delta\lambda_w)(1 - \delta\lambda_\theta)} \frac{(\overline{w'\theta'_l})^2}{\overline{w'^2} \overline{\theta'^2_l}} \right). \quad (4.7.1)$$

While the scalar third moments (e.g., $\overline{\theta'^3_l}$) may not directly participate in solving the prognostic equations within CLUBB, they hold an indirect yet crucial role in shaping cloud properties within atmospheric simulations. This influence is coming from the connection between the pdf and the cloud formation.

Cumulus cloud formation mostly occurs at the edges, or “tails” of the pdf for a specific variable. These tails represent regions where the probability of encountering extreme values of the variable is relatively higher. As the relative “width” of the normal distribution representing w increases, the magnitude of $\overline{\theta'^3_l}$ also grows (refer to (4.7.1) for details). In simpler terms, a larger value of $\overline{\theta'^3_l}$ corresponds to a broader pdf for the up-/downdraft variable. This broader pdf deviates more significantly from a double delta function, which is a construct with two spikes at zero.

Unlike the scalar third moment, $\overline{w'^4}$ does not depend on the thermodynamic scalar moments (such as $\overline{\theta'^3_l}$). Consequently, it is independent of the parameter β .

To derive the explicit formula for $\overline{w'^4}$, we can substitute the previously stated expressions for \widehat{w}_1 (equation (4.4.2)) and \widehat{w}_2 (equation (4.4.3)) into equation (4.5.1).

$$\begin{aligned} \frac{1}{(1 - \tilde{\sigma}_w^2)^2} \frac{(1 - \delta)}{(1 - \delta\lambda_w)^2} \frac{\overline{w'^4}}{\left(\overline{w'^2}\right)^2} &= 3 \frac{\tilde{\sigma}_w^4}{(1 - \tilde{\sigma}_w^2)^2} + 6 \frac{\tilde{\sigma}_w^2}{(1 - \tilde{\sigma}_w^2)} + 1 \\ &+ \widehat{Sk}_w^2 + \frac{1}{(1 - \tilde{\sigma}_w^2)^2} \frac{(1 - \delta)}{(1 - \delta\lambda_w)^2} \delta 3\lambda_w^2, \end{aligned} \quad (4.7.2)$$

and also

$$\begin{aligned} \overline{w'^4} &= \left(\overline{w'^2}\right)^2 \frac{(1 - \delta\lambda_w)^2}{(1 - \delta)} \left(3\tilde{\sigma}_w^4 + 6(1 - \tilde{\sigma}_w^2)\tilde{\sigma}_w^2 + (1 - \tilde{\sigma}_w^2)^2\right) \\ &+ \frac{1}{(1 - \tilde{\sigma}_w^2)} \frac{1}{(1 - \delta\lambda_w)} \frac{\left(\overline{w'^3}\right)^2}{\overline{w'^2}} \\ &+ \delta 3\lambda_w^2 \left(\overline{w'^2}\right)^2. \end{aligned} \quad (4.7.3)$$

As observed with $\overline{w'^4}$, $\overline{w'^2\theta'_l}$ displays independence from the parameter β within the context of the chosen pdf.

To proceed, we can substitute the previously derived expressions for \widehat{w}_1 (equation (4.4.2)), \widehat{w}_2 (equation (4.4.3)), \widehat{Sk}_w (equation (4.3.6)), $\widehat{\theta}_{l1}$ (equation (4.4.5)), and $\widehat{\theta}_{l2}$ (equation (4.4.6)) into equation (4.5.2). This substitution process will yield an explicit formula for $\overline{w'^2\theta'_l}$ that solely relies on known quantities, such as the prognostic moments directly calculated by the model.

$$\frac{1}{(1 - \tilde{\sigma}_w^2)} \frac{(1 - \delta)^{1/2}}{(1 - \delta\lambda_w)(1 - \delta\lambda_\theta)^{1/2}} \frac{\overline{w'^2\theta'_l}}{\overline{w'^2} \left(\overline{\theta'^2_l}\right)^{1/2}} = \widehat{c}_{w\theta_l} \widehat{Sk}_w, \quad (4.7.4)$$

and

$$\overline{w'^2\theta'_l} = \frac{1}{(1 - \tilde{\sigma}_w^2)} \frac{1 - \delta\lambda_{w\theta}}{1 - \delta\lambda_w} \frac{\overline{w'^3}}{\overline{w'^2}} \overline{w'\theta'_l}. \quad (4.7.5)$$

$\overline{w'\theta'^2_l}$ depends explicitly on Sk_{θ_l} . Substituting equation (4.4.2) - equation (4.4.8) into equation (4.5.4) yields

$$\frac{1}{(1 - \tilde{\sigma}_w^2)^{1/2}} \frac{(1 - \delta)^{1/2}}{(1 - \delta\lambda_w)^{1/2}(1 - \delta\lambda_\theta)} \frac{\overline{w'\theta'^2_l}}{\left(\overline{w'^2}\right)^{1/2} \overline{\theta'^2_l}} = \frac{2}{3} \widehat{c}_{w\theta_l}^2 \widehat{Sk}_w + \frac{1}{3} \frac{\widehat{Sk}_{\theta_l}}{\widehat{c}_{w\theta_l}}, \quad (4.7.6)$$

and

$$\begin{aligned}\overline{w'\theta_l'^2} &= \frac{2}{3} \frac{(1 - \delta\lambda_{w\theta})^2}{(1 - \delta\lambda_w)^2} \frac{1}{(1 - \tilde{\sigma}_w^2)^2} \frac{\overline{w'^3}}{(\overline{w'^2})^2} (\overline{w'\theta_l'})^2 \\ &+ \frac{1}{3} \frac{(1 - \delta\lambda_w)}{(1 - \delta\lambda_{w\theta})} (1 - \tilde{\sigma}_w^2) \frac{\overline{w'^2} \overline{\theta_l'^3}}{\overline{w'\theta_l'}}.\end{aligned}\quad (4.7.7)$$

PROCEED HERE!

The formula has a problem because $\overline{w'\theta_l'}$ is in the denominator. As $\overline{w'\theta_l'}$ gets closer to zero, the formula becomes infinitely large, which is called a singularity. This can cause issues if we use the formula directly with real-world measurements of $\overline{\theta_l'^3}$ and $\overline{w'\theta_l'}$. The resulting diagnosis of $\overline{w'\theta_l'^2}$ would be very sensitive to small changes in the measurements and might not be reliable (noisy). We can fix this singularity by either

- substitute in the ansatz for Sk_{θ_l} (equation (4.3.10)) into the original formula (equation (4.5.3)),
- or, equivalently, substitute equation (4.7.1) for $\overline{\theta_l'^3}$. This is possible because equation (4.7.1) shows that $\overline{\theta_l'^3}$ is proportional to $\overline{w'\theta_l'}$.

Both approaches effectively remove the singularity from the formula. Therefore, we find:

$$\frac{1}{(1 - \tilde{\sigma}_w^2)^{1/2}} \frac{(1 - \delta)^{1/2}}{(1 - \delta\lambda_w)^{1/2}(1 - \delta\lambda_\theta)} \frac{\overline{w'\theta_l'^2}}{(\overline{w'^2})^{1/2} \overline{\theta_l'^2}} = \widehat{Sk}_w \left[\frac{1}{3}\beta + \left(1 - \frac{1}{3}\beta\right) \widehat{c}_{w\theta_l}^2 \right], \quad (4.7.8)$$

and

$$\overline{w'\theta_l'^2} = \frac{1}{(1 - \tilde{\sigma}_w^2)} \frac{(1 - \delta\lambda_\theta)}{(1 - \delta\lambda_w)} \frac{\overline{w'^3}}{\overline{w'^2}} \left[\frac{1}{3}\beta \overline{\theta_l'^2} + \frac{(1 - \frac{1}{3}\beta)}{(1 - \tilde{\sigma}_w^2)} \frac{(1 - \delta\lambda_{w\theta})^2}{(1 - \delta\lambda_w)(1 - \delta\lambda_\theta)} \frac{(\overline{w'\theta_l'})^2}{\overline{w'^2}} \right]. \quad (4.7.9)$$

Finally, substituting equation (4.4.2) - equation (4.4.9) into equation (4.5.4) yields the fol-

lowing formula for the turbulent flux of $\overline{r'_t \theta'_l}$, $\overline{w' r'_t \theta'_l}$:

$$\begin{aligned} & \frac{(1 - \delta)^{1/2}}{(1 - \delta \lambda_w)^{1/2} (1 - \delta \lambda_\theta)^{1/2} (1 - \delta \lambda_{r_t})^{1/2}} \frac{\overline{w' r'_t \theta'_l}}{(1 - \tilde{\sigma}_w^2)^{1/2} \left(\overline{w'^2}\right)^{1/2} \left(\overline{r_t'^2}\right)^{1/2} \left(\overline{\theta_l'^2}\right)^{1/2}} \\ &= \widehat{c}_{wr_t} \widehat{c}_{w\theta_l} \widehat{Sk}_w + E(w, q_t, \theta_l) \frac{1}{2} \widehat{Sk}_w (c_{q_t \theta_l} - \widehat{c}_{wr_t} \widehat{c}_{w\theta_l}), \end{aligned} \quad (4.7.10)$$

and

$$\begin{aligned} \overline{w' r'_t \theta'_l} &= \frac{\frac{1}{2} E}{(1 - \tilde{\sigma}_w^2)} \frac{(1 - \delta \lambda_{\theta_q})}{(1 - \delta \lambda_w)} \overline{r'_t \theta'_l} \frac{\overline{w'^3}}{\overline{w'^2}} \\ &+ \frac{1 - \frac{1}{2} E}{(1 - \tilde{\sigma}_w^2)^2} \frac{(1 - \delta \lambda_{wq})(1 - \delta \lambda_{w\theta})}{(1 - \delta \lambda_w)^2} \overline{w' r'_t} \overline{w' \theta'_l} \frac{\overline{w'^3}}{\left(\overline{w'^2}\right)^2}. \end{aligned} \quad (4.7.11)$$

The function $E(w, r_t, \theta_l)$ is

$$E = \frac{1 - \frac{1}{2} \frac{2\alpha}{1-2\alpha} \xi}{1 + \frac{1}{2} \xi}, \quad (4.7.12)$$

where

$$1 + \xi = \frac{1 - \alpha}{\alpha} \frac{\tilde{\sigma}_{r_{t2}}}{\tilde{\sigma}_{r_{t1}}} \frac{\tilde{\sigma}_{\theta_{l2}}}{\tilde{\sigma}_{\theta_{l1}}} = \left(\frac{A_{r_t} - B_{r_t}}{-A_{r_t} - B_{r_t}} \right)^{1/2} \left(\frac{A_{\theta_l} - B_{\theta_l}}{-A_{\theta_l} - B_{\theta_l}} \right)^{1/2}, \quad (4.7.13)$$

and

$$A_{\theta_l} = Sk_{\theta_l} - \frac{3}{2} \widehat{c}_{w\theta_l} \widehat{Sk}_w + \frac{1}{2} \widehat{c}_{w\theta_l}^3 \widehat{Sk}_w, \quad (4.7.14)$$

$$B_{\theta_l} = \frac{3}{2} \left(4 + \widehat{Sk}_w^2 \right)^{1/2} \widehat{c}_{w\theta_l} (1 - \widehat{c}_{w\theta_l}^2), \quad (4.7.15)$$

and A_{r_t} and B_{r_t} are analogous. We now list two cases in which the expression for E simplifies.

First, if

$$\frac{\tilde{\sigma}_{r_{t2}}}{\tilde{\sigma}_{r_{t1}}} \frac{\tilde{\sigma}_{\theta_{l2}}}{\tilde{\sigma}_{\theta_{l1}}} = 1$$

then $E = 0$. This would occur, for instance, if the “widths” of the first and second normal were equal to each other for both r_t and θ_l , that is, if $\tilde{\sigma}_{r_{t2}} = \tilde{\sigma}_{r_{t1}}$ and $\tilde{\sigma}_{\theta_{l2}} = \tilde{\sigma}_{\theta_{l1}}$. Second, if we use the diagnostic ansatz (equation (4.3.10)) for the scalar skewnesses, then

$$\xi = \frac{1 - 2\zeta}{\zeta}, \quad (4.7.16)$$

where

$$\zeta = \alpha + \frac{1}{3}\beta(1 - 2\alpha). \quad (4.7.17)$$

Then we find

$$E = \frac{2}{3}\beta, \quad (4.7.18)$$

and finally

$$\begin{aligned} \overline{w'r'_t\theta'_l} &= \frac{\frac{1}{3}\beta}{(1 - \tilde{\sigma}_w^2)} \frac{(1 - \delta\lambda_{\theta r})}{(1 - \delta\lambda_w)} \overline{r'_t\theta'_l} \frac{\overline{w'^3}}{\overline{w'^2}} \\ &+ \frac{1 - \frac{1}{3}\beta}{(1 - \tilde{\sigma}_w^2)^2} \frac{(1 - \delta\lambda_{wr})(1 - \delta\lambda_{w\theta})}{(1 - \delta\lambda_w)^2} \overline{w'r'_t} \overline{w'\theta'_l} \frac{\overline{w'^3}}{\left(\overline{w'^2}\right)^2}. \end{aligned} \quad (4.7.19)$$

5 Integration using SymPy

Throughout this thesis, symbolic manipulation plays a crucial role in verifying mathematical expressions, particularly integrals. To achieve this, we rely on SymPy¹ – a powerful Python library for symbolic mathematics. This chapter dives into the world of SymPy, showcasing its capabilities through a detailed example.

