

Constraints on halo formation from cross-correlations with correlated variables

Emanuele Castorina^{1,2,3*}, Aseem Paranjape⁴ & Ravi K. Sheth^{5,6}

¹ *Berkeley Center for Cosmological Physics, University of California, Berkeley, CA 94720, USA*

² *Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 93720, USA*

³ *SISSA - International School For Advanced Studies, Via Bonomea, 265 34136 Trieste, Italy*

⁴ *Inter-University Centre for Astronomy and Astrophysics, Ganeshkhind, Post Bag 4, Pune 411007, India*

⁵ *Center for Particle Cosmology, University of Pennsylvania, 209 S. 33rd St., Philadelphia, PA 19104, USA*

⁶ *The Abdus Salam International Center for Theoretical Physics, Strada Costiera, 11, Trieste 34151, Italy*

draft

ABSTRACT

Cross-correlations between biased tracers and the dark matter field encode information about the physical variables which characterize these tracers. However, if the physical variables of interest are correlated with one another, then extracting this information is not as straightforward as one might naively have thought. We show how to exploit these correlations so as to estimate scale-independent bias factors of *all* orders in a model-independent way. We also show that failure to account for this will lead to incorrect conclusions about which variables matter and which do not. Moreover, accounting for this allows one to use the scale dependence of bias to constrain the physics of halo formation; to date the argument has been phrased the other way around. We illustrate by showing that the scale dependence of linear and nonlinear bias, measured on nonlinear scales, can be used to provide consistent estimates of how the critical density for halo formation depends on halo mass. Our methods work even when the bias is nonlocal and stochastic, such as when, in addition to the spherically averaged density field and its derivatives, the quadrupolar shear field also matters for halo formation. In such models, the nonlocal bias factors are closely related to the more familiar local nonlinear bias factors, which are much easier to measure. Our analysis emphasizes the fact that biased tracers are biased because they do not sample fields (density, velocity, shear, etc.) at all positions in space in the same way that the dark matter does.

Key words: cosmology: theory, large-scale structure of Universe – methods: analytical, numerical

1 INTRODUCTION

The point process defined by a galaxy sample is a biased tracer of the underlying dark matter distribution. This is a consequence of the fact that galaxies populate dark matter haloes, and the haloes themselves are biased tracers (e.g. the review in Cooray & Sheth 2002). Nevertheless, because halo bias is reasonably well understood, the clustering of galaxies in large scale sky surveys encodes information about galaxy formation and also provides competitive constraints on cosmological parameters.

Halo bias is commonly defined as the ratio of $\langle \Delta | \text{halo} \rangle$, the cross correlation of the large scale matter fluctuation Δ with the halo distribution (at distance r), to the cross

correlation of Δ with the matter (rather than halo) distribution (at the same distance r): $\langle \Delta | \text{matter} \rangle = \langle \Delta \delta \rangle$. Similarly, if $P_{\Delta|h}(k)$ denotes the Fourier transform of $\langle \Delta | \text{halo} \rangle$, then the Fourier space bias $b_\delta(k)$ is defined by the ratio $P_{\Delta|h}(k)/P_{\Delta\delta}(k)$. The main goal of the present note is to show that there is much to be gained if one generalizes this definition by replacing Δ with any other quantity, say Y , associated with the field at distance r from each halo. The associated Fourier space bias will be $b_Y(k) \equiv P_{Y|h}(k)/P_{Y\delta}(k)$. We will also argue that it is useful to replace the notion of “bias with respect to the dark matter overdensity field” with the slightly more general notion of bias with respect to an average over all space whatever the field. This makes explicit that biased tracers are biased because they do not sample (density, velocity, shear, etc.) fields at all positions in space in the same way that the dark matter does. Although

* E-mail: ecastorina@berkeley.edu

this was the sense in which bias was initially meant, this particular view of bias has gone out of vogue.

In Section 2 we show that, in some cases, the b_y may be the same function of k for more than one choice of Y . When this occurs, we show how cross-correlating the halo field with these different choices for Y allows for less model-dependent parameter constraints than were previously possible. We first show that cross-correlating with different parameters Y provides a simple way of estimating the coefficients which determine the scale dependence of linear bias. We then show how to combine these to provide information about halo formation. Finally, we show how to generalize the method to estimate nonlinear bias factors of all orders. In Section 3 we use Monte-Carlo simulations to show that our approach works well when the density field and its derivatives are the only parameters which matter for halo formation. Section 4 shows that it continues to work well when the shear field also matters. A final section summarizes.

2 BIAS OF CONSTRAINED REGIONS

2.1 Overdense patches in the initial conditions

To begin, suppose that we identify those positions in the initial field which, when smoothed on scale R_p have $\delta_p \geq \delta_c$. This is not a particularly good model of protohaloes – patches in the initial overdensity fluctuation field which are destined to form haloes – but it allows us to highlight a number of interesting points before we move on to more realistic models.

The cross-correlation between the overdensity δ_q smoothed on scale $R_q \neq R_p$ and δ_p , when it is known that $\delta_p \geq \delta_c$ (the subscript c denotes ‘critical’; it does not denote a scale R_c) is given by

$$\langle \delta_q | \mathcal{C}_p \rangle = \int_{\delta_c}^{\infty} d\delta_p p(\delta_p) \frac{\langle \delta_q | \delta_p \rangle}{p(\mathcal{C}_p)}, \quad (1)$$

where $\mathcal{C}_p \equiv (\delta_p \geq \delta_c)$. Since

$$\langle \delta_q | \delta_p \rangle \equiv \frac{\langle \delta_q \delta_p \rangle}{\langle \delta_p \delta_p \rangle} \delta_p, \quad (2)$$

we have that

$$\langle \delta_q | \mathcal{C}_p \rangle = b_L(\nu) \langle \delta_q \delta_p \rangle, \quad (3)$$

where

$$b_L(\nu) \equiv \frac{\nu}{\delta_c} \frac{e^{-\nu^2/2}/\sqrt{2\pi}}{\operatorname{erfc}(\nu/\sqrt{2})/2} \quad \text{with} \quad \nu \equiv \frac{\delta_c}{\sqrt{\langle \delta_p \delta_p \rangle}}. \quad (4)$$

Evidently, the cross-correlation between the positions identified as satisfying the constraint \mathcal{C}_p on scale R_p and the field smoothed on scale R_q is linearly proportional to the cross-correlation of the field itself when smoothed on the two scales R_p and R_q . (Frusciante & Sheth 2012, show that this is a generic feature of local Lagrangian bias models such as this one.) The constant of proportionality defines the linear bias factor. It is a complicated function of the threshold, but is otherwise just a number which is independent of the scale R_q .

For what follows, it is useful to also study the Fourier transform of the cross-correlation, which is given by

$$P_{\delta|c}(k) = b_L(\nu) P_{\delta\delta}(k) W(kR_q) W(kR_p), \quad (5)$$

where the W denote the Fourier transforms of the smoothing windows. Although the most natural choice has $W(x)$ being the same function of x on both scales R_p and R_q , the formalism is sufficiently general that different functional forms W for the different scales are permitted: we could have written the two W s in the expression above as $W_q(kR_q)$ and $W_p(kR_p)$.