We begin by demonstrating the analytical approach to solving an integral. Next, we will explore the numerical side of integration. We are going to demonstrate how SymPy can be seamlessly integrated with numerical computing libraries to evaluate the integral for specific input values. This combined approach allows us to not only verify our analytical solution but also gain valuable insights into the integral’s behavior for different scenarios. By following this step-by-step example, the reader will gain a solid understanding of how SymPy can be used, not only for this thesis.

5.1 Analytic integration

For simplicity and readability, we choose the check for the formula of $\overline{w'^2}$ (this is item 3 from section 2.6). One starts by importing and – obviously – installing the packages if they are not there yet. Importing the package `display` is useful for later on printing the equations. Thus, this results in the code in listing 5.1. In this listing, `sympy` was defined to be called `sp` and from `sympy` we directly imported some packages, too, which are needed later on.

¹Aaron Meurer et al. “SymPy: symbolic computing in Python”. In: *PeerJ Computer Science* 3 (Jan. 2017), e103. ISSN: 2376-5992. DOI: 10.7717/peerj-cs.103. URL: <https://doi.org/10.7717/peerj-cs.103>.

Listing 5.1: Import statements

```
import sympy as sp
from IPython.display import display
from sympy import abc, oo, Symbol, Integral
from sympy.stats import Normal, density
```

Next, we define all symbols which are needed to calculate the given integral and therefore also to print the equations nicely. Since we are checking $\overline{w'^2}$, we need the (self-defined) symbols listed in listing 5.2. Having defined the symbols, we can proceed with defining

Listing 5.2: Defining symbols

```
sigma_w = Symbol('\sigma_w')
w_1 = Symbol('w_1')
w_2 = Symbol('w_2')
w_bar = Symbol('\overline{w}')
sigma_lambda_w = Symbol('\sigma_{\lambda w}')
w_prime_2_bar = Symbol('\overline{w\prime^2}')
```

the marginal distribution. Now, we are also using `sympy.abc` for displaying some standard symbols (listing 5.3). Having done that we can actually display the integral which we want

Listing 5.3: Defining the marginals

```
G_1_w = Normal(name='G_1_w', mean=w_1, std=sigma_w)
G_1_w_density = density(G_1_w)(sp.abc.w)
G_2_w = Normal(name='G_2_w', mean=w_2, std=sigma_w)
G_2_w_density = density(G_2_w)(sp.abc.w)
G_3_w = Normal(name='G_3_w', mean=w_bar, std=sigma_lambda_w)
G_3_w_density = density(G_3_w)(sp.abc.w)
G_w = ((1 - sp.abc.delta) * sp.abc.alpha * G_1_w_density +
        (1 - sp.abc.delta) * (1 - sp.abc.alpha) * G_2_w_density +
        sp.abc.delta * G_3_w_density)
```

to compute (listing 5.4). Looking at figure 5.1, this is exactly the integral which we want to compute. Using the command `.doit(conds='none')` in listing 5.5, we can actually calculate the given integral, where we assume that all given constants are real. We are also using `.simplify()` here to make the output more readable as well as more comparable to

Listing 5.4: Defining and displaying the needed integral

```
w_prime_2_bar_int = sp.Integral((sp.abc.w - w_bar) ** 2 * G_w, [sp.abc.w, -oo, oo])
display(sp.Eq(w_prime_2_bar, w_prime_2_bar_int))
```

Figure 5.1: Output of listing 5.4

$$\overline{w'^2} = \int_{-\infty}^{\infty} (-\bar{w} + w)^2 \left(\frac{\sqrt{2}\delta e^{-\frac{(-\bar{w}+w)^2}{2\sigma_{\lambda w}^2}}}{2\sqrt{\pi}\sigma_{\lambda w}} + \frac{\sqrt{2}\alpha(1-\delta)e^{-\frac{(w-w_1)^2}{2\sigma_w^2}}}{2\sqrt{\pi}\sigma_w} + \frac{\sqrt{2} \cdot (1-\alpha)(1-\delta)e^{-\frac{(w-w_2)^2}{2\sigma_w^2}}}{2\sqrt{\pi}\sigma_w} \right) dw$$

Listing 5.5: Calculating and printing the integral

```
w_prime_2_bar_int_val = w_prime_2_bar_int.doit(conds='none').simplify()
display(sp.Eq(w_prime_2_bar, w_prime_2_bar_int_val))
```

Figure 5.2: Output of listing 5.5

$$\begin{aligned} \overline{w'^2} = & -\bar{w}^2\delta + \bar{w}^2 + 2\bar{w}\alpha\delta w_1 - 2\bar{w}\alpha\delta w_2 - 2\bar{w}\alpha w_1 + 2\bar{w}\alpha w_2 + 2\bar{w}\delta w_2 - 2\bar{w}w_2 \\ & -\sigma_w^2\delta + \sigma_w^2 + \sigma_{\lambda w}^2\delta - \alpha\delta w_1^2 + \alpha\delta w_2^2 + \alpha w_1^2 - \alpha w_2^2 - \delta w_2^2 + w_2^2, \end{aligned}$$

the actual function we want to check. We can now compare figure 5.2 to the given equation.

To do this, we first need to define the equation for equation (4.3.3) in listing 5.6. We can

Listing 5.6: Python function for the second order moment

```
def w_prime_2_bar_check(delta=sp.abc.delta, alpha=sp.abc.alpha, w_1=w_1, w_2=w_2,
w_bar=w_bar, sigma_w=sigma_w, sigma_lambda_w=sigma_lambda_w):
    return (((1 - delta) * alpha * ((w_1 - w_bar) ** 2 + sigma_w ** 2))
            + ((1 - delta) * (1 - alpha) * ((w_2 - w_bar) ** 2 + sigma_w ** 2))
            + (delta * sigma_lambda_w ** 2))
```

print this equation using `display` again (listing 5.7). The last step is to check if those two

Listing 5.7: Printing the symbolic equation

```
display(sp.Eq(w_prime_2_bar, w_prime_2_bar_check()))
```

Figure 5.3: Output of listing 5.7

$$\overline{w'^2} = \sigma_{\lambda w}^2 \delta + \alpha (1 - \delta) (\sigma_w^2 + (-\overline{w} + w_1)^2) + (1 - \alpha) (1 - \delta) (\sigma_w^2 + (-\overline{w} + w_2)^2)$$

formulas are equivalent to each other. We can do this by using `Eq(...)` from the package `SymPy`. `factor(...)` tries to factor the given variables to make the comparison easier. All of this can be seen in listing 5.8. This code (listing 5.8) just displays `True`, which is exactly

Listing 5.8: Check if the integral and the given formula are the same

```
display(sp.factor(sp.Eq(w_prime_2_bar_int_val, w_prime_2_bar_check()),
    sp.abc.alpha, sp.abc.delta))
```

what we wanted to have.

5.2 Numeric integration

Again, for better readability, we choose to check the formula for $\overline{w'^2 \theta'_l}$ (this is item 6 from section 2.6). As in section 5.2, there needs to be some packages imported. We are importing the same packages as in listing 5.1 together with some more (listing 5.9). Since we are going

Listing 5.9: Import statements

```
from itertools import product
import pandas as pd
import numpy as np
```

to need some more symbols, we also need to define those. We still use the symbols as in

listing 5.2, together with the ones in listing 5.10.

Listing 5.10: Defining symbols

```
sigma_lambda_theta_1 = Symbol('\sigma_{\lambda\theta_1}')
theta_1_1 = Symbol('\theta_{11}')
theta_1_2 = Symbol('\theta_{12}')
theta_1_bar = Symbol('\overline{\theta_1}')
sigma_theta_1_1 = Symbol('\sigma_{\theta_{11}}')
sigma_theta_1_2 = Symbol('\sigma_{\theta_{12}}')
rho_w_theta_1 = Symbol('\rho_{w\theta_1}')
w_prime_3_bar = Symbol('\overline{w^3}')
w_prime_theta_1_prime_bar = Symbol('\overline{w^{\theta_1}}')
w_prime_2_theta_prime_1_bar = Symbol('\overline{w^2\theta_1}')
sigma_tilde_w = Symbol('\Tilde{\sigma}_w')
lambda_w_theta = Symbol('\lambda_{w\theta}')
lambda_w = Symbol('\lambda_w')
```

We start defining the integral by defining the marginals (listing 5.11).

Listing 5.11: Defining the marginals

```
G_1_w_theta = Normal(name='G_1_w_theta', mean=sp.Matrix([w_1, theta_1_1]),
    std=sp.Matrix([[sigma_w ** 2, 0], [0, sigma_theta_1_1 ** 2]]))
G_1_w_theta_density = density(G_1_w_theta)(sp.abc.w, sp.abc.theta)
G_2_w_theta = Normal(name='G_2_w_theta', mean=sp.Matrix([w_2, theta_1_2]),
    std=sp.Matrix([[sigma_w ** 2, 0], [0, sigma_theta_1_2 ** 2]]))
G_2_w_theta_density = density(G_2_w_theta)(sp.abc.w, sp.abc.theta)
G_3_w_theta = Normal(name='G_3_w_theta', mean=sp.Matrix([w_bar, theta_1_bar]),
    std=sp.Matrix([[sigma_lambda_w ** 2,
        rho_w_theta_1 * sigma_lambda_w * sigma_lambda_theta_1,
        [rho_w_theta_1 * sigma_lambda_w * sigma_lambda_theta_1,
        sigma_lambda_theta_1 ** 2]]]))
G_3_w_theta_density = sp.simplify(density(G_3_w_theta)(sp.abc.w, sp.abc.theta))
G_w_theta = (
    (1 - sp.abc.delta) * sp.abc.alpha * density(G_1_w_theta)(sp.abc.w, sp.abc.theta)
    + (1 - sp.abc.delta) * (1 - sp.abc.alpha) * density(G_2_w_theta)(sp.abc.w, sp.abc.theta)
    + sp.abc.delta * G_3_w_theta_density)
```

The integral which needs to be computed is then defined as in listing 5.12.

Here (listing 5.12), the output is omitted for better readability. We do not yet compute the integral, because due to the complexity, unfortunately this is not working with SymPy.

Listing 5.12: Defining and displaying the needed integral

```
w_prime_2_theta_l_prime_bar = sp.Integral((sp.abc.w - w_bar) ** 2 *
    (sp.abc.theta - theta_l_bar) * G_w_theta,
    [sp.abc.w, -oo, oo], [sp.abc.theta, -oo, oo])
display(sp.Eq(w_prime_2_theta_prime_l_bar, w_prime_2_theta_l_prime_bar))
```

Since there is still the equation to check needed, we proceed by defining a function for that in listing 5.13. Looking at figure 5.4, there are some other equations needed like

Listing 5.13: Python function for $\overline{w'^2\theta_l}$

```
def w_prime_2_theta_l_prime_bar_check(sigma_tilde_w = sigma_tilde_w,
    delta = sp.abc.delta, lambda_w_theta = lambda_w_theta, lambda_w = lambda_w,
    w_prime_3_bar = w_prime_3_bar, w_prime_2_bar = w_prime_2_bar,
    w_prime_theta_l_prime_bar = w_prime_theta_l_prime_bar):
    return ((1 / (1 - sigma_tilde_w ** 2)) *
        ((1 - delta * lambda_w_theta) / (1 - delta * lambda_w)) *
        (w_prime_3_bar / w_prime_2_bar) *
        w_prime_theta_l_prime_bar)
display(sp.Eq(w_prime_2_theta_prime_l_bar, w_prime_2_theta_l_prime_bar_check()))
```

Figure 5.4: Output of listing 5.13

$$\overline{w'^2\theta_l} = \frac{\overline{w'\theta_l'} \cdot \overline{w'^3} (-\lambda_{w\theta}\delta + 1)}{\overline{w'^2} \cdot (1 - \tilde{\sigma}_w^2) (-\lambda_w\delta + 1)}$$

equation (4.3.15), equation (4.3.5), equation (4.3.3), equation (4.2.3), and equation (4.1.4). We do not list the functions to those equations here, because they are defined the same way like the other equations are defined as functions.

Instead, since we cannot compute the integral analytically, we can create a **dataframe** using **pandas**.² The columns for this dataframe are going to be all the inputs we have. To get

²Wes McKinney. “Data Structures for Statistical Computing in Python”. In: *Proceedings of the 9th Python in Science Conference*. Ed. by Stéfan van der Walt and Jarrod Millman. 2010, pp. 56–61. DOI: 10.25080/Majora-92bf1922-00a.

all permutations, this code (listing 5.14) is also using `product(..)` from the `itertools` package.

Listing 5.14: Create a dataframe and putting in arbitrary numbers

```
df = pd.DataFrame(product([0, 1],[-2, 2],[-1, 2],[0, 3], [Rational(1, 10)],
    [Rational(3, 10)], [Rational(4, 10)], [Rational(7, 10)], [Rational(6, 10)],
    [Rational(5, 10)], [Rational(1, 10), Rational(5, 10)], [Rational(5, 10)]),
    columns=[w_1, w_2, theta_l_1, theta_l_2, sigma_theta_l_1, sigma_theta_l_2,
    sigma_lambda_theta_l, sigma_w, sigma_lambda_w, sp.abc.alpha, sp.abc.delta,
    rho_w_theta_l])
```

We append another column which is called “checkval” and lists the values for the given equation to check. This code also uses the function defined in listing 5.13, where all other

Listing 5.15: Attaching the “checkval” column to the dataframe

```
df['checkval'] = (df.apply(lambda x: w_prime_2_theta_l_prime_bar_check_val.subs({
    w_1: x[w_1], w_2: x[w_2], theta_l_1: x[theta_l_1], theta_l_2: x[theta_l_2],
    sigma_theta_l_1: x[sigma_theta_l_1], sigma_theta_l_2: x[sigma_theta_l_2],
    sigma_lambda_theta_l: x[sigma_lambda_theta_l], sigma_w: x[sigma_w],
    sigma_lambda_w: x[sigma_lambda_w], sp.abc.alpha: x[sp.abc.alpha],
    sp.abc.delta: x[sp.abc.delta], rho_w_theta_l: x[rho_w_theta_l]}), axis=1))
```

equations are substituted into. The function `df.apply(..)` is used to apply the function given in the parenthesis to all rows of the dataframe by specifying a `lambda x`, where `x` is corresponding to the given dataframe, `df`. Lastly, there is also the `axis=1` parameter, which specifies the direction of applying the function.

Next, we are actually computing $\overline{w'^2\theta_l}$ numerically by using the quadrature method and applying the values of this integrals to a new column in the dataframe (listing 5.16). Here, we are using the integral which has been specified earlier and adding the parameter `method='quad'` to the function `.doit(..)`. After that, `.evalf(..)` just gives the numerical value. We try to prove that the integral value equals the function value, hence we are computing the error between those two columns (listing 5.17) and take the mean (`numpy.mean(..)` from the pack-

Listing 5.16: Attaching the “numint” column to the dataframe

```
df['numint'] = (df.apply(lambda x: Rational(w_prime_2_theta_1_prime_bar.subs({
    w_1: x[w_1], w_2: x[w_2], theta_1_1: x[theta_1_1], theta_1_2: x[theta_1_2],
    sigma_theta_1_1: x[sigma_theta_1_1], sigma_theta_1_2: x[sigma_theta_1_2],
    sigma_lambda_theta_1: x[sigma_lambda_theta_1], sigma_w: x[sigma_w],
    sigma_lambda_w: x[sigma_lambda_w], sp.abc.alpha: x[sp.abc.alpha],
    sp.abc.delta: x[sp.abc.delta], rho_w_theta_1: x[rho_w_theta_1]
})).doit(conds='none', method='quad').evalf()), axis=1))
```

age NumPy³) of these new columns (listing 5.18) to see if the error is actually numerically 0.

Listing 5.17: Attaching the “diffnum” column to the dataframe

```
df['diffnum'] = abs(df['checkval'].astype(float) - df['numint'].astype(float))
```

Listing 5.18: Calculating the mean difference

```
print('The mean error between the rhs and the lhs is:', np.mean(df['diffnum']))
```

Figure 5.5: Output of listing 5.18

The mean error between the rhs and the lhs is: 1.3753423344481015e-124

In figure 5.5, we see that the mean error is basically 0 which we wanted. It should be noted that based on the configuration of each individual computer, the solutions can slightly differ due to floating point arithmetic.

³Charles R. Harris et al. “Array programming with NumPy”. In: *Nature* 585.7825 (Sept. 2020), pp. 357–362. DOI: 10.1038/s41586-020-2649-2. URL: <https://doi.org/10.1038/s41586-020-2649-2>.

6 Asymptotics

Once we defined all functions, we see that we want certain behaviors for certain values as well as there is a need to restrict some parameter values.

We start with the “obvious” restrictions for the pdf parameters. The mixture fractions α and δ are meant to be $\alpha \in [0, 1]$ and $\delta \in [0, 1)$. But since the code tries to simplify a lot of things, the binormal representation also does not revert back to a single normal distribution. Therefore we have $\alpha \in (0, 1)$ due to the code. The restriction to δ makes sense in a way that we do not really want just the third normal to predict the whole shape. Also, most of the formulas, e.g. equation (4.3.6) have a $1 - \delta$ in the denominator.

In section 2.3, we saw, how the transformation between the sum of two normal distributions and the sum of three normal distributions are working. From those transformations, we see that we want That is, for instance, that the variance of w over the whole pdf has to be strictly

$$\begin{aligned} & 0 < \delta\lambda_w < 1, \quad 0 < \delta\lambda_\theta < 1, \quad 0 < \delta\lambda_r < 1, \\ \iff & 0 < \delta\frac{\sigma_{w3}^2}{w'^2} < 1, \quad 0 < \delta\frac{\sigma_{\theta l3}^2}{\theta_l'^2} < 1, \quad 0 < \delta\frac{\sigma_{r_t3}^2}{r_t'^2} < 1, \\ \iff & 0 < \delta\sigma_{w3}^2 < \overline{w'^2}, \quad 0 < \delta\sigma_{\theta l3}^2 < \overline{\theta_l'^2}, \quad 0 < \delta\sigma_{r_t3}^2 < \overline{r_t'^2}. \end{aligned}$$

greater than δ times the squared standard deviation in w of the third normal distribution.

For realizability, it turns out that we want $-1 < \hat{c}_{w\theta_l}, \hat{c}_{wr_t}, \hat{c}_{r_t\theta_l} < 1$. This is for instance:

$$c_{wr_t}^2 < (1 - \tilde{\sigma}_w^2) \left(\frac{(1 - \delta\lambda_w)(1 - \delta\lambda_r)}{(1 - \delta\lambda_{wr})^2} \right) \quad (6.0.1)$$

So it might be safer to set

$$\lambda_w, \lambda_r < \lambda_{wr} \iff \frac{\sigma_{w3}^2}{w'^2}, \frac{\sigma_{r_t3}^2}{r_t'^2} < \frac{\rho_{wr_t}\sigma_{w3}\sigma_{r_t3}}{w'r_t'} \quad (6.0.2)$$

so that the rhs is greater and the bound is less restrictive. If we assume $\lambda_\theta = \lambda_r$ and $\lambda_{w\theta} = \lambda_{wr}$, then we have 5 new pdf parameters: $\delta, \lambda_w, \lambda_\theta, \lambda_{w\theta}, \lambda_{\theta r}$. If we want the pdf to revert to a single normal distribution in the limit of zero skewness, then we need

$$\delta, \lambda_w, \lambda_r, \lambda_\theta, \lambda_{wr}, \lambda_{w\theta}, \lambda_{\theta r} \rightarrow 1, \quad (6.0.3)$$

as $Sk_w \rightarrow 0$. That being said, we can have a look at equation (4.3.6) and take the limit.

$$\begin{aligned} & \lim_{\delta \rightarrow 1, \lambda_w \rightarrow 1} \frac{1}{(1 - \tilde{\sigma}_w^2)^{3/2}} \frac{\overline{w'^3}}{\left(\overline{w'^2}\right)^{3/2}} \frac{1}{\left(\frac{1-\delta\lambda_w}{1-\delta}\right)^{3/2}} \frac{1}{1-\delta} \\ &= \lim_{\delta \rightarrow 1, \lambda_w \rightarrow 1} \frac{1}{(1 - \tilde{\sigma}_w^2)^3} \frac{\overline{w'^3}^2}{\left(\overline{w'^2}\right)^3} \frac{1}{\left(\frac{1-\delta\lambda_w}{1-\delta}\right)^3} \frac{1}{(1-\delta)^2} \\ &= \lim_{\delta \rightarrow 1, \lambda_w \rightarrow 1} \frac{1}{(1 - \tilde{\sigma}_w^2)^3} \frac{\overline{w'^3}^2}{\left(\overline{w'^2}\right)^3} \frac{(1-\delta)^3}{(1-\delta\lambda_w)^3} \frac{1}{(1-\delta)^2} \\ &= \lim_{\delta \rightarrow 1, \lambda_w \rightarrow 1} \frac{1}{(1 - \tilde{\sigma}_w^2)^3} \frac{\overline{w'^3}^2}{\left(\overline{w'^2}\right)^3} \underbrace{\frac{1-\delta}{(1-\delta\lambda_w)^3}}_{\rightarrow 0} \end{aligned} \quad (6.0.4)$$

Hence, if $\delta \rightarrow 1$ we get $Sk_w \rightarrow 0$. Of course, one needs to pay attention when computing those equations in the code, since we could run into a division by zero error, depending on which values are computed first.

7 Summary

In this thesis we saw how adding a third normal distribution to the CLUBB model would improve the representation of certain shapes. While it is not entirely clear at first why this is an actual improvement, there are some plots describing certain behaviors in section 2.1. Once we saw the graphical advantages,

8 Outlook

We have seen that adding a third normal distribution right in the middle of the two defined (simplified) normal distributions does not change the “closed” formulas too much, neither it makes it too complicated. So for CLUBB there are new parameters, e.g. δ or λ_w , which can be chosen to “tweak” the representation of the underlying pdf for the prognosed moments. Ultimately one would like to fit this resulting pdf to real data, to get better relationships, as well as thresholds for some variables. To do this there are some approaches which unfortunately have not been discussed. A following thesis could e.g. incorporate some machine learning approach to learn optimal values for certain parameters.

The methodology established in chapter 5 extends far beyond the immediate application within the CLUBB model. This chapter serves as a blueprint for a generalizable approach to verifying and analyzing integral expressions. Its core strength lies in the utilization of SymPy, a powerful and well-supported cas. SymPy’s community-driven nature provides continuous development and a vast library of mathematical capabilities. By leveraging this versatile tool, we can tackle a wide range of integral expressions, both analytically and numerically. This approach offers several advantages:

- *Symbolic Verification:* SymPy allows us to perform symbolic manipulations, enabling the derivation of exact solutions for integrals whenever possible.
- *Numerical Approximation:* For integrals that are analytically intractable, SymPy seamlessly integrates with numerical computing libraries. This allows us to efficiently approximate the integral’s value for specific parameter choices. This combined approach ensures we can handle a broader range of integral expressions.

- *Generalizability and Reusability:* The framework outlined in chapter 5 is not specific to the context of CLUBB. By focusing on the core functionalities of SymPy, this approach can be adapted to various scientific disciplines.

Overall, the methods we developed in this thesis using SymPy are not just useful for the CLUBB model. These methods can be applied to many other scientific problems because they can both solve integrals exactly (symbolically) and get close answers (numerically) for a wide range of equations. SymPy, being a powerful and widely-used tool, makes this possible.

Bibliography

- Harris, Charles R. et al. “Array programming with NumPy”. In: *Nature* 585.7825 (Sept. 2020), pp. 357–362. DOI: 10.1038/s41586-020-2649-2. URL: <https://doi.org/10.1038/s41586-020-2649-2>.
- Izenman, Alan Julian. *Modern multivariate statistical techniques: regression, classification, and manifold learning*. Springer texts in statistics. OCLC: ocn225427579. New York ; London: Springer, 2008. ISBN: 9780387781884.
- Larson, Vincent E and Jean-Christophe Golaz. “Using probability density functions to derive consistent closure relationships among higher-order moments”. In: *Monthly Weather Review* 133.4 (2005), pp. 1023–1042.
- Larson, Vincent E. *CLUBB-SILHS: A parameterization of subgrid variability in the atmosphere*. 2022. arXiv: 1711.03675 [physics.ao-ph].
- McKinney, Wes. “Data Structures for Statistical Computing in Python”. In: *Proceedings of the 9th Python in Science Conference*. Ed. by Stéfan van der Walt and Jarrod Millman. 2010, pp. 56–61. DOI: 10.25080/Majora-92bf1922-00a.
- Meurer, Aaron et al. “SymPy: symbolic computing in Python”. In: *PeerJ Computer Science* 3 (Jan. 2017), e103. ISSN: 2376-5992. DOI: 10.7717/peerj-cs.103. URL: <https://doi.org/10.7717/peerj-cs.103>.

A Code

A.1 User Guide

A.1.1 Accessing the code

The code used for checking functions mentioned in the thesis is attached and can be accessed alongside this document.

A.1.2 Key files and their purposes

- `checked_functions.py`: This file houses the definitions of all functions used for checking integrals.
- `symbols.py`: This file contains definitions of all symbols employed for computations with SymPy, a powerful library for symbolic mathematics.