In the example above, δ_p is special, since the constraint \mathcal{C}_p depends only on δ on the scale R_p . This raises the question of whether or not we can infer, from cross-correlation measurements alone, that the constraint depends only on the physical parameter δ smoothed on scale R_p . (In anticipation of the discussion in the next section, it may be useful to think of $y_p \equiv d\delta_p/d\ln \langle \delta_p \delta_p \rangle$ as a specific example of another variable on scale R_p .) For example, one might measure the cross-correlation $\langle \delta_q \delta_r \rangle$, where δ_r denotes the field smoothed on scale R_r and the average is over all positions, not just those which have $\delta_p \geq \delta_c$. Then the ratio

$$\frac{\langle \delta_q | \mathcal{C}_p \rangle}{\langle \delta_q \delta_r \rangle} = b_L(\nu) \frac{\langle \delta_q \delta_p \rangle}{\langle \delta_q \delta_r \rangle}. \quad (6)$$

If one repeated this analysis for a range of R_r and R_q , one would find that the real-space ratio was a function of R_q except when $R_r = R_p$, when the ratio would equal the constant $b_L(\nu)$. One might then conclude that R_p was special. Analysis of the ratio of the associated Fourier space quantities would lead to a similar conclusion, since

$$\frac{P_{\delta|c}(k)}{P_{\delta\delta}(k)W(kR_q)W(kR_r)} = b_L(\nu) \frac{W(kR_p)}{W(kR_r)} \quad (7)$$

is a function of k except when $R_r = R_p$, when it equals $b_L(\nu)$.

This is not quite as trivial as it sounds of course, since one is also free to vary the shape of the smoothing filter – which may not be known *a priori* – but the logic of the approach is clear. In this case, the Fourier space approach is more illuminating, since if $W(kR_r) \rightarrow 1$ when $R_r \rightarrow 0$ (as is true for the uncompensated filters which are commonly used in this context), then the Fourier transform of $\langle \delta_q \delta_r \rangle \rightarrow P_{\delta\delta}(k)W(kR_q)$. In this limit, the ratio $P_{\delta|c}(k)/[P_{\delta\delta}(k)W(kR_q)] = b_L(\nu)W(kR_p)$, and since b_L does not depend on k , one has determined the shape of $W(kR_p)$.

Of course, in practice one may not know that δ is the relevant physical variable, so one may not know that the relevant cross-correlation to study is $\langle \delta_q \delta_r \rangle$ rather than, say, $\langle \delta_q y_r \rangle$. However, scale independence is a useful guide here too. For example, if $W(kR_r) = \exp(-k^2 R_r^2/2)$ and $y_r = d\delta_r/d\ln \langle \delta_r^2 \rangle$ then $\langle \delta_q y_r \rangle$ will differ from $\langle \delta_q \delta_r \rangle$ by a term that is proportional to k^2 . So, even if $R_r = R_p$, the ratio $P_{\delta|c}(k)/[P_{y\delta}(k)W(kR_q)]$ will be k -dependent.

One might have thought that since \mathcal{C}_p depends only on δ_p , cross-correlating with the overdensity on another scale is a natural choice, and this is why the analysis yields a ‘linear’, ‘scale-independent’ bias factor. This would be a concern, since it suggests that to measure the linear bias factor, one needs to know *a priori* which variable (or variables) were important for determining what the constraint was.

To see why this is incorrect, suppose that, instead of cross-correlating the special positions with δ_q , one cross-correlated with some other variable y_q defined on scale R_q . (Again, it may be useful to think of $y_q \equiv d\delta_q/d\ln \langle \delta_q \delta_q \rangle$ as

a specific example.) Then the same logic as above yields

$$\langle y_q | \mathcal{C}_p \rangle = b_L(\nu) \langle y_q \delta_p \rangle, \quad (8)$$

with associated Fourier transform

$$P_{y|c}(k) = b_L(\nu) P_{y\delta}(k) W(kR_q) W(kR_p). \quad (9)$$

In this case, because one has cross-correlated with y , one would study the ratio

$$\frac{\langle y_q | \mathcal{C}_p \rangle}{\langle y_q \delta_r \rangle} = b_L(\nu) \frac{\langle \delta_q \delta_p \rangle}{\langle y_q \delta_r \rangle} \quad (10)$$

or

$$\frac{P_{y|c}(k)}{P_{y\delta}(k)W(kR_q)W(kR_r)} = b_L(\nu) \frac{W(kR_p)}{W(kR_r)} \quad (11)$$

for a range of R_r , and would find the ratio to be independent of R_q or k only if $R_r = R_p$. Notice that the bias factor is the *same* for all choices of y_q . In particular, b_L is the same as when $y_q = \delta_q$, showing that there is nothing special about δ_q . I.e., cross-correlating with *any* variable which correlates with δ_p will yield the same ‘linear’ bias factor – this generalizes the argument in [Frusciante & Sheth \(2012\)](#). The lesson is that the bias factor itself contains information about the important variable (i.e., that it is δ_p which must exceed a threshold), and this information can be obtained from *any* quantity which correlates with it.

Since it is not necessary that one choose the same variable smoothed on another scale, one no longer need know *a priori* that it was δ_p which mattered. This is a significant simplification. There are, of course, signal-to-noise issues, since one estimates the bias by dividing $\langle y_q | \delta_p \geq \delta_c \rangle$ by $\langle y_q \delta_p \rangle$, so it is important that $\langle y_q \delta_p \rangle$ be far from zero. This suggests one should choose y_q for which $\langle y_q \delta_p \rangle$ is larger, so in this sense knowing that δ_p matters is important. But note that this is a measurement issue, and not one of principle. In principle, one can determine the linear bias factor b_L having much less knowledge of which variable actually matters (in this case δ_p) than one might naively have thought.

Finally, from the analysis above, it should be clear that there was nothing special about δ . For example, one might instead have set $\mathcal{C}_p = (y_p \geq y_c)$ for some other physical parameter y . Then, the analysis above implies that it is cross-correlations with y which will play a fundamental role; cross-correlations with δ will be k -dependent; and the linear scale independent bias b_L factor will be a function of $\nu = y_c / \langle y_p^2 \rangle^{1/2}$. This is the sense in which it is the bias with respect to unconstrained averages (of the y field, in this case) which is simple and linear; the bias with respect to the δ field will appear more complicated.

2.2 When more than one variable matters

The previous subsection studied constrained averages in which the only constraint was on the overdensity field on one scale (or, more generally, on one variable on one scale). The excursion set approach asserts that protohalo patches satisfy more complex constraints. In particular, in the upcrossing approximation associated with the simplest excursion set approach, protohaloes of mass m are associated with regions of size $R_p \propto m^{1/3}$ in the initial conditions where $\delta_p \geq \delta_c(R_p)$ and $\delta(R_p + \Delta R) < \delta_c(R_p + \Delta R)$ ([Bond et al. 1991](#); [Musso & Sheth 2012](#)). Notice that this is like requiring that both δ

and its derivative exceed a threshold; it is in this sense that this is a more complex model than the previous one.

In this approximation, the cross-correlation between protohaloes and some other quantity Y in the initial fluctuation field is

$$\langle Y | \mathcal{C}_p \rangle = f(s)^{-1} \int_{\delta'_c}^{\infty} dv (v - \delta'_c) p(\delta_c, v) \langle Y | \delta_c, v \rangle \quad (12)$$

where $v \equiv d\delta_p/ds$ with $s \equiv \langle \delta_p^2 \rangle$, and

$$f(s) = \int_{\delta'_c}^{\infty} dv (v - \delta'_c) p(\delta_c, v), \quad (13)$$

where $\delta'_c \equiv d\delta_c(s)/ds$.