A.1.3 Working with functions

To obtain a function with symbols already incorporated, call it with an empty list of arguments (e.g., `function_name()`).

A.1.4 Displaying equations effectively

1. Import the `display` function from `IPython.display`.
2. Use `display(..)` to present statements or even equations visually.

A.1.5 Handling integrals

1. *Displaying an integral*: Use `sympy.Integral(..)` and put it into `display(..)` to visualize the integral.

2. *Computing an integral symbolically:* Apply `.doit(..)` to the integral object for symbolic computation.
3. *Approximating an integral numerically:* Add the option **"method=quad"** within the `.doit(..)` call to calculate the integral using the quadrature approximation method.

A.2 symbols.py

```

1  from sympy import Symbol, symbols
2
3  # sigma
4  sigma_w = Symbol('\sigma_w')
5
6  sigma_r_t_i, sigma_r_t_1, sigma_r_t_2 = (
7      symbols('\sigma_{r_{ti}} \sigma_{r_{t1}} \sigma_{r_{t2}}'))
8
9  sigma_theta_l_i, sigma_theta_l_1, sigma_theta_l_2 = (
10     symbols('\sigma_{\theta_{li}} \sigma_{\theta_{l1}} \sigma_{\theta_{l2}}'))
11
12  sigma_tilde_r_t_i, sigma_tilde_r_t_1, sigma_tilde_r_t_2 = (
13     symbols('\tilde{\sigma}_{r_{ti}} \tilde{\sigma}_{r_{t1}} \tilde{\sigma}_{r_{t2}}'))
14  )
15
16  sigma_tilde_theta_l_i, sigma_tilde_theta_l_1, sigma_tilde_theta_l_2 = (
17     symbols(
18         '\tilde{\sigma}_{\theta_{li}} \tilde{\sigma}_{\theta_{l1}}
19         \rightarrow \tilde{\sigma}_{\theta_{l2}}')
20  )
21
22  sigma_tilde_lambda_w = Symbol('\tilde{\sigma}_{\lambda_w}')
23  sigma_tilde_w = Symbol('\tilde{\sigma}_w')
24  sigma_lambda_w = Symbol('\sigma_{\lambda_w}')
25  sigma_lambda_theta_l = Symbol('\sigma_{\lambda\theta_l}')
26  sigma_lambda_r_t = Symbol('\sigma_{\{\lambda\}r_t}')
27
28  # w
29  w_bar = Symbol('\overline{w}')
30  w_prime = Symbol('w\prime')
31  w_i, w_1, w_2 = symbols('w_i w_1 w_2')
32
33  w_hat_i, w_hat_1, w_hat_2, w_hat_3, w_hat_prime = (
34     symbols('\hat{w}_i \hat{w}_1 \hat{w}_2 \hat{w}_3 \hat{w}\prime')
35  )
36
37  w_prime_2_bar, w_prime_3_bar, w_prime_4_bar = symbols(
38     '\overline{w\prime^2} \overline{w\prime^3} \overline{w\prime^4}')
39  )
40  w_prime_r_t_prime_bar = Symbol('\overline{w\prime_r\prime_t}')

```

```

41 w_prime_2_theta_prime_1_bar = Symbol('\overline{w}^{\prime 2} \theta \prime_1')
42 w_prime_theta_prime_1_2_bar = Symbol('\overline{w} \theta \prime_1^2')
43 w_prime_theta_1_prime_bar = Symbol('\overline{w} \theta \prime_1')
44
45 w_prime_r_t_prime_theta_1_prime_bar = Symbol('\overline{w} \{r_t\} \{ \theta \prime_1 \}')
46
47 # r
48 r_t = Symbol('r_t')
49
50 r_t_bar, r_t_prime_2_bar = symbols(
51     '\overline{r_t} \overline{r_t}^{\prime 2}')
52 )
53
54 r_t_i, r_t_1, r_t_2 = symbols('r_{ti} r_{t1} r_{t2}')
55
56 r_t_tilde_prime, r_t_i_tilde, r_t_1_tilde, r_t_2_tilde = symbols(
57     '\tilde{r_t} \tilde{r_{ti}} \tilde{r_{t1}} \tilde{r_{t2}}')
58 )
59
60 # theta
61 theta_1 = Symbol('\theta_1')
62
63 theta_1_bar, theta_1_prime_2_bar, theta_1_prime_3_bar = symbols(
64     '\overline{\theta_1} \overline{\theta_1}^{\prime 2} \overline{\theta_1}^{\prime 3}')
65 )
66
67 theta_1_prime_r_t_prime_bar = Symbol('\overline{\theta_1} r_t \prime')
68
69 theta_1_i, theta_1_1, theta_1_2 = symbols(
70     '\theta_{1i} \theta_{11} \theta_{12}')
71 )
72
73 theta_tilde_prime_1 = Symbol('\tilde{\theta}_1')
74
75 theta_tilde_1_i, theta_tilde_1_1, theta_tilde_1_2 = symbols(
76     '\tilde{\theta}_{1i} \tilde{\theta}_{11} \tilde{\theta}_{12}')
77 )
78
79 # lambda
80 lambda_theta = Symbol('\lambda_{\theta}')
81 lambda_theta_r = Symbol('\lambda_{\theta_r}')
82 lambda_r = Symbol('\lambda_r')
83 lambda_w = Symbol('\lambda_w')
84 lambda_w_theta = Symbol('\lambda_{w \theta}')
85 lambda_w_r = Symbol('\lambda_{wr}')
86
87 r_r_t_theta_1 = Symbol('r_{r_t \theta_1}')
88 triple_gaussian = Symbol('P_{tmg}(\hat{w}, \tilde{\theta}_1, \tilde{r}_{\prime t})')
89 sk_hat_w = Symbol('\widehat{Sk}_w')
90 sk_theta_1_hat = Symbol('\widehat{Sk}_{\theta_1}')
91 c_w_theta_1_hat = Symbol('\hat{c}_{w \theta_1}')
92 G_3 = Symbol('G_3')
93 G_3_hat = Symbol('\hat{G}_3')
94

```

```

95 G_w, G_1_w, G_2_w, G_3_w = symbols(
96     'G_{w}(w) G_{1w}(w) G_{2w}(w) G_{3w}(w)'
97 )
98
99 G_w_1_2 = Symbol('G_{w_{12}}')
100
101 G_w_theta = Symbol('G_{w\\theta}(w,\\theta)')
102 G_1_w_theta = Symbol('G_{1w\\theta}(w,\\theta)')
103 G_2_w_theta = Symbol('G_{2w\\theta}(w,\\theta)')
104 G_3_w_theta = Symbol('G_{3w\\theta}(w,\\theta)')
105
106 G_theta_r = Symbol('G_{\\theta r}(\\theta, r)')
107 G_1_theta_r = Symbol('G_{1\\theta r}(\\theta, r)')
108 G_2_theta_r = Symbol('G_{2\\theta r}(\\theta, r)')
109 G_3_theta_r = Symbol('G_{3\\theta r}(\\theta, r)')
110
111 G_w_theta_l_r_t = Symbol('G_{w\\theta_l}{r_t}(w,\\theta_l,{r_t})')
112 G_1_w_theta_l_r_t = Symbol('G_{1w\\theta_l}{r_t}(w,\\theta_l,{r_t})')
113 G_2_w_theta_l_r_t = Symbol('G_{2w\\theta_l}{r_t}(w,\\theta_l,{r_t})')
114 G_3_w_theta_l_r_t = Symbol('G_{3w\\theta_l}{r_t}(w,\\theta_l,{r_t})')
115
116 G_theta, G_1_theta, G_2_theta, G_3_theta = symbols(
117     'G_{\\theta}(\\theta) G_{1\\theta}(\\theta) G_{2\\theta}(\\theta)
118     ↦ G_{3\\theta}(\\theta)'
119 )
120
121 rho_w_theta_l = Symbol('\\rho_{w\\theta_l}')
122 rho_w_r_t = Symbol('\\rho_{wr_t}')
123 rho_theta_l_r_t = Symbol('\\rho_{\\theta_lr_t}')

```

Listing A.1: symbols.py

A.3 checked_functions.py

```

1  import sympy as sp
2  from sympy import Rational
3  from sympy.stats import Normal, density
4
5  import symbols as sym
6
7
8  # w equations
9  # -----
10
11 def w_bar(alpha=sp.abc.alpha, delta=sp.abc.delta, w_1=sym.w_1, w_2=sym.w_2):
12     return ((1 - delta) * alpha * w_1
13             + (1 - delta) * (1 - alpha) * w_2
14             + delta * (alpha * w_1 + (1 - alpha) * w_2))
15

```



```

16
17 def w_prime_2_bar(delta=sp.abc.delta, alpha=sp.abc.alpha, w_1=sym.w_1, w_2=sym.w_2,
18                 w_bar=sym.w_bar, sigma_w=sym.sigma_w,
19                 ↪ sigma_lambda_w=sym.sigma_lambda_w):
19     return (((1 - delta) * alpha * ((w_1 - w_bar) ** 2 + sigma_w ** 2)) +
20            ((1 - delta) * (1 - alpha) * ((w_2 - w_bar) ** 2 + sigma_w ** 2)) +
21            (delta * sigma_lambda_w ** 2))
22
23
24 def w_prime_3_bar(delta=sp.abc.delta, alpha=sp.abc.alpha, w_1=sym.w_1, w_2=sym.w_2,
25                 w_bar=sym.w_bar, sigma_w=sym.sigma_w):
26     return (((1 - delta) * alpha * ((w_1 - w_bar) ** 3 +
27                                     3 * sigma_w ** 2 * (w_1 - w_bar))) +
28            ((1 - delta) * (1 - alpha) * ((w_2 - w_bar) ** 3 +
29                                     3 * sigma_w ** 2 * (w_2 - w_bar))))
30
31
32 def w_prime_4_bar(w_prime_2_bar=sym.w_prime_2_bar,
33                 w_prime_3_bar=sym.w_prime_3_bar,
34                 delta=sp.abc.delta,
35                 sigma_tilde_w=sym.sigma_tilde_w,
36                 sigma_lambda_w=sym.sigma_lambda_w):
37     return (w_prime_2_bar ** 2 *
38            ((1 - delta * (sigma_lambda_w ** 2 / w_prime_2_bar)) ** 2 / (1 - delta)) *
39            (3 * sigma_tilde_w ** 4 +
40             6 * (1 - sigma_tilde_w ** 2) *
41             sigma_tilde_w ** 2 +
42             (1 - sigma_tilde_w ** 2) ** 2) +
43            ((1 / (1 - sigma_tilde_w ** 2)) *
44             (1 / (1 - delta * (sigma_lambda_w ** 2 / w_prime_2_bar)))) *
45            (w_prime_3_bar ** 2 / w_prime_2_bar)) +
46            (delta * 3 * sigma_lambda_w ** 4))
47
48
49 # -----
50
51 # theta_l equations
52 # -----
53
54 def theta_l_bar(alpha=sp.abc.alpha, delta=sp.abc.delta,
55                theta_l_1=sym.theta_l_1, theta_l_2=sym.theta_l_2):
56     return ((1 - delta) * alpha * theta_l_1
57            + (1 - delta) * (1 - alpha) * theta_l_2
58            + delta * (alpha * theta_l_1 + (1 - alpha) * theta_l_2))
59
60
61 def theta_l_prime_2_bar(delta=sp.abc.delta,
62                        alpha=sp.abc.alpha,
63                        theta_l_1=sym.theta_l_1,
64                        theta_l_2=sym.theta_l_2,
65                        theta_l_bar=sym.theta_l_bar,
66                        sigma_theta_l_1=sym.sigma_theta_l_1,
67                        sigma_theta_l_2=sym.sigma_theta_l_2,
68                        sigma_lambda_theta_l=sym.sigma_lambda_theta_l):

```

```

69     return (((1 - delta) * alpha * ((theta_l_1 - theta_l_bar) ** 2 + sigma_theta_l_1 **
70         ↪ 2)) +
71         ((1 - delta) * (1 - alpha) *
72         ((theta_l_2 - theta_l_bar) ** 2 + sigma_theta_l_2 ** 2)) +
73         (delta * sigma_lambda_theta_l ** 2))
74
75 def theta_l_prime_3_bar(delta=sp.abc.delta,
76     alpha=sp.abc.alpha,
77     theta_l_1=sym.theta_l_1,
78     theta_l_2=sym.theta_l_2,
79     theta_l_bar=sym.theta_l_bar,
80     sigma_theta_l_1=sym.sigma_theta_l_1,
81     sigma_theta_l_2=sym.sigma_theta_l_2):
82     return (((1 - delta) * alpha *
83         ((theta_l_1 - theta_l_bar) ** 3 +
84         3 * sigma_theta_l_1 ** 2 * (theta_l_1 - theta_l_bar))) +
85         ((1 - delta) * (1 - alpha) *
86         ((theta_l_2 - theta_l_bar) ** 3 +
87         3 * sigma_theta_l_2 ** 2 * (theta_l_2 - theta_l_bar))))
88
89
90 # -----
91
92 # r_t equations
93 # -----
94
95 def r_t_bar(alpha=sp.abc.alpha, delta=sp.abc.delta, r_t_1=sym.r_t_1, r_t_2=sym.r_t_2):
96     return ((1 - delta) * alpha * r_t_1
97         + (1 - delta) * (1 - alpha) * r_t_2
98         + delta * (alpha * r_t_1 + (1 - alpha) * r_t_2))
99
100
101 def r_t_prime_2_bar(delta=sp.abc.delta,
102     alpha=sp.abc.alpha,
103     r_t_1=sym.r_t_1,
104     r_t_2=sym.r_t_2,
105     r_t_bar=sym.r_t_bar,
106     sigma_r_t_1=sym.sigma_r_t_1,
107     sigma_r_t_2=sym.sigma_r_t_2,
108     sigma_lambda_r_t=sym.sigma_lambda_r_t):
109     return (((1 - delta) * alpha * ((r_t_1 - r_t_bar) ** 2 + sigma_r_t_1 ** 2)) +
110         ((1 - delta) * (1 - alpha) * ((r_t_2 - r_t_bar) ** 2 + sigma_r_t_2 ** 2)) +
111         (delta * sigma_lambda_r_t ** 2))
112
113
114 # -----
115
116 # Mixed equations
117 # -----
118
119 def w_prime_theta_l_prime_bar(delta=sp.abc.delta,
120     alpha=sp.abc.alpha,
121     w_1=sym.w_1,

```

```

122         w_2=sym.w_2,
123         w_bar=sym.w_bar,
124         theta_l_1=sym.theta_l_1,
125         theta_l_2=sym.theta_l_2,
126         theta_l_bar=sym.theta_l_bar,
127         cov_lambda_w_theta=sym.rho_w_theta_l * sym.sigma_lambda_w
            ↪ *
            sym.sigma_lambda_theta_l):
128
129     return (((1 - delta) * alpha * ((w_1 - w_bar) * (theta_l_1 - theta_l_bar))) +
130            ((1 - delta) * (1 - alpha) * ((w_2 - w_bar) * (theta_l_2 - theta_l_bar)))
131            + delta * cov_lambda_w_theta)
132
133
134 def w_prime_r_t_prime_bar(delta=sp.abc.delta,
135                            alpha=sp.abc.alpha,
136                            w_1=sym.w_1,
137                            w_2=sym.w_2,
138                            w_bar=sym.w_bar,
139                            r_t_1=sym.r_t_1,
140                            r_t_2=sym.r_t_2,
141                            r_t_bar=sym.r_t_bar,
142                            cov_lambda_w_r=sym.rho_w_r_t * sym.sigma_lambda_w *
143                            sym.sigma_lambda_r_t):
144     return (((1 - delta) * alpha * ((w_1 - w_bar) * (r_t_1 - r_t_bar))) +
145            ((1 - delta) * (1 - alpha) * ((w_2 - w_bar) * (r_t_2 - r_t_bar)))
146            + delta * cov_lambda_w_r)
147
148
149 def w_prime_2_theta_l_prime_bar(sigma_tilde_w=sym.sigma_tilde_w,
150                                 delta=sp.abc.delta,
151                                 lambda_w_theta=sym.lambda_w_theta,
152                                 lambda_w=sym.lambda_w,
153                                 w_prime_3_bar=sym.w_prime_3_bar,
154                                 w_prime_2_bar=sym.w_prime_2_bar,
155                                 w_prime_theta_l_prime=sym.w_prime_theta_l_prime_bar):
156     return ((1 / (1 - sigma_tilde_w ** 2)) *
157            ((1 - delta * lambda_w_theta) / (1 - delta * lambda_w)) *
158            (w_prime_3_bar / w_prime_2_bar) *
159            w_prime_theta_l_prime)
160
161
162 def w_prime_theta_l_prime_2_bar(delta=sp.abc.delta,
163                                 lambda_w_theta=sym.lambda_w_theta,
164                                 lambda_w=sym.lambda_w,
165                                 sigma_tilde_w=sym.sigma_tilde_w,
166                                 w_prime_3_bar=sym.w_prime_3_bar,
167                                 w_prime_2_bar=sym.w_prime_2_bar,
168                                 w_prime_theta_l_prime_bar=sym.w_prime_theta_l_prime_bar,
169                                 theta_l_prime_3_bar=sym.theta_l_prime_3_bar):
170     return (Rational(2, 3) *
171            ((1 - delta * lambda_w_theta) ** 2 / (1 - delta * lambda_w) ** 2) *
172            (1 / (1 - sigma_tilde_w ** 2) ** 2) *
173            (w_prime_3_bar / w_prime_2_bar ** 2) *
174            w_prime_theta_l_prime_bar ** 2 +

```

```

175         Rational(1, 3) *
176         ((1 - delta * lambda_w) / (1 - delta * lambda_w_theta)) *
177         (1 - sigma_tilde_w ** 2) *
178         ((w_prime_2_bar * theta_l_prime_3_bar) / w_prime_theta_l_prime_bar))
179
180
181 def w_prime_theta_l_prime_2_bar_beta(
182     sigma_tilde_w=sym.sigma_tilde_w,
183     delta=sp.abc.delta,
184     lambda_theta=sym.lambda_theta,
185     lambda_w=sym.lambda_w,
186     w_prime_3_bar=sym.w_prime_3_bar,
187     w_prime_2_bar=sym.w_prime_2_bar,
188     beta=sp.abc.beta,
189     theta_l_prime_2_bar=sym.theta_l_prime_2_bar,
190     lambda_w_theta=sym.lambda_w_theta,
191     w_prime_theta_l_prime_bar=sym.w_prime_theta_l_prime_bar):
192     from sympy import Rational
193     return (
194         (1 / (1 - sigma_tilde_w ** 2)) *
195         ((1 - delta * lambda_theta) / (1 - delta * lambda_w)) *
196         (w_prime_3_bar / w_prime_2_bar) *
197         (
198             Rational(1, 3) * beta * theta_l_prime_2_bar +
199             (
200                 ((1 - Rational(1, 3) * beta) / (1 - sigma_tilde_w ** 2)) *
201                 ((1 - delta * lambda_w_theta) ** 2 /
202                  ((1 - delta * lambda_w) * (1 - delta * lambda_theta))) *
203                 (w_prime_theta_l_prime_bar ** 2 / w_prime_2_bar)
204             )
205         )
206     )
207
208
209 def w_prime_r_t_prime_theta_l_prime_bar(delta=sp.abc.delta,
210     alpha=sp.abc.alpha,
211     w_1=sym.w_1,
212     w_2=sym.w_2,
213     w_bar=sym.w_bar,
214     r_t_1=sym.r_t_1,
215     r_t_2=sym.r_t_2,
216     r_t_bar=sym.r_t_bar,
217     theta_l_1=sym.theta_l_1,
218     theta_l_2=sym.theta_l_2,
219     theta_l_bar=sym.theta_l_bar,
220     r_r_t_theta_l=sym.r_r_t_theta_l,
221     sigma_r_t_1=sym.sigma_r_t_1,
222     sigma_r_t_2=sym.sigma_r_t_2,
223     sigma_theta_l_1=sym.sigma_theta_l_1,
224     sigma_theta_l_2=sym.sigma_theta_l_2):
225     return (((1 - delta) * alpha * (w_1 - w_bar) *
226         (
227             (r_t_1 - r_t_bar) * (theta_l_1 - theta_l_bar) +
228             r_r_t_theta_l * sigma_r_t_1 * sigma_theta_l_1

```

```

229         )) +
230         ((1 - delta) * (1 - alpha) * (w_2 - w_bar) *
231         (
232             (r_t_2 - r_t_bar) * (theta_l_2 - theta_l_bar) +
233             r_r_t_theta_l * sigma_r_t_2 * sigma_theta_l_2
234         )))
235
236
237 def w_prime_r_t_prime_theta_l_prime_bar_beta(
238     beta=sp.abc.beta,
239     sigma_tilde_w=sym.sigma_tilde_w,
240     delta=sp.abc.delta,
241     lambda_theta_r=sym.lambda_theta_r,
242     lambda_w=sym.lambda_w,
243     theta_l_prime_r_t_prime_bar=sym.theta_l_prime_r_t_prime_bar,
244     w_prime_3_bar=sym.w_prime_3_bar,
245     w_prime_2_bar=sym.w_prime_2_bar,
246     lambda_w_r=sym.lambda_w_r,
247     lambda_w_theta=sym.lambda_w_theta,
248     w_prime_r_t_prime_bar=sym.w_prime_r_t_prime_bar,
249     w_prime_theta_l_prime_bar=sym.w_prime_theta_l_prime_bar):
250     from sympy import Rational
251     return (
252         (
253             ((Rational(1, 3) * beta) / (1 - sigma_tilde_w ** 2)) *
254             ((1 - delta * lambda_theta_r) / (1 - delta * lambda_w)) *
255             theta_l_prime_r_t_prime_bar *
256             (w_prime_3_bar / w_prime_2_bar)
257         ) +
258         (
259             ((1 - Rational(1, 3) * beta) / (1 - sigma_tilde_w ** 2) ** 2) *
260             (((1 - delta * lambda_w_r) * (1 - delta * lambda_w_theta)) /
261              ((1 - delta * lambda_w) ** 2)) *
262             w_prime_r_t_prime_bar *
263             w_prime_theta_l_prime_bar *
264             (w_prime_3_bar / w_prime_2_bar ** 2)
265         )
266     )
267
268
269 def r_t_prime_theta_l_prime_bar(alpha=sp.abc.alpha, delta=sp.abc.delta,
270     r_t_1=sym.r_t_1, r_t_2=sym.r_t_2,
271     ↪ r_t_prime_bar=sym.r_t_bar,
272     theta_l_1=sym.theta_l_1, theta_l_2=sym.theta_l_2,
273     theta_l_bar=sym.theta_l_bar,
274     r_r_t_theta_l=sym.r_r_t_theta_l,
275     sigma_r_t_1=sym.sigma_r_t_1,
276     ↪ sigma_r_t_2=sym.sigma_r_t_2,
277     sigma_theta_l_1=sym.sigma_theta_l_1,
278     sigma_theta_l_2=sym.sigma_theta_l_2,
279     cov_lambda_r_theta=sym.rho_theta_l_r_t *
280     sym.sigma_lambda_theta_l *
281     sym.sigma_lambda_r_t):
282     return ((1 - delta) * alpha * (

```

```

281         (r_t_1 - r_t_prime_bar) * (theta_l_1 - theta_l_bar) +
282         r_r_t_theta_l * sigma_r_t_1 * sigma_theta_l_1) +
283         ((1 - delta) * (1 - alpha) * (
284             (r_t_2 - r_t_prime_bar) * (theta_l_2 - theta_l_bar) +
285             r_r_t_theta_l * sigma_r_t_2 * sigma_theta_l_2)) +
286         delta * cov_lambda_r_theta)
287
288
289 # ---
290
291 # Distributions
292 # ---
293
294 G_1_theta_l = Normal(name='G_1_theta_l', mean=sym.theta_l_1, std=sym.sigma_theta_l_1)
295 G_1_theta_l_density = density(G_1_theta_l)(sym.theta_l)
296
297 G_2_theta_l = Normal(name='G_2_theta_l', mean=sym.theta_l_2, std=sym.sigma_theta_l_2)
298 G_2_theta_l_density = density(G_2_theta_l)(sym.theta_l)
299
300 G_3_theta_l = Normal(name='G_3_theta_l', mean=sym.theta_l_bar,
301     ↪ std=sym.sigma_lambda_theta_l)
302 G_3_theta_l_density = density(G_3_theta_l)(sym.theta_l)
303
304 G_theta = ((1 - sp.abc.delta) * sp.abc.alpha * G_1_theta_l_density +
305             (1 - sp.abc.delta) * (1 - sp.abc.alpha) * G_2_theta_l_density +
306             sp.abc.delta * G_3_theta_l_density)
307
308 # ---
309
310 G_1_w = Normal(name='G_1_w', mean=sym.w_1, std=sym.sigma_w)
311 G_1_w_density = density(G_1_w)(sp.abc.w)
312
313 G_2_w = Normal(name='G_2_w', mean=sym.w_2, std=sym.sigma_w)
314 G_2_w_density = density(G_2_w)(sp.abc.w)
315
316 G_3_w = Normal(name='G_3_w', mean=sym.w_bar, std=sym.sigma_lambda_w)
317 G_3_w_density = density(G_3_w)(sp.abc.w)
318
319 G_w = ((1 - sp.abc.delta) * sp.abc.alpha * G_1_w_density +
320         (1 - sp.abc.delta) * (1 - sp.abc.alpha) * G_2_w_density +
321         sp.abc.delta * G_3_w_density)
322
323 # ---
324
325 G_1_w_theta = Normal(name='G_1_w_theta', mean=sp.Matrix([sym.w_1, sym.theta_l_1]),
326     ↪ std=sp.Matrix([[sym.sigma_w ** 2, 0], [0, sym.sigma_theta_l_1 **
327     ↪ 2]]))
328 G_1_w_theta_density = density(G_1_w_theta)(sp.abc.w, sp.abc.theta)
329
330 G_2_w_theta = Normal(name='G_2_w_theta', mean=sp.Matrix([sym.w_2, sym.theta_l_2]),
331     ↪ std=sp.Matrix([[sym.sigma_w ** 2, 0], [0, sym.sigma_theta_l_2 **
332     ↪ 2]]))
333 G_2_w_theta_density = density(G_2_w_theta)(sp.abc.w, sp.abc.theta)