Setting $x \equiv v/\sqrt{\langle v^2 \rangle} \equiv v/\sqrt{s_v}$ and $\gamma \equiv \langle v \delta_p \rangle / \sqrt{\langle \delta_p^2 \rangle \langle v^2 \rangle} = (2\sqrt{s s_v})^{-1}$ yields

$$\langle Y | \delta_c, v \rangle = \langle Y \delta_p \rangle \frac{\nu - \gamma x}{\sqrt{s(1 - \gamma^2)}} + \langle Y v \rangle \frac{x - \gamma \nu}{\sqrt{s_v(1 - \gamma^2)}}. \quad (14)$$

The terms $\langle Y \delta_p \rangle$ and $\langle Y v \rangle$ depend differently on scale because $\langle Y v \rangle = d\langle Y \delta_p \rangle/ds$. Since we define bias as the ratio of the halo- Y correlation with respect to $\langle Y \delta_p \rangle$, the term which is proportional to $\langle Y \delta_p \rangle$ will be independent of scale. To match notation with previous work, we call this scale-independent term b_{10} . Note that its amplitude is set by the variables which determine halo formation (in this case δ_p and v) and *not* by the variable Y with which we chose to cross correlate the halo field. In the example above,

$$b_{10} = \frac{\delta_c - \langle v | \mathcal{C}_p \rangle / 2s_v}{s(1 - \gamma^2)} \quad (15)$$

where $\langle v | \mathcal{C}_p \rangle = f(s)^{-1} \int_{\delta'_c}^{\infty} dv v (v - \delta'_c) p(\delta_c, v)$. A little algebra then shows that

$$b_1 \equiv \frac{\langle Y | \mathcal{C}_p \rangle}{\langle Y \delta_p \rangle} = b_{10} + \epsilon_{Y\delta_p} b_{11}, \quad (16)$$

where

$$\epsilon_{Y\delta_p} \equiv 2 \frac{d \ln \langle Y \delta_p \rangle}{d \ln s} \quad (17)$$

and

$$b_{11} \equiv \frac{\delta_c}{s} - b_{10}. \quad (18)$$

Whereas b_{11} , like b_{10} , depends only on the quantities which determine \mathcal{C}_p and not on the Y field, the term it multiplies is scale dependent in real space and k -dependent in Fourier space. This is particularly easy to see if we write the ratio of the constrained to unconstrained power spectra:

$$\frac{P_{Y|c}(k)}{P_{Y\delta}(k)} = b_{10} W(kR_p) + 2 \frac{dW(kR_p)}{d \ln s} b_{11}, \quad (19)$$

where we have assumed that the δ in $P_{Y\delta}(k)$ was unsmoothed (i.e. it is δ_r when $R_r \rightarrow 0$), since neither the shape of $W(kR_p)$ nor the scale R_p are known *a priori*. E.g., if $W(kR_p) = \exp(-k^2 R_p^2 / 2)$, then the right hand side equals b_{10} plus a term which is proportional to $R_p^2 k^2$ all multiplied by $W(kR_p)$.

Notice that the right hand side of equation (19) is independent of anything to do with Y . (Things are not as simple in real-space, since correlations with Y appear on the right-hand side of equation 16, a fact we exploit shortly.) This shows explicitly that, even though the constraints \mathcal{C}_p are

more complicated than in the previous section, and they result in k -dependent bias (previously the ratio of constrained to unconstrained correlations was independent of k), one gets the *same* k -dependent bias by cross-correlating the protohaloes with *any* field Y which is correlated with one of the variables which determine the constraints \mathcal{C}_p .

Although this is a straightforward generalization of what we found in the previous section – so the same caveats about signal-to-noise apply here too – it has an important consequence. This is because the traditional estimator of b_{10} sets $Y = \delta_q$ with $R_q \gg R_p$, and measures the ratio with respect to $\langle Y \delta_p \rangle$. The choice of large R_q is motivated by the fact that the b_{10} term dominates in this limit, as can be seen by studying the $kR_p \ll 1$ limit of the Fourier space expressions. But the decision to set $Y = \delta$ rather than some other field is motivated by the fact that \mathcal{C}_p depends on δ_p , and the b_{10} term is the prefactor of the $\langle \delta_q \delta_p \rangle$ correlation. Since b_{11} is the prefactor of the term involving the correlations with $d\delta_p/ds$ (the other variable which matters for \mathcal{C}_p), one might naively have thought that setting $Y = d\delta_q/d\langle \delta_q^2 \rangle$ and looking for the k -independent part is the way to isolate b_{11} . The analysis above shows that if one defines the bias by ratioing to $\langle Y \delta_p \rangle$ as is conventional, then this will not work: the k -independent part will still be b_{10} .

To get b_{11} from scale-independence, equation (16) shows that one must ratio to $\langle Y d\delta_p/d \ln s \rangle$ instead. Doing so will make the term which multiplies b_{10} increase at $k \ll R_p$, so it is only at larger k that the b_{11} term may dominate. However, at these larger k , the term which multiplies b_{11} in equation (19) will show some scale dependence because we are potentially in the regime where the k -dependence of $W(kR_p)$ matters (it may help to think of a Gaussian filter here). As a result, k -independent bias will not be as trivial to identify as for b_{10} . Nevertheless, the distinction between these two procedures – which we may write schematically as $\langle v_q | \mathcal{C}_p \rangle / \langle v_q \delta_p \rangle$ and $\langle \delta_q | \mathcal{C}_p \rangle / \langle \delta_q v_p \rangle$ – has led to some confusion in the literature, as we discuss in the next section.

We turn, therefore, to the question of estimating b_{11} in some other way. We have already noted that the traditional estimator of b_{10} sets $Y = \delta_q$ with $R_q \gg R_p$. Musso et al. (2012) note that b_{11} is then simply determined by equation (18). However, equation (18) assumes prior knowledge of the physics of collapse – in this case, that δ_c is the same constant for all values of s , and that δ is the important variable. Hence, it is interesting to see if we can determine b_{11} without prior knowledge of how δ_c depends on s .

One possibility is to fit the Fourier space ratio using equation (19). This requires prior knowledge that only $W(x)$ and $dW(x)/dx$ matter, which boils down to knowing that δ and its derivative both mattered for \mathcal{C}_p . Since W is usually close to a tophat with slightly rounded edges, it is relatively straightforward to fit for the scale R_p as well as the bias factors b_{10} and b_{11} (e.g. Chan et al. 2015).

In real space, equation (16) shows that we can isolate b_{11} , the term which controls the scale dependence, if we compute the (real space) bias for two choices of Y , and subtract the two expressions. (This cannot be done in Fourier space, since the right hand side of equation 19 is independent of Y , so it would yield zero!) Doing so yields

$$\frac{\langle Y_1 | \mathcal{C}_p \rangle}{\langle Y_1 \delta_p \rangle} - \frac{\langle Y_2 | \mathcal{C}_p \rangle}{\langle Y_2 \delta_p \rangle} = 2b_{11} \frac{d \ln[\langle Y_1 \delta_p \rangle / \langle Y_2 \delta_p \rangle]}{d \ln s}, \quad (20)$$

which can be rearranged to provide a practical estimator for b_{11} which *does not require* prior knowledge of $\delta_c(s)$. With b_{11} in hand, one can go on to estimate b_{10} from equation (16), again without prior knowledge of $\delta_c(s)$. In fact, equation (18) shows that if one adds the estimators of b_{10} and b_{11} , then this furnishes an estimate of $\delta_c(s)$.