```

```

332 G_3_w_theta = Normal(name='G_3_w_theta', mean=sp.Matrix([sym.w_bar, sym.theta_l_bar]),
333                      std=sp.Matrix([
334                          [sym.sigma_lambda_w ** 2,
335                           sym.rho_w_theta_l * sym.sigma_lambda_w *
336                             ↪ sym.sigma_lambda_theta_l],
337                          [sym.rho_w_theta_l * sym.sigma_lambda_w *
338                             ↪ sym.sigma_lambda_theta_l,
339                           sym.sigma_lambda_theta_l ** 2]
340                      ]))
341 G_3_w_theta_density = sp.simplify(density(G_3_w_theta)(sp.abc.w, sp.abc.theta))
342
343 G_w_theta = ((1 - sp.abc.delta) * sp.abc.alpha * G_1_w_theta_density +
344              (1 - sp.abc.delta) * (1 - sp.abc.alpha) * G_2_w_theta_density +
345              sp.abc.delta * G_3_w_theta_density)
346
347 # -----
348 mu_1_theta_l_r_t = sp.Matrix([sym.theta_l_1, sym.r_t_1])
349 Sigma_1_theta_l_r_t = sp.Matrix([[sym.sigma_theta_l_1 ** 2,
350                                   sym.r_r_t_theta_l * sym.sigma_theta_l_1 *
351                                     ↪ sym.sigma_r_t_1],
352                                   [sym.r_r_t_theta_l * sym.sigma_theta_l_1 *
353                                     ↪ sym.sigma_r_t_1,
354                                   sym.sigma_r_t_1 ** 2]])
355
356 G_1_theta_l_r_t = Normal(name='G_1_theta_l_r_t',
357                           mean=mu_1_theta_l_r_t,
358                           std=Sigma_1_theta_l_r_t)
359
360 G_1_theta_l_r_t_density = density(G_1_theta_l_r_t)(sym.theta_l, sym.r_t)
361
362 mu_2_theta_l_r_t = sp.Matrix([sym.theta_l_2, sym.r_t_2])
363 Sigma_2_theta_l_r_t = sp.Matrix([
364     [sym.sigma_theta_l_2 ** 2,
365      sym.r_r_t_theta_l * sym.sigma_theta_l_2 * sym.sigma_r_t_2],
366     [sym.r_r_t_theta_l * sym.sigma_theta_l_2 * sym.sigma_r_t_2,
367      sym.sigma_r_t_2 ** 2]])
368
369 G_2_theta_l_r_t = Normal(name='G_2_theta_l_r_t',
370                           mean=mu_2_theta_l_r_t,
371                           std=Sigma_2_theta_l_r_t)
372
373 G_2_theta_l_r_t_density = density(G_2_theta_l_r_t)(sym.theta_l, sym.r_t)
374
375 mu_3_theta_l_r_t = sp.Matrix([sym.theta_l_bar, sym.r_t_bar])
376 Sigma_3_theta_l_r_t = sp.Matrix([
377     [sym.sigma_lambda_theta_l ** 2,
378      sym.rho_theta_l_r_t * sym.sigma_lambda_theta_l * sym.sigma_lambda_r_t],
379     [
380         sym.rho_theta_l_r_t * sym.sigma_lambda_theta_l * sym.sigma_lambda_r_t,
381         sym.sigma_lambda_r_t ** 2]])
382
383 G_3_theta_l_r_t = Normal(name='G_3_theta_l_r_t',
384                           mean=mu_3_theta_l_r_t,

```

```

382         std=Sigma_2_theta_l_r_t)
383 G_3_theta_l_r_t_density = density(G_3_theta_l_r_t)(sym.theta_l, sym.r_t)
384
385 G_theta_l_r_t = ((1 - sp.abc.delta) * sp.abc.alpha * G_1_theta_l_r_t_density +
386                 (1 - sp.abc.delta) * (1 - sp.abc.alpha) * G_2_theta_l_r_t_density +
387                 sp.abc.delta * G_3_theta_l_r_t_density)
388
389 # ---
390
391 mu_1_w_theta_l_r_t = sp.Matrix([sym.w_1, sym.theta_l_1, sym.r_t_1])
392 Sigma_1_w_theta_l_r_t = sp.Matrix(
393     [[sym.sigma_w ** 2,
394      0,
395      0],
396     [0,
397      sym.sigma_theta_l_1 ** 2,
398      sym.r_r_t_theta_l * sym.sigma_theta_l_1 * sym.sigma_r_t_1],
399     [0,
400      sym.r_r_t_theta_l * sym.sigma_theta_l_1 * sym.sigma_r_t_1,
401      sym.sigma_r_t_1 ** 2]])
402
403 G_1_w_theta_l_r_t = Normal(name='G_1_w_theta_l_r_t',
404                             mean=mu_1_w_theta_l_r_t,
405                             std=Sigma_1_w_theta_l_r_t)
406
407 G_1_w_theta_l_r_t_density = density(G_1_w_theta_l_r_t)(sp.abc.w, sym.theta_l, sym.r_t)
408
409 mu_2_w_theta_l_r_t = sp.Matrix([sym.w_2, sym.theta_l_2, sym.r_t_2])
410 Sigma_2_w_theta_l_r_t = sp.Matrix([[sym.sigma_w ** 2,
411                                     0,
412                                     0],
413                                    [0,
414                                     sym.sigma_theta_l_2 ** 2,
415                                     sym.r_r_t_theta_l * sym.sigma_theta_l_2 *
416                                     ↪ sym.sigma_r_t_2],
417                                    [0,
418                                     sym.r_r_t_theta_l * sym.sigma_theta_l_2 *
419                                     ↪ sym.sigma_r_t_2,
420                                     sym.sigma_r_t_2 ** 2]])
421
422 G_2_w_theta_l_r_t = Normal(name='G_2_w_theta_l_r_t',
423                             mean=mu_2_w_theta_l_r_t,
424                             std=Sigma_2_w_theta_l_r_t)
425
426 G_2_w_theta_l_r_t_density = density(G_2_w_theta_l_r_t)(sp.abc.w, sym.theta_l, sym.r_t)
427
428 mu_3_w_theta_l_r_t = sp.Matrix([sym.w_bar, sym.theta_l_bar, sym.r_t_bar])
429 Sigma_3_w_theta_l_r_t = sp.Matrix(
430     [[sym.sigma_lambda_w ** 2,
431      sym.rho_w_theta_l * sym.sigma_lambda_w * sym.sigma_lambda_theta_l,
432      sym.rho_w_r_t * sym.sigma_lambda_theta_l * sym.sigma_lambda_r_t],
433     [
434      sym.rho_w_theta_l * sym.sigma_lambda_w * sym.sigma_lambda_theta_l,
435      sym.sigma_lambda_theta_l ** 2,

```



```

434         sym.rho_theta_l_r_t * sym.sigma_lambda_w * sym.sigma_lambda_r_t],
435     [sym.rho_w_r_t * sym.sigma_lambda_theta_l * sym.sigma_lambda_r_t,
436     sym.rho_theta_l_r_t * sym.sigma_lambda_w * sym.sigma_lambda_r_t,
437     sym.sigma_lambda_r_t ** 2]])
438
439 G_3_w_theta_l_r_t = Normal(name='G_3_w_theta_l_r_t',
440                             mean=mu_3_w_theta_l_r_t,
441                             std=Sigma_2_w_theta_l_r_t)
442 G_3_w_theta_l_r_t_density = density(G_3_w_theta_l_r_t)(sp.abc.w, sym.theta_l, sym.r_t)
443
444 G_w_theta_l_r_t = ((1 - sp.abc.delta) * sp.abc.alpha * G_1_w_theta_l_r_t_density +
445                    (1 - sp.abc.delta) * (1 - sp.abc.alpha) * G_2_w_theta_l_r_t_density +
446                    sp.abc.delta * G_3_w_theta_l_r_t_density)
447
448
449 # -----
450
451 # sigma equations
452 # -----
453
454 def sigma_tilde_w(sigma_w=sym.sigma_w, w_prime_2_bar=sym.w_prime_2_bar,
455                  delta=sp.abc.delta, lambda_w=sym.lambda_w):
456     from sympy import sqrt
457     return ((sigma_w / sqrt(w_prime_2_bar)) *
458            (1 / sqrt((1 - delta * lambda_w) / (1 - delta))))
459
460
461 # -----
462
463 # lambda equations
464 # -----
465
466 def lambda_w(sigma_lambda_w=sym.sigma_lambda_w, w_prime_2_bar=sym.w_prime_2_bar):
467     return sigma_lambda_w ** 2 / w_prime_2_bar
468
469
470 def lambda_theta(sigma_lambda_theta=sym.sigma_lambda_theta_l,
471                 theta_l_prime_2_bar=sym.theta_l_prime_2_bar):
472     return sigma_lambda_theta ** 2 / theta_l_prime_2_bar
473
474
475 def lambda_r(sigma_lambda_r=sym.sigma_lambda_r_t,
476             r_t_prime_2_bar=sym.r_t_prime_2_bar):
477     return sigma_lambda_r ** 2 / r_t_prime_2_bar
478
479
480 def lambda_w_theta(cov_lambda_w_theta=
481                   sym.rho_w_theta_l * sym.sigma_lambda_w * sym.sigma_lambda_theta_l,
482                   w_prime_theta_l_prime_bar=sym.w_prime_theta_l_prime_bar):
483     return cov_lambda_w_theta / w_prime_theta_l_prime_bar
484
485
486 def lambda_w_r(cov_lambda_w_r=
487               sym.rho_w_r_t * sym.sigma_lambda_w * sym.sigma_lambda_r_t,

```

```

488         w_prime_r_t_prime_bar=sym.w_prime_r_t_prime_bar):
489     return cov_lambda_w_r / w_prime_r_t_prime_bar
490
491
492 def lambda_r_theta(cov_lambda_r_theta=
493     sym.rho_theta_l_r_t * sym.sigma_lambda_r_t *
494     ↪ sym.sigma_lambda_theta_l,
495     r_t_prime_theta_l_prime_bar=sym.theta_l_prime_r_t_prime_bar):
496     return cov_lambda_r_theta / r_t_prime_theta_l_prime_bar
497
498 # ---
499
500 # sk
501 # ---
502
503 def sk_theta_l_hat_beta(sk_hat_w=sym.sk_hat_w, c_hat_w_theta_l=sym.c_w_theta_l_hat,
504     beta=sp.abc.beta):
505     return sk_hat_w * c_hat_w_theta_l * (beta + (1 - beta) * c_hat_w_theta_l ** 2)
506
507
508 def sk_theta_l_hat(theta_l_prime_3_bar=sym.theta_l_prime_3_bar,
509     theta_l_prime_2_bar=sym.theta_l_prime_2_bar,
510     delta=sp.abc.delta,
511     lambda_theta=sym.lambda_theta):
512     from sympy import Rational
513     return ((theta_l_prime_3_bar / theta_l_prime_2_bar ** Rational(3, 2)) *
514         (1 / ((1 - delta * lambda_theta) / (1 - delta)) ** Rational(3, 2)) *
515         (1 / (1 - delta)))
516
517
518 def sk_w_hat(sigma_tilde_w=sym.sigma_tilde_w,
519     w_prime_3_bar=sym.w_prime_3_bar,
520     w_prime_2_bar=sym.w_prime_2_bar,
521     delta=sp.abc.delta,
522     lambda_w=sym.lambda_w):
523     from sympy import Rational
524     return ((1 / (1 - sigma_tilde_w ** 2) ** Rational(3, 2)) *
525         (w_prime_3_bar / (w_prime_2_bar ** Rational(3, 2))) *
526         (1 / ((1 - delta * lambda_w) / (1 - delta)) ** Rational(3, 2)) *
527         (1 / (1 - delta)))
528
529
530 def c_w_theta_l_hat(sigma_w_tilde=sym.sigma_tilde_w,
531     w_prime_theta_l_prime_bar=sym.w_prime_theta_l_prime_bar,
532     w_prime_2_bar=sym.w_prime_2_bar,
533     theta_l_prime_2_bar=sym.theta_l_prime_2_bar,
534     delta=sp.abc.delta,
535     lambda_w=sym.lambda_w,
536     lambda_theta=sym.lambda_theta,
537     lambda_w_theta=sym.lambda_w_theta):
538     from sympy import sqrt
539     return ((1 / sqrt(1 - sigma_w_tilde ** 2)) *