Note that the Y_i can be $\delta(R_i)$ on two different scales R_i , or one can be δ_i and the other $d\delta_j/d\langle \delta_j^2 \rangle$, and so on. In practice, some of these combinations will provide higher signal-to-noise estimators of b_{11} than others. For example, if Y_1 is the first variable, then the natural choice for the second variable Y_2 would actually be $Y_2 - \langle Y_2 | Y_1 \rangle$. But the point is that each pair furnishes an estimate of b_{11} (and b_{10}) which does not depend on any assumptions about $\delta_c(s)$. However, they all require prior knowledge of $W(kR_p)$ (to estimate the $\langle Y_i \delta_p \rangle$ terms).

Of course, we could have played this the other way round. Had we started from $\langle Y | \mathcal{C}_p \rangle / \langle Y v \rangle$ rather than $\langle Y | \mathcal{C}_p \rangle / \langle Y \delta_p \rangle$, then the constrained and unconstrained averages for the two different Y_i could be rearranged to provide an estimator of b_{10} rather than b_{11} . Equation (16) then yields b_{11} , and equation (18) yields $\delta_c(s)$.

For what follows, it is useful to formulate the analysis above in matrix notation. We have argued that, for any two variables, Y_1 and Y_2 ,

$$\begin{pmatrix} b_{2Y_1} \\ b_{2Y_2} \end{pmatrix} = \begin{pmatrix} 1 & \epsilon_{Y_1 \delta_p} \\ 1 & \epsilon_{Y_2 \delta_p} \end{pmatrix} \begin{pmatrix} b_{10} \\ b_{11} \end{pmatrix}, \quad (21)$$

Therefore, estimates of the left hand side, say \widehat{b}_{1Y_1} and \widehat{b}_{1Y_2} , can be turned into estimates of b_{10} and b_{11} because

$$\begin{pmatrix} b_{10} \\ b_{11} \end{pmatrix} = \begin{pmatrix} 1 & \epsilon_{Y_1 \delta_p} \\ 1 & \epsilon_{Y_2 \delta_p} \end{pmatrix}^{-1} \begin{pmatrix} \widehat{b}_{1Y_1} \\ \widehat{b}_{1Y_2} \end{pmatrix}. \quad (22)$$

2.3 Generalization to higher order bias

Define $y \equiv Y/\langle Y^2 \rangle^{1/2}$. Then equation (12) can be generalized to

$$\begin{aligned} b_n &\equiv (-1)^n \frac{\langle Y^2 \rangle^{n/2}}{\langle Y \delta_p \rangle^n} \langle H_n(y) | \mathcal{C}_p \rangle \\ &= \int_{\delta'_c}^{\infty} dv (v - \delta'_c) \left(\frac{\partial}{\partial \delta_c} + \frac{\langle Y v \rangle}{\langle Y \delta_p \rangle} \frac{\partial}{\partial v} \right)^n \frac{p(\delta_c, v)}{f(s)}, \\ &= \sum_{j=0}^n \binom{n}{j} b_{nj} \epsilon_{Y \delta_p}^j, \end{aligned} \quad (23)$$

where the $H_n(y)$ are the probabilist's Hermite polynomials. The second line follows from recognizing that all the steps in Appendix A2 of Musso et al. (2012) go through unchanged if one replaces their δ_0 (our Δ) with any other variable Y . Of course, one must recognize that their $S_x \rightarrow \langle Y \delta_p \rangle$ and their $\epsilon_x \rightarrow \epsilon_{Y \delta_p}$.

The key point here is that the derivatives with respect to δ_c and v generate the *same* bias coefficients b_{nj} for all choices of Y . E.g., for $n = 2$,

$$b_2 = b_{20} + 2b_{21} \epsilon_{Y \delta_p} + b_{22} \epsilon_{Y \delta_p}^2. \quad (24)$$

We can estimate these from the data, with no knowledge of δ_c , using the same approach as for b_1 , except that now we need three rather than two different Y s. Following the logic

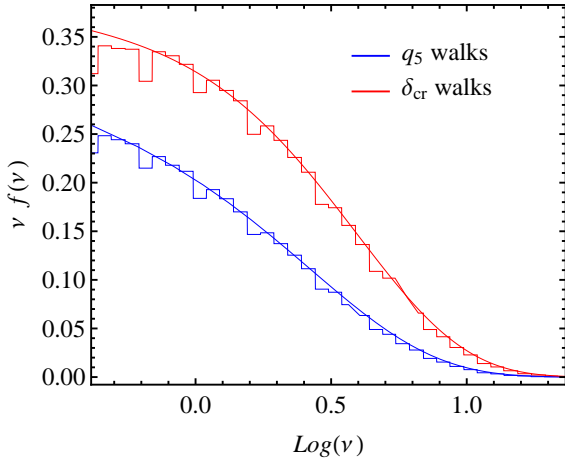


Figure 1. First crossing distribution of a barrier of height $\delta_c (1 + q/q_c)$ for Gaussian smoothing of walks having a Λ CDM $P(k)$. Histograms show Monte-Carlo first crossing distributions for $q_c = 8$ (lower) and ∞ (upper); curves show the analytic prediction. The agreement indicates that our upcrossing approach is a good approximation, even for the stochastic barrier model which we consider in the next section.

of the previous section, we first write

$$\begin{pmatrix} b_{2Y_1} \\ b_{2Y_2} \\ b_{2Y_3} \end{pmatrix} = \begin{pmatrix} 1 & 2\epsilon_{Y_1\delta_p} & \epsilon_{Y_1\delta_p}^2 \\ 1 & 2\epsilon_{Y_2\delta_p} & \epsilon_{Y_2\delta_p}^2 \\ 1 & 2\epsilon_{Y_3\delta_p} & \epsilon_{Y_3\delta_p}^2 \end{pmatrix} \begin{pmatrix} b_{20} \\ b_{21} \\ b_{22} \end{pmatrix}, \quad (25)$$

which shows that

$$\begin{pmatrix} b_{20} \\ b_{21} \\ b_{22} \end{pmatrix} = \begin{pmatrix} 1 & 2\epsilon_{Y_1\delta_p} & \epsilon_{Y_1\delta_p}^2 \\ 1 & 2\epsilon_{Y_2\delta_p} & \epsilon_{Y_2\delta_p}^2 \\ 1 & 2\epsilon_{Y_3\delta_p} & \epsilon_{Y_3\delta_p}^2 \end{pmatrix}^{-1} \begin{pmatrix} \widehat{b}_{2Y_1} \\ \widehat{b}_{2Y_2} \\ \widehat{b}_{2Y_3} \end{pmatrix}. \quad (26)$$

We describe our estimators for the various \widehat{b}_{2Y} in the next section. The consistency relation – the analogue of equation (18) – becomes

$$b_{20} + 2b_{21} + b_{22} = H_2 (\delta_c / \sqrt{s_p}) / s_p. \quad (27)$$

Written this way, the generalization to b_n is trivial: One makes $n + 1$ estimates of b_{nY_i} and then inverts the associated matrix (whose ij th element is $\epsilon_{i \times}^j n! / j!(n - j)!$). The consistency relation is

$$\sum_{j=0}^n \binom{n}{j} b_{nj} = H_n (\delta_c / \sqrt{s_p}) / s_p^{n/2}. \quad (28)$$

Finally, although we do not exploit this in what follows, it is worth noting that, with these replacements, equation (34) of [Musso et al. \(2012\)](#) expresses the conditional distribution $f(s|y)$ as a Taylor series in y (these replacements must also be made in their expression for the quantity they call \bar{c}). Previous work has only really considered $f(s|\Delta)$ as a Taylor series in Δ . It is in this sense that our analysis generalizes the usual notion of bias from being with respect to the large scale environment to any other variable.

3 NUMERICAL TESTS

We now use measurements in Monte-Carlo realizations of random walks crossing a barrier to illustrate some of these

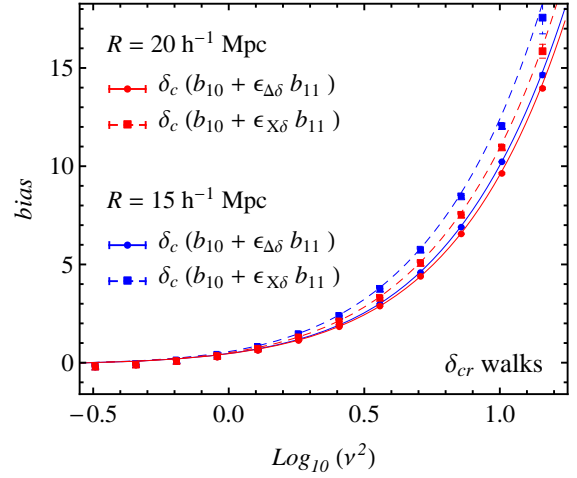


Figure 2. Bias factor b_1 , shown as a function of $\nu^2 \equiv \delta_c^2/s$, estimated from the mean value of Y on the scale R when the first crossing scale was $s > S(R)$ (symbols). Filled circles show measurements for $Y_1 = \Delta$ for two different smoothing scales R , while filled squares are for $Y_2 = d\Delta/d \ln S = X$. Curves show equation (16), where $\epsilon_{Y_i\delta} \equiv 2d \ln \langle Y_i \Delta \rangle / d \ln S$.

ideas. In all cases, walks were generated using a Λ CDM $P(k)$ (Planck cosmology) with correlations between steps being due to Top-Hat smoothing filters. The upper histogram in Figure 1 shows the first crossing distribution of a barrier of height δ_c (the quantity $\nu \equiv \delta_c / \sqrt{s}$ where s is the first crossing scale s). The curve passing through shows the prediction which is based on the assumption that it is only the walk height and its derivative on the upcrossing scale which matter. This agreement is important, since the bias formulae which we wish to test also result from this ‘upcrossing’ approximation, our equation (13).

The lower histogram and associated curve show results when more than one independent variable determines if the barrier δ_c has been crossed (a model we describe in more detail in Section 4).

With a slight abuse of notation, let b_Y denote the mean value of Y when smoothed on scale S around a walk which first crossed the barrier on scale $s \neq S$ (typically $s > S$), in units of the unconstrained cross correlation between Y and δ . I.e., b_Y is b_1 when the cross-correlation is with the variable Y . Then, we can estimate it from the walks by measuring ([Szalay 1988](#); [Musso et al. 2012](#))

$$\widehat{b}_Y = \frac{1}{N} \sum_{\alpha=1}^N \frac{Y_\alpha}{\langle Y \delta_p \rangle} \quad (29)$$

where the sum is over those walks which first crossed $\delta_c(s)$ on scale s . Note that \widehat{b}_Y will be a function of the scale on which Y was estimated, as well as of the constraint: s and $\delta_c(s)$. The denominator shows that it requires knowledge of the fact that the scale s is special. In contrast, $\widehat{b}_Y \langle Y \delta_p \rangle / \langle Y^2 \rangle$, the quantity studied by [Musso et al. \(2012\)](#), does not require this knowledge.

Figure 2 shows \widehat{b}_Y for two choices of Y : $Y_1 = \Delta$ (filled circles) and $Y_2 = d\Delta/d \ln S$ (filled squares), and two choices of smoothing scale R (as indicated). In all cases, δ is in units of δ_c so the normalization factor in equation (29) is $\langle Y \delta_p \rangle / \delta_c$, and we show s as $\nu^2 \equiv \delta_c^2/s$.

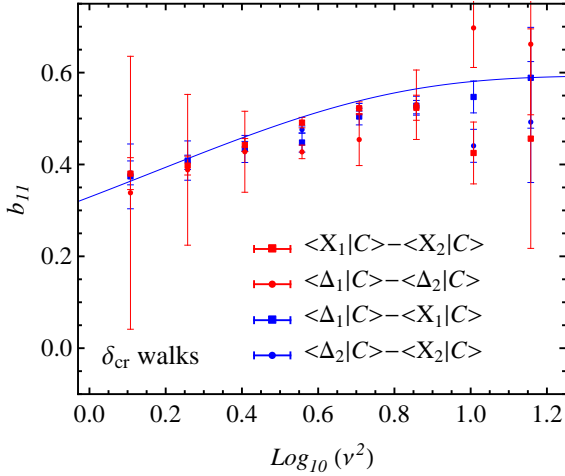


Figure 3. Bias factor b_{11} estimated using different combinations of $Y_1 = \Delta$ and $Y_2 = d\Delta/dS$, on the scales $R = 15 h^{-1} \text{Mpc}$ and/or $R = 20 h^{-1} \text{Mpc}$. Curve shows the expected value (equation 20).

First, notice that both choices of Y result in similar signals, and this signal is rather similar for different R . To show that we understand the small differences between Y 's, as well as the (smaller!) differences between smoothing scales, the curves show equation (16) for the appropriate values of Y , R and s . Since we have normalized our measurements by $\langle Y\delta_p \rangle / \delta_c$ rather than $\langle Y\delta_p \rangle$ itself, the curves actually show the right hand side of equation (16) multiplied by δ_c .

In all cases the curves pass through the associated symbols, indicating that both choices of Y estimate b_{10} plus a small correction factor which depends on b_{11} (and slightly on scale R). Whereas it is well-known that the cross-correlation with Δ yields what is essentially an estimate of b_{10} , this shows explicitly that the cross-correlation with $2d\Delta/d \ln S$ also estimates b_{10} (rather than b_{11}). This illustrates one of our main results.

Previous work (e.g. Musso et al. 2012) turns this estimate of b_1 into an estimate of b_{10} using the consistency relation (equation 18). However, we argued that one can estimate b_{11} directly – with no assumptions about δ_c – by differencing b_1 estimates derived from two different choices of Y , or the same Y at different smoothing scales. E.g., if $Y_1 = \Delta$ and $Y_2 = X \equiv d\Delta/dS$, then

$$\hat{b}_{11}(s) = \frac{1}{N} \sum_{\alpha=1}^N \frac{\Delta_{\alpha} / \langle \Delta \delta \rangle - X_{\alpha} / \langle X \delta \rangle}{\epsilon_{\Delta\delta} - \epsilon_{X\delta}} \quad (30)$$

where $\epsilon_{Y\delta}$ was defined in equation (17).

Figure 3 illustrates this for a number of different combinations of Y . The symbols show four different estimates of b_{11} for the walks with constant barrier δ_c and the curve shows equation (20). Measurements in blue were obtained from equation (30) at the same smoothing scales that were used in Figure 2. The other sets of symbols (red circles and squares) show the result of estimating the bias from a single field but at two different smoothing scales. Clearly, some estimates are noisier than others, but over all, there is general agreement with the prediction.