```

```

540      (w_prime_theta_l_prime_bar / ((sqrt(w_prime_2_bar)) *
541      ↪ sqrt(theta_l_prime_2_bar))) *
542      (1 / sqrt((1 - delta * lambda_w) / (1 - delta))) *
543      (1 / sqrt((1 - delta * lambda_theta) / (1 - delta))) *
544      ((1 - delta * lambda_w_theta) / (1 - delta)))
545
546 # -- -- -- -- --
547
548 # sk
549 # -- -- -- -- --
550
551 def beta(c_w_theta_l_hat=sym.c_w_theta_l_hat,
552         sk_w_hat=sym.sk_hat_w,
553         sk_theta_l_hat=sym.sk_theta_l_hat):
554     return (((c_w_theta_l_hat ** 3 * sk_w_hat) - sk_theta_l_hat) /
555            (c_w_theta_l_hat * sk_w_hat * (c_w_theta_l_hat ** 2 - 1)))
556
557 # -- -- -- -- --

```

Listing A.2: checked_functions.py

A.4 w_prime_4_bar.py

```

1  #!/usr/bin/env python
2  # coding: utf-8
3
4  # In[2]:
5
6
7  import sympy as sp
8  from IPython.display import display
9  from sympy import abc, oo
10
11  import checked_functions as c_f
12  import symbols as sym
13
14  # # This document aims to analytically check  $\overline{w^4}f$ 
15
16  # ## Define the marginal distributions with those parameters.
17
18  # In[2]:
19
20
21  display(sp.Eq(sym.G_1_w, c_f.G_1_theta_l_density))
22
23  # In[3]:
24
25

```

```

26 display(sp.Eq(sym.G_2_w, c_f.G_2_theta_1_density))
27
28 # In[4]:
29
30
31 display(sp.Eq(sym.G_3_w, c_f.G_3_theta_1_density))
32
33 # In[5]:
34
35
36 display(sp.Eq(sym.G_w, c_f.G_w))
37
38 # Calculate the moment analytically:
39
40 # In[6]:
41
42
43 w_prime_4_bar_int = sp.Integral((sp.abc.w - sym.w_bar) ** 4 * c_f.G_w, [sp.abc.w, -oo,
44 ↪ oo])
45 display(sp.Eq(sym.w_prime_4_bar, w_prime_4_bar_int))
46
47 # In[7]:
48
49
50 w_prime_4_bar_int_val = w_prime_4_bar_int.doit(conds='none').simplify()
51 display(sp.Eq(sym.w_prime_4_bar, w_prime_4_bar_int_val))
52
53 # The equation in the document is:
54
55 # In[8]:
56
57
58 display(sp.Eq(sym.w_prime_4_bar, c_f.w_prime_4_bar()))
59
60 # where
61
62 # In[9]:
63
64
65 display(sp.Eq(sym.sigma_tilde_w, c_f.sigma_tilde_w()))
66
67 # and
68
69 # In[10]:
70
71
72 display(sp.Eq(sym.w_prime_2_bar, c_f.w_prime_2_bar()))
73
74 # and
75
76 # In[11]:
77
78
79 display(sp.Eq(sym.w_prime_3_bar, c_f.w_prime_2_bar()))

```

```

79
80 # So,
81
82 # In[12]:
83
84
85 lambda_w_val = c_f.lambda_w().subs({
86     sym.w_prime_2_bar: c_f.w_prime_2_bar()
87 })
88 display(sp.Eq(sym.lambda_w, lambda_w_val))
89
90 # In[13]:
91
92
93 w_prime_4_bar_check_val = c_f.w_prime_4_bar().subs({
94     sym.w_prime_2_bar: c_f.w_prime_2_bar(),
95     sym.w_prime_3_bar: c_f.w_prime_3_bar(),
96     sym.sigma_tilde_w: c_f.sigma_tilde_w().subs({
97         sym.w_prime_2_bar: c_f.w_prime_2_bar(),
98         sym.lambda_w: c_f.lambda_w()
99     })
100 })
101
102 display(sp.Eq(sym.w_prime_4_bar, w_prime_4_bar_check_val))
103
104 # In[14]:
105
106
107 display(sp.Eq(w_prime_4_bar_int_val, w_prime_4_bar_check_val, evaluate=True)
108     .subs({sym.w_bar: c_f.w_bar()}).simplify())

```

Listing A.3: w_prime_4_bar.py

A.5 w_prime_2_theta_l_prime_bar.py

```

1  #!/usr/bin/env python
2  # coding: utf-8
3
4  # In[2]:
5
6
7  import sympy as sp
8  from IPython.display import display
9  from sympy import abc, oo
10
11  import checked_functions as c_f
12  import symbols as sym
13
14  # # This document aims to analytically check  $\overline{w'^4} \ell$ 

```

```

15
16 # ## Define the marginal distributions with those parameters.
17
18 # In[2]:
19
20
21 display(sp.Eq(sym.G_1_w, c_f.G_1_theta_l_density))
22
23 # In[3]:
24
25
26 display(sp.Eq(sym.G_2_w, c_f.G_2_theta_l_density))
27
28 # In[4]:
29
30
31 display(sp.Eq(sym.G_3_w, c_f.G_3_theta_l_density))
32
33 # In[5]:
34
35
36 display(sp.Eq(sym.G_w, c_f.G_w))
37
38 # Calculate the moment analytically:
39
40 # In[6]:
41
42
43 w_prime_4_bar_int = sp.Integral((sp.abc.w - sym.w_bar) ** 4 * c_f.G_w, [sp.abc.w, -oo,
44 ↪ oo])
45 display(sp.Eq(sym.w_prime_4_bar, w_prime_4_bar_int))
46
47 # In[7]:
48
49
50 w_prime_4_bar_int_val = w_prime_4_bar_int.doit(conds='none').simplify()
51 display(sp.Eq(sym.w_prime_4_bar, w_prime_4_bar_int_val))
52
53 # The equation in the document is:
54
55 # In[8]:
56
57
58 display(sp.Eq(sym.w_prime_4_bar, c_f.w_prime_4_bar()))
59
60 # where
61
62 # In[9]:
63
64
65 display(sp.Eq(sym.sigma_tilde_w, c_f.sigma_tilde_w()))
66
67 # and

```

```

68 # In[10]:
69
70
71 display(sp.Eq(sym.w_prime_2_bar, c_f.w_prime_2_bar()))
72
73 # and
74
75 # In[11]:
76
77
78 display(sp.Eq(sym.w_prime_3_bar, c_f.w_prime_2_bar()))
79
80 # So,
81
82 # In[12]:
83
84
85 lambda_w_val = c_f.lambda_w().subs({
86     sym.w_prime_2_bar: c_f.w_prime_2_bar()
87 })
88 display(sp.Eq(sym.lambda_w, lambda_w_val))
89
90 # In[13]:
91
92
93 w_prime_4_bar_check_val = c_f.w_prime_4_bar().subs({
94     sym.w_prime_2_bar: c_f.w_prime_2_bar(),
95     sym.w_prime_3_bar: c_f.w_prime_3_bar(),
96     sym.sigma_tilde_w: c_f.sigma_tilde_w().subs({
97         sym.w_prime_2_bar: c_f.w_prime_2_bar(),
98         sym.lambda_w: c_f.lambda_w()
99     })
100 })
101
102 display(sp.Eq(sym.w_prime_4_bar, w_prime_4_bar_check_val))
103
104 # In[14]:
105
106
107 display(sp.Eq(w_prime_4_bar_int_val, w_prime_4_bar_check_val, evaluate=True)
108     .subs({sym.w_bar: c_f.w_bar()}).simplify())

```

Listing A.4: w_prime_2_theta_1_prime_bar.py

A.6 w_prime_theta_1_prime_2_bar.py

```

1 #!/usr/bin/env python
2 # coding: utf-8
3

```

```

4  # In[1]:
5
6
7  from itertools import product
8
9  import pandas as pd
10
11  pd.set_option('display.max_columns', None)
12  pd.set_option('display.max_rows', None)
13
14  import sympy as sp
15  from IPython.display import display
16  from sympy import abc, oo, Rational
17
18  import checked_functions as c_f
19  import symbols as sym
20
21  # # This document aims to numerically check  $\int \overline{w' \theta_l'^2} \ell$ 
22
23  # ## Define the marginal distributions.
24
25  # In[2]:
26
27
28  display(sp.Eq(sym.G_1_w_theta, c_f.G_1_w_theta_density))
29
30  # In[3]:
31
32
33  display(sp.Eq(sym.G_2_w_theta, c_f.G_2_w_theta_density))
34
35  # In[4]:
36
37
38  display(sp.Eq(sym.G_3_w_theta, c_f.G_3_w_theta_density))
39
40  # In[5]:
41
42
43  display(sp.Eq(sym.G_w_theta, c_f.G_w_theta, evaluate=False))
44
45  # In[6]:
46
47
48  w_prime_theta_l_2_prime_bar = sp.Integral(
49      (sp.abc.w - sym.w_bar) * (sp.abc.theta - sym.theta_l_bar) ** 2 * c_f.G_w_theta,
50      [sp.abc.w, -oo, oo],
51      [sp.abc.theta, -oo, oo])
52
53  display(sp.Eq(sym.w_prime_theta_prime_l_2_bar, w_prime_theta_l_2_prime_bar))
54
55  # In[7]:
56
57

```



```

58 w_prime_theta_l_2_prime_bar = w_prime_theta_l_2_prime_bar.subs({
59     sym.w_bar: c_f.w_bar(),
60     sym.theta_l_bar: c_f.theta_l_bar()
61 })
62
63 display(sp.Eq(sym.w_prime_theta_prime_l_2_bar, w_prime_theta_l_2_prime_bar))
64
65 # The equation in the document is:
66
67 # In[8]:
68
69
70 display(sp.Eq(sym.w_prime_theta_prime_l_2_bar, c_f.w_prime_theta_l_prime_2_bar()))
71
72 # For this we still need  $\mathcal{L}\tilde{\sigma}_w$ :
73
74 # In[9]:
75
76
77 display(sp.Eq(sym.sigma_tilde_w, c_f.sigma_tilde_w()))
78
79 # And  $\overline{w'^2}$ :
80
81 # In[10]:
82
83
84 display(sp.Eq(sym.w_prime_2_bar, c_f.w_prime_2_bar()))
85
86 # And  $\overline{w'^3}$ :
87
88 # In[11]:
89
90
91 display(sp.Eq(sym.w_prime_3_bar, c_f.w_prime_3_bar()))
92
93 # And  $\overline{w'\theta'_l}$ :
94
95 # In[12]:
96
97
98 display(sp.Eq(sym.w_prime_theta_l_prime_bar, c_f.w_prime_theta_l_prime_bar()))
99
100 # And  $\mathcal{L}\lambda_w$ :
101
102 # In[13]:
103
104
105 display(sp.Eq(sym.lambda_w, c_f.lambda_w()))
106
107 # And  $\mathcal{L}\lambda_{w\theta}$ :
108
109 # In[14]:
110
111

```

```

112 display(sp.Eq(sym.lambda_w_theta, c_f.lambda_w_theta()))
113
114 # And  $\overline{\theta_l^3}$ :
115
116 # In[15]:
117
118
119 display(sp.Eq(sym.theta_l_prime_3_bar, c_f.theta_l_prime_3_bar()))
120
121 # Putting those all together yields:
122
123 # In[16]:
124
125
126 w_prime_theta_l_prime_2_bar_check_val = c_f.w_prime_theta_l_prime_2_bar().subs({
127     sym.theta_l_prime_3_bar: c_f.theta_l_prime_3_bar(),
128     sym.sigma_tilde_w: c_f.sigma_tilde_w(),
129     sym.lambda_w: c_f.lambda_w(),
130     sym.w_prime_2_bar: c_f.w_prime_2_bar(),
131     sym.w_prime_3_bar: c_f.w_prime_3_bar(),
132     sym.lambda_w_theta: c_f.lambda_w_theta(),
133     sym.w_prime_theta_l_prime_bar: c_f.w_prime_theta_l_prime_bar()
134 })
135
136 w_prime_theta_l_prime_2_bar_check_val = w_prime_theta_l_prime_2_bar_check_val.subs({
137     sym.w_prime_2_bar: c_f.w_prime_2_bar(),
138     sym.lambda_w: c_f.lambda_w(),
139     sym.w_bar: c_f.w_bar(),
140     sym.theta_l_bar: c_f.theta_l_bar()
141 })
142
143 display(sp.Eq(sym.w_prime_theta_prime_l_2_bar, w_prime_theta_l_prime_2_bar_check_val))
144
145 # Since the integral is too difficult to be calculated analytically, at least with
146 → sympy, we try to put in some arbitrary numbers for the pdf parameters, to simplify
147 → the equations.
148
149 # We create a dataframe to get all possible permutations and therefore also all possible
150 → evaluations of the integrals.
151
152 # In[17]:
153
154
155 df = pd.DataFrame(
156     product([0, 1],
157             [-2, 2],
158             [-1, 2],
159             [0, 3],
160             [Rational(1, 10)],
161             [Rational(3, 10)],
162             [Rational(4, 10)],
163             [Rational(7, 10)],
164             [Rational(6, 10)],
165             [Rational(5, 10)],