Most of the measurement error comes from the smallness of the $\epsilon_{Y\delta}$ factors. E.g., estimates from Δ at two (not

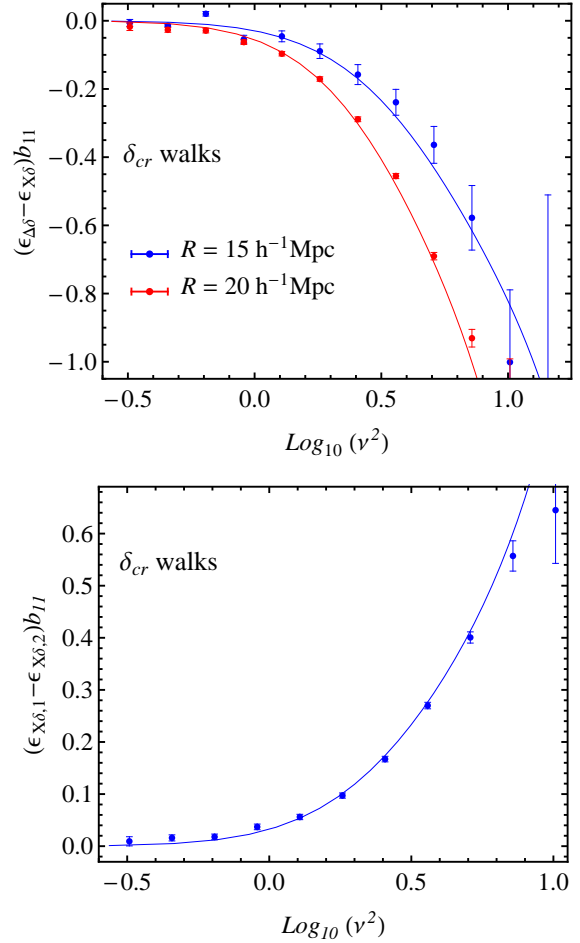


Figure 4. Same as previous Figure, but now \hat{b}_{11} has been multiplied by the appropriate $\epsilon_{Y\delta}$ coefficient (the determinant of the 2×2 matrix in equation (22)). Top panel shows equation (30), which uses Δ and X on the same smoothing scale R ; the two sets of symbols are for two choices of R . Bottom panel uses X on the two different scales. Smooth curves show the associated predictions.

very different) scales do not provide good estimates of b_{11} , because the differences between the two sets of measurements are small (c.f. Figure 2). To study this further, we rescaled the measurements by the determinant of the matrix in equation (22) (for equation 30, this factor is $[\epsilon_{\Delta\delta} - \epsilon_{X\delta}]$). Figure 4 shows the results. The top panel shows the estimator of equation (30), which uses Δ and X on the same scale R (the two sets of symbols are for two choices of R); the bottom panel uses X on two different scales for which the multiplicative factor is $[\epsilon_{X\delta}(R_1) - \epsilon_{X\delta}(R_2)]$; and the smooth curves show the associated predictions. Most of the noise in the previous Figure is gone, so it is easy to see the predictions and measurements are in excellent agreement.

The next step is to use the estimates of b_{11} shown in Figure 2 to correct the b_1 measurements shown in Figure 2 (using equation 16) and so obtain b_{10} without any assumptions about $\delta_c(s)$. Since our estimator for b_{10} involves subtracting our estimators for b_1 and b_{11} , it can be written as a weighted sum of the two profiles \hat{b}_{Δ} and \hat{b}_X that we used

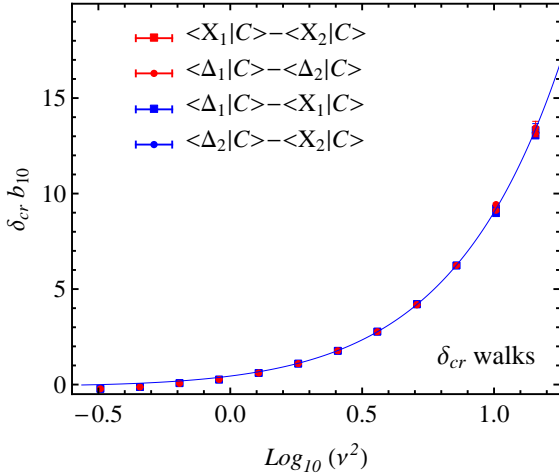


Figure 5. Comparison of b_{10} with a number of estimates which were obtained from combining the Hermite polynomial weighted estimate of b_1 shown in Figure 2 with the estimates of b_{11} based on equation (20), for the different choices of Y shown in the previous Figure (overdensity or its derivative measured on the same or different scales).

to construct b_{11} :

$$\widehat{b}_{10}(s) = \frac{\epsilon_{X\delta}}{\epsilon_{X\delta} - \epsilon_{\Delta\delta}} \widehat{b}_{\Delta} + \frac{\epsilon_{\Delta\delta}}{\epsilon_{\Delta\delta} - \epsilon_{X\delta}} \widehat{b}_X. \quad (31)$$

Figure 5 shows the results. The agreement between the four estimators we have tried is remarkable. (Measurements of the full b_1 have greater signal-to-noise; estimating the scale independent part b_{10} requires the use of a different field, in our case X , and this may reduce the signal-to-noise of the measurement.) We conclude that we are now able to estimate b_{10} with no prior knowledge of δ_c .

We have made the point that our estimates of b_{10} and b_{11} were made without knowledge of $\delta_c(s)$. Therefore, it is interesting to simply add these estimates, as equation (18) indicates that they should sum to give $\delta_c(s)/s$. In terms of the two profiles, our estimator reads

$$\widehat{\delta}_c(s) = s \left(\frac{1 - \epsilon_{X\delta}}{\epsilon_{\Delta\delta} - \epsilon_{X\delta}} \widehat{b}_{\Delta} + \frac{1 - \epsilon_{\Delta\delta}}{\epsilon_{X\delta} - \epsilon_{\Delta\delta}} \widehat{b}_X \right). \quad (32)$$

Figure 6 shows that this procedure works quite well. In the next section we show that it also works when δ_c is stochastic, with a mean value which is scale dependent ($\delta_c(s)$ is a function of s). We conclude that we are able to infer the value of $\delta_c(s)$, and potentially its dependence on s , from measurements of the surrounding density field, without any a priori information about the physics of halo formation.

It is worth emphasizing how remarkable this is: to date, it has been thought that one must know something about the physics of halo formation – $\delta_c(s)$ – to correctly predict halo abundances and bias. We have shown that one can turn the argument around: One can obtain interesting constraints on this physics from measurements of halo bias.

We have also tested our methodology for estimating the second order bias parameters. I.e., we first set

$$\widehat{b}_{2Y} = \frac{\langle Y^2 \rangle}{\langle Y \delta_p \rangle^2} \frac{1}{N} \sum_{\alpha=1}^N \left[\frac{Y_{\alpha}^2}{\langle Y_{\alpha}^2 \rangle} - 1 \right] \quad (33)$$

for three choices of Y , and then use equation (26) to estimate

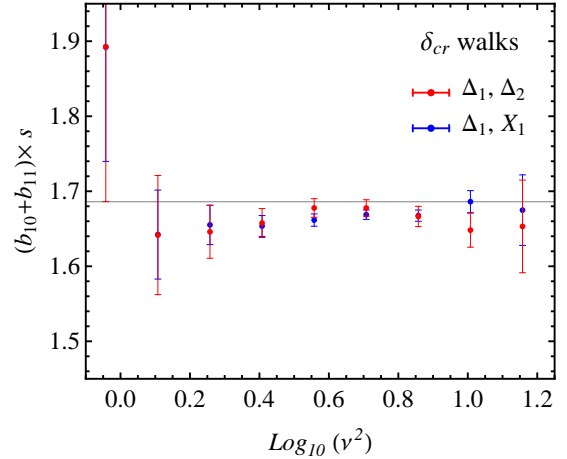


Figure 6. Estimate of δ_c from summing our previous estimates of b_{10} and b_{11} for the same choices of Y shown in the previous Figure (overdensity or its derivative measured on the same or different scales). Horizontal line shows the correct value; symbols show our various estimates.

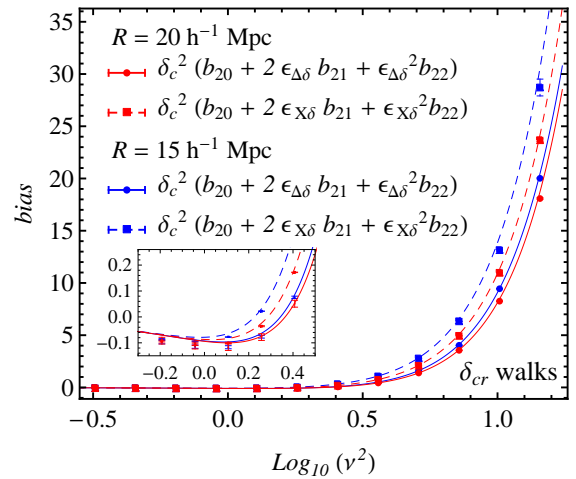


Figure 7. Estimate of b_2 (equation 33) for the choices of Y indicated (overdensity or its derivative measured on the same or different scales). Curves show the predicted (scale-dependent) value.

b_{20} , b_{21} and b_{22} . Figure 7 shows that our various estimators of b_2 depend on scale in the expected way: scale-independent coefficients are multiplied by scale-dependent factors $\epsilon_{Y\delta}$ – it is only these scale-dependent factors which depend on the choice of Y .

Figure 8 shows our estimates of the scale-independent coefficients, obtained from using the \widehat{b}_2 values shown in Figure 7 in equation (26). These show that we recover b_{20} very well, b_{21} less well, and b_{22} not well at all. Our experience with b_{11} means this is not completely unexpected – we are differencing similar numbers and then normalizing by a small number. (Notice that b_{21} is about $10\times$ smaller than b_{20} , and b_{22} is $10\times$ smaller still.) Presumably, multiplying by the determinant of the matrix in equation (26) would reduce some of this systematic, for the same reason that theory and measurement are in better agreement in Figure 4 than in Figure 3.

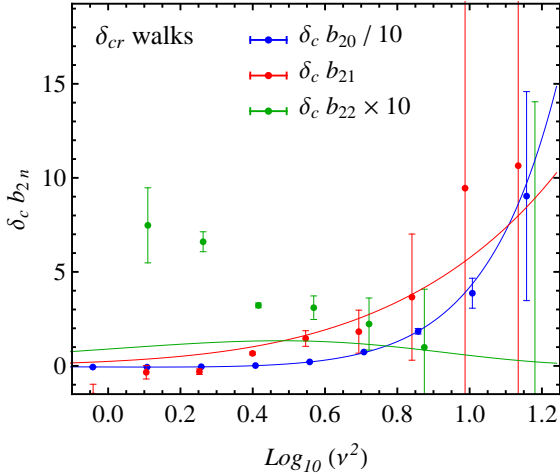


Figure 8. Estimates of the scale-independent bias coefficients b_{20} , b_{21} and b_{22} from inserting the various estimates of b_2 shown in the previous Figure in equation (26). Curves show the predicted values.

Despite the disagreement for b_{22} , it is worth pausing to appreciate what Figure 8 shows. The value of b_{20} is usually estimated from measurements on large scales. We have been able to estimate it remarkably well from measurements on much smaller scales – scales which are almost nonlinear. And b_{21} has never been measured before.

While the smallness of b_{21} and b_{22} has made estimating them difficult, it also means that we should be able to make a relatively clean test of the consistency relation, equation (27). The symbols in Figure 9 show the result of inserting the estimates shown in Figure 8 into the left hand side of equation (27) (and multiplying by s_p). The solid curve shows the predicted value – the right hand side of equation (27) (times s_p).

One might argue that we were only able to show the theory curve because we knew the value of δ_c to begin with. We could, of course, have assumed we did not know it, and then fit the measurements to $H_2(av)$, to see if the parameter a is constant or not. We leave such tests for future work. For the present purposes, we think it is sufficient that the agreement between theory and measurement in Figure 9 indicates that our methodology has indeed recovered the correct value of δ_c , and the fact that it is independent of scale.

As another consistency check, the dashed curve shows the result of using the estimate of δ_c from Figure 6 – that based on the linear bias factors b_{10} and b_{11} – in the right hand side of equation (27). The agreement is very good, again suggesting that our methodology works well.

Before moving on, we note again that these consistency checks are unprecedented. Never before has the nonlinear bias factor b_2 , measured on nonlinear scales, been used to estimate δ_c . Perhaps more remarkably, our estimate of δ_c is made from *linear* combinations of nonlinear bias factors measured on nonlinear scales.

4 STOCHASTICITY

The previous section showed that one gets the same k -dependent bias by cross-correlating the protohaloes with any

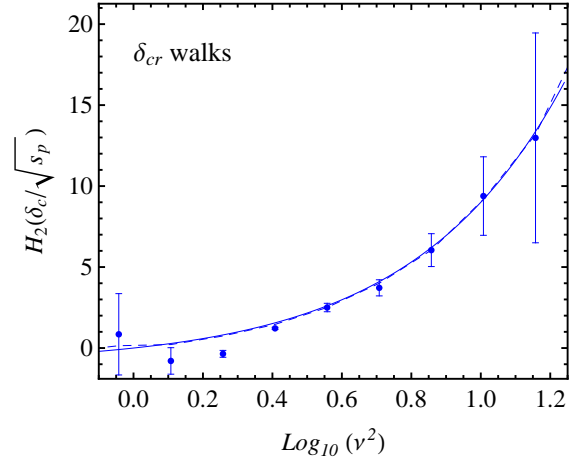


Figure 9. Test of consistency relation (equation 27). Symbols show the result of inserting the b_{20} , b_{21} and b_{22} estimates shown in the previous Figure into the left hand side of equation (27). Solid curve shows the predicted value; dashed curve shows the result of inserting the estimate of δ_c shown in Figure 6 into the right hand side of equation 27. The agreement shows that we recover the same δ_c values from our measurements of b_2 that we did from b_1 .

field Y that is correlated with one of the variables which determine the protohalo constraints \mathcal{C}_p , if bias is defined (as is conventional) with respect to $\langle Y\delta_p \rangle$. It also showed that this allows one to estimate the k -independent bias factor b_{10} (and, by extension, all the bias factors b_{nj} as well) with no prior assumptions about the value of $\delta_c(s)$.

One might have wondered if this only works if the barrier is a deterministic function of s . To show that it is more general, we now consider a model in which first crossing is determined by two variables δ and q , and their derivatives:

$$\begin{aligned} \delta(R_p) &\geq \delta_c(1 + \sqrt{s_p}q(R_p)/q_c) \\ \delta(R_p + \Delta R) &\leq \delta_c[1 + \sqrt{s_{p+\Delta}}q(R_p + \Delta R)/q_c]. \end{aligned} \quad (34)$$

Here, q may or may not have the same statistical properties as δ . E.g., even if both δ and q are Gaussian, they may have different correlation properties. The case in which δ is Gaussian but q is not can be related to models in which both the overdensity and the traceless shear determine halo formation (e.g. Sheth & Tormen 2002). In such models q is independent of δ , and one may think of the conditions above as defining a model in which the critical overdensity for collapse varies stochastically (Sheth et al. 2013; Musso & Sheth 2014).