```

```

163         [Rational(1, 10), Rational(5, 10)],
164         [Rational(5, 10)]),
165     columns=[sym.w_1,
166             sym.w_2,
167             sym.theta_l_1,
168             sym.theta_l_2,
169             sym.sigma_theta_l_1,
170             sym.sigma_theta_l_2,
171             sym.sigma_lambda_theta_l,
172             sym.sigma_w,
173             sym.sigma_lambda_w,
174             sp.abc.alpha,
175             sp.abc.delta,
176             sym.rho_w_theta_l])
177
178 # In[18]:
179
180
181 df['check_val'] = (
182     df.apply(lambda x:
183             w_prime_theta_l_prime_2_bar_check_val.subs({
184                 sym.w_1: x[sym.w_1],
185                 sym.w_2: x[sym.w_2],
186                 sym.theta_l_1: x[sym.theta_l_1],
187                 sym.theta_l_2: x[sym.theta_l_2],
188                 sym.sigma_theta_l_1: x[sym.sigma_theta_l_1],
189                 sym.sigma_theta_l_2: x[sym.sigma_theta_l_2],
190                 sym.sigma_lambda_theta_l: x[sym.sigma_lambda_theta_l],
191                 sym.sigma_w: x[sym.sigma_w],
192                 sym.sigma_lambda_w: x[sym.sigma_lambda_w],
193                 sp.abc.alpha: x[sp.abc.alpha],
194                 sp.abc.delta: x[sp.abc.delta],
195                 sym.rho_w_theta_l: x[sym.rho_w_theta_l]
196             }).evalf(), axis=1))
197
198 # Calculate the moment analytically:
199
200 # In[19]:
201
202
203 df['num_int'] = (
204     df.apply(lambda x: w_prime_theta_l_2_prime_bar.subs({
205         sym.w_1: x[sym.w_1],
206         sym.w_2: x[sym.w_2],
207         sym.theta_l_1: x[sym.theta_l_1],
208         sym.theta_l_2: x[sym.theta_l_2],
209         sym.sigma_theta_l_1: x[sym.sigma_theta_l_1],
210         sym.sigma_theta_l_2: x[sym.sigma_theta_l_2],
211         sym.sigma_lambda_theta_l: x[sym.sigma_lambda_theta_l],
212         sym.sigma_w: x[sym.sigma_w],
213         sym.sigma_lambda_w: x[sym.sigma_lambda_w],
214         sp.abc.alpha: x[sp.abc.alpha],
215         sp.abc.delta: x[sp.abc.delta],
216         sym.rho_w_theta_l: x[sym.rho_w_theta_l]

```

```

217     }).doit(conds='none', method='quad').evalf(), axis=1))
218
219 # In[20]:
220
221
222 df['diff'] = abs(df['check_val'] - df['num_int'])
223
224 # In[21]:
225
226
227 df['diff_num'] = abs(df['check_val'].astype(float) - df['num_int'].astype(float))
228
229 # In[22]:
230
231
232 display(df)
233
234 # In[23]:
235
236
237 import numpy as np
238
239 print('The mean error between the rhs and the lhs is:', np.mean(df['diff_num']))

```

Listing A.5: w_prime_theta_l_prime_2_bar.py

A.7 w_prime_r_t_prime_theta_l_prime_bar.py

```

1  #!/usr/bin/env python
2  # coding: utf-8
3
4  # In[1]:
5
6
7  from itertools import product
8
9  import pandas as pd
10
11 pd.set_option('display.max_columns', None)
12 pd.set_option('display.max_rows', None)
13
14 import sympy as sp
15 from IPython.display import display
16 from sympy import abc, oo, Rational
17
18 import checked_functions as c_f
19 import symbols as sym
20
21 # # This document aims to numerically check  $\overline{w'r_t'\theta_l'}$ 

```

```

22
23 # ## Define the normal distributions.
24
25 # In[2]:
26
27
28 display(sp.Eq(sp.symbols('\mu_1'), c_f.mu_1_w_theta_l_r_t, evaluate=False))
29 display(sp.Eq(sp.symbols('\Sigma_1'), c_f.Sigma_1_w_theta_l_r_t, evaluate=False))
30 # display(sp.Eq(sym.G_1_w_theta_l_r_t, c_f.G_1_w_theta_l_r_t_density))
31
32
33 # In[3]:
34
35
36 display(sp.Eq(sp.symbols('\mu_2'), c_f.mu_2_w_theta_l_r_t, evaluate=False))
37 display(sp.Eq(sp.symbols('\Sigma_2'), c_f.Sigma_2_w_theta_l_r_t, evaluate=False))
38 # display(sp.Eq(sym.G_2_w_theta_l_r_t, c_f.G_2_w_theta_l_r_t_density))
39
40
41 # In[4]:
42
43
44 display(sp.Eq(sp.symbols('\mu_3'), c_f.mu_3_w_theta_l_r_t, evaluate=False))
45 display(sp.Eq(sp.symbols('\Sigma_3'), c_f.Sigma_3_w_theta_l_r_t, evaluate=False))
46 # display(sp.Eq(sym.G_3_w_theta_l_r_t, c_f.G_3_w_theta_l_r_t_density))
47
48
49 # In[5]:
50
51
52 display(sp.Eq(sym.G_w_theta_l_r_t, c_f.G_w_theta_l_r_t))
53
54 # In[6]:
55
56
57 w_prime_r_t_prime_theta_l_prime_bar = sp.Integral(
58     (sp.abc.w - sym.w_bar) * (sym.r_t - sym.r_t_bar) * (
59         sp.abc.theta - sym.theta_l_bar) * c_f.G_w_theta_l_r_t,
60     [sp.abc.w, -oo, oo],
61     [sym.theta_l, -oo, oo],
62     [sym.r_t, -oo, oo])
63 display(sp.Eq(sym.w_prime_r_t_prime_theta_l_prime_bar,
64     ↪ w_prime_r_t_prime_theta_l_prime_bar))
65
66 # In[7]:
67
68 w_prime_r_t_prime_theta_l_prime_bar = (
69     w_prime_r_t_prime_theta_l_prime_bar.subs({
70         sym.w_bar: c_f.w_bar(),
71         sym.r_t_bar: c_f.r_t_bar(),
72         sym.theta_l_bar: c_f.theta_l_bar()
73     }))

```

```

74 display(sp.Eq(sym.w_prime_r_t_prime_theta_l_prime_bar,
75 ↪ w_prime_r_t_prime_theta_l_prime_bar))
76
77 # The equation in the document is:
78
79 # In[8]:
80
81 display(sp.Eq(sym.w_prime_r_t_prime_theta_l_prime_bar,
82 ↪ c_f.w_prime_r_t_prime_theta_l_prime_bar()))
83
84 # Since the integral is too difficult to be calculated analytically, at least with
85 ↪ sympy, we try to put in some arbitrary numbers for the pdf parameters, to simplify
86 ↪ the equations.
87
88 # We also use a the method `nquad(..)` from `sympy` to get a numerical evaluation of the
89 ↪ 3d integral.
90
91 # We create a dataframe to get all possible permutations and therefore also all possible
92 ↪ evaluations of the integrals.
93
94 # In[9]:
95
96 df = pd.DataFrame(
97     product([3, 1],
98             [-2],
99             [-1, 3],
100             [2],
101             [1, 4],
102             [2],
103             [1.1],
104             [1.3],
105             [1.4],
106             [1.7],
107             [1.2],
108             [1.5],
109             [1.9],
110             [1.6],
111             [.55],
112             [.8],
113             [.65],
114             [.45],
115             [.35],
116             [.5]),
117     columns=[sym.w_1,
118             sym.w_2,
119             sym.theta_l_1,
120             sym.theta_l_2,
121             sym.r_t_1,
122             sym.r_t_2,
123             sym.sigma_theta_l_1,
124             sym.sigma_theta_l_2,
125             sym.sigma_lambda_theta_l,

```

```

122         sym.sigma_w,
123         sym.sigma_r_t_1,
124         sym.sigma_r_t_2,
125         sym.sigma_lambda_r_t,
126         sym.sigma_lambda_w,
127         sp.abc.alpha,
128         sp.abc.delta,
129         sym.rho_w_theta_l,
130         sym.rho_w_r_t,
131         sym.rho_theta_l_r_t,
132         sym.r_r_t_theta_l])
133
134 # In[10]:
135
136
137 w_prime_r_t_prime_theta_l_prime_bar_check_sym_val = (
138     c_f.w_prime_r_t_prime_theta_l_prime_bar().subs({
139         sym.w_bar: c_f.w_bar(),
140         sym.r_t_bar: c_f.r_t_bar(),
141         sym.theta_l_bar: c_f.theta_l_bar()
142     }))
143
144 # In[11]:
145
146
147 df['check_val'] = (
148     df.apply(lambda x: Rational(c_f.w_prime_r_t_prime_theta_l_prime_bar().subs({
149         sym.w_bar: c_f.w_bar(),
150         sym.theta_l_bar: c_f.theta_l_bar(),
151         sym.r_t_bar: c_f.r_t_bar(),
152         sym.w_1: x[sym.w_1],
153         sym.w_2: x[sym.w_2],
154         sym.theta_l_1: x[sym.theta_l_1],
155         sym.theta_l_2: x[sym.theta_l_2],
156         sym.r_t_1: x[sym.r_t_1],
157         sym.r_t_2: x[sym.r_t_2],
158         sym.sigma_theta_l_1: x[sym.sigma_theta_l_1],
159         sym.sigma_theta_l_2: x[sym.sigma_theta_l_2],
160         sym.sigma_r_t_1: x[sym.sigma_r_t_1],
161         sym.sigma_r_t_2: x[sym.sigma_r_t_2],
162         sp.abc.alpha: x[sp.abc.alpha],
163         sp.abc.delta: x[sp.abc.delta],
164         sym.r_r_t_theta_l: x[sym.r_r_t_theta_l]
165     })).evalf(), axis=1))
166
167 # Calculate the moment numerically:
168
169 # In[12]:
170
171
172 import scipy
173
174 df['num_int'] = df.apply(lambda x: scipy.integrate.nquad(
175     sp.lambdify(

```

```

176 [sp.abc.w, sym.r_t, sym.theta_l],
177 (((sp.abc.w - c_f.w_bar()) *
178 (sym.r_t - c_f.r_t_bar()) *
179 (sym.theta_l - c_f.theta_l_bar()) *
180 c_f.G_w_theta_l_r_t))
181 .subs({
182     sym.w_bar: c_f.w_bar(),
183     sym.r_t_bar: c_f.r_t_bar(),
184     sym.theta_l_bar: c_f.theta_l_bar(),
185     sym.w_1: x[sym.w_1],
186     sym.w_2: x[sym.w_2],
187     sym.theta_l_1: x[sym.theta_l_1],
188     sym.theta_l_2: x[sym.theta_l_2],
189     sym.r_t_1: x[sym.r_t_1],
190     sym.r_t_2: x[sym.r_t_2],
191     sym.sigma_w: x[sym.sigma_w],
192     sym.sigma_theta_l_1: x[sym.sigma_theta_l_1],
193     sym.sigma_theta_l_2: x[sym.sigma_theta_l_2],
194     sym.sigma_lambda_theta_l: x[sym.sigma_lambda_theta_l],
195     sym.sigma_lambda_w: x[sym.sigma_lambda_w],
196     sym.sigma_lambda_r_t: x[sym.sigma_lambda_r_t],
197     sym.sigma_r_t_1: x[sym.sigma_r_t_1],
198     sym.sigma_r_t_2: x[sym.sigma_r_t_2],
199     sp.abc.alpha: x[sp.abc.alpha],
200     sp.abc.delta: x[sp.abc.delta],
201     sym.rho_w_theta_l: x[sym.rho_w_theta_l],
202     sym.rho_w_r_t: x[sym.rho_w_r_t],
203     sym.rho_theta_l_r_t: x[sym.rho_theta_l_r_t],
204     sym.r_r_t_theta_l: x[sym.r_r_t_theta_l]
205     })),
206 ranges=[[-30, 30], [-30, 30], [-30, 30]][0],
207         axis=1)
208
209 # In[13]:
210
211
212 df['diff'] = abs(df['check_val'] - df['num_int'])
213
214 # In[14]:
215
216
217 df['diff_num'] = abs(df['check_val'].astype(float) - df['num_int'])
218
219 # In[15]:
220
221
222 display(df)
223
224 # In[16]:
225
226
227 import numpy as np
228

```


229

```
print('The mean error between the rhs and the lhs is:', np.mean(df['diff_num']))
```

Listing A.6: w_prime_r_t_prime_theta_l_prime_bar.py