This is a nice model to explore in the present context since the distribution $p(q^2)$ is chi-squared with 5 degrees of freedom,

$$q^2 \equiv \frac{1}{5} \sum_{i=1}^5 \frac{g_i^2}{s}, \quad (35)$$

where the g_i are Gaussian random variables with zero mean and variance s (we will sometimes use the notation q_5 instead of q). Therefore, unlike δ (which is Gaussian distributed), it has non-zero mean. Thus, if one wished to ignore the stochasticity arising from the distribution of q_5 by replacing $q_5 \rightarrow \langle q_5 \rangle$ in equation (34), then we would have a problem

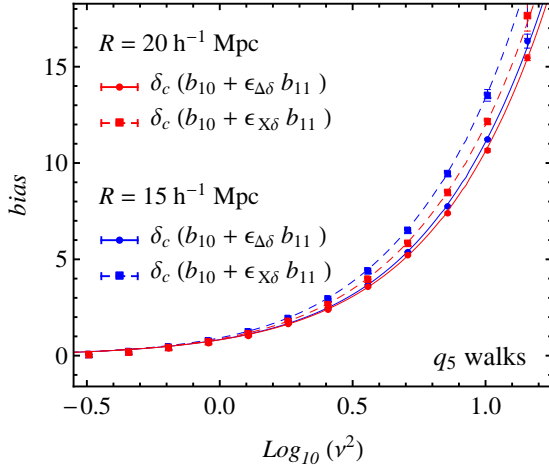


Figure 10. Same as Figure 2 but for walks crossing a stochastic boundary (equation 34) that has $q_c = 6.25$.

in which δ must cross a deterministic barrier whose height depends on s . The deterministic and constant barrier of the previous section corresponds to $q_c \rightarrow \infty$. Since our more general model is more complex in both respects, we now wish to explore the consequences of the fact that \mathcal{C}_p now depends on δ_p and v_p (as in the previous section) as well as q_p and its derivative.

4.1 Consistency relations for density bias

For the same reasons as before, we expect this model to have bias factors b_{10} and b_{11} which depend on δ_c and s . However, because the field q also matters, we expect the values of b_{10} and b_{11} to depend on q_c as well. In particular, we expect equation (18) to become

$$b_{10} + b_{11} = \frac{\langle \delta_{1\times} \rangle}{s} = \frac{\delta_c}{s} \left(1 + \sqrt{s} \frac{\langle q_p | \mathcal{C}_p \rangle}{q_c} \right) \quad (36)$$

where $\langle \delta_{1\times} \rangle \equiv \langle \delta_p | \mathcal{C}_p \rangle$ is the mean value of δ on the first crossing scale. The final equality shows that this mean value is related to the mean value of q at first crossing, which we might write in more suggestive notation as $\langle q_p | \mathcal{C}_p \rangle \equiv \langle q_{1\times} \rangle$. More generally, we expect equation (28) to become

$$\sum_{j=0}^n \binom{n}{j} b_{nj} = s^{-n/2} \langle H_n(\nu_{1\times}) \rangle. \quad (37)$$

(Castorina et al. 2016), where we have set $\nu_{1\times} \equiv \delta_{1\times} / \sqrt{s}$.

On the other hand, because q is independent of δ , we do not expect q to contribute scale (or k -) dependence to correlations with δ . Therefore all the technology of the previous section should go through unchanged: we should be able to estimate the scale independent b_{10} from the scale dependent b_{11} using any combination of cross-correlations of Δ and its derivatives around the first crossing scales (i.e. centered on the protohalo positions).

Figure 10 shows b_1 for this stochastic barrier model in which $q_c = 6.25$. The agreement between the theoretical curves and the measurements is as good as in Figure 2 (where the barrier was deterministic). Figure 11 shows estimates of b_{10} obtained analogously to the constant deterministic barrier case shown in Figure 5. The agreement between

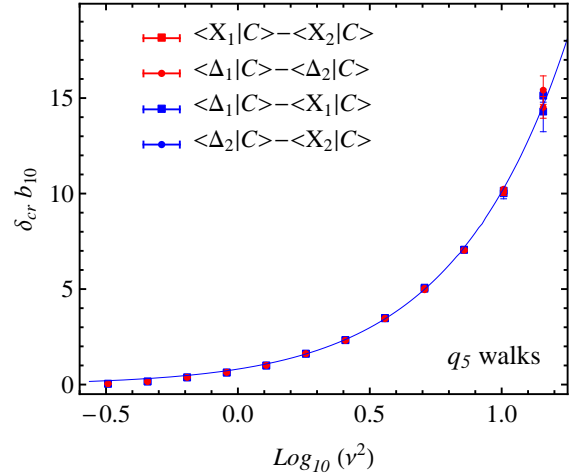


Figure 11. Same as Figure 5, but now for the case in which $q_c = 6.25$.

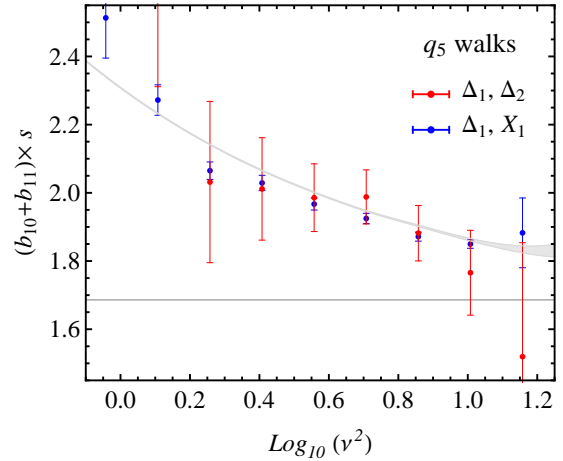


Figure 12. Same as Figure 6, but now for the case in which $q_c = 6.25$. In this model, the mean value of $\delta_c(s)$ increases with s because of stochasticity (grey band). Our methodology correctly reproduces this trend with no prior knowledge of either the stochasticity or the scale dependence.

our various estimators is again excellent. This is a nontrivial extension of the constant barrier one shown in Figure 5 since, in this case, not only is the barrier stochastic, but the fact that the mean of q is non-zero makes the effective value of $\delta_c(s)$ increase with s . Thus, we conclude that we are now able to estimate b_{10} with no prior knowledge of δ_c (constant or not? stochastic or not? etc.).

Following the steps laid out in the previous section, we can also estimate b_{11} . This works well, so we have not shown it. Rather, we have combined it with the estimate of b_{10} shown in Figure 11 to estimate $\delta_c(s)$. Figure 12 shows that this works very well: in particular, our methodology is able to recover the (stochasticity induced) mass dependence of δ_c quite faithfully.

Finally, the higher order bias coefficients b_n can also be measured in the same way as before, by averaging higher-order Hermite polynomials centered on the constrained regions. Figures 13–15 respectively show the scale dependent

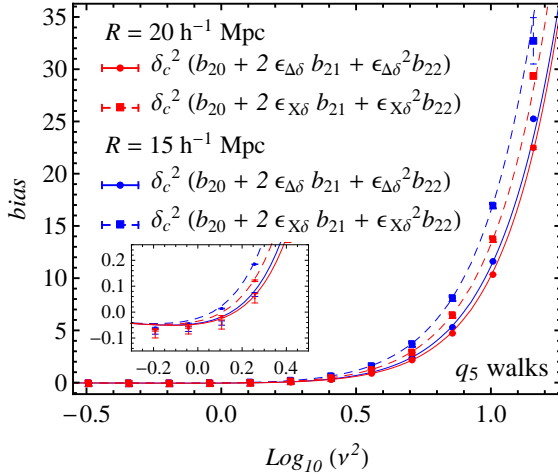


Figure 13. Same as Figure 7, but now for the case in which $q_c = 6.5$.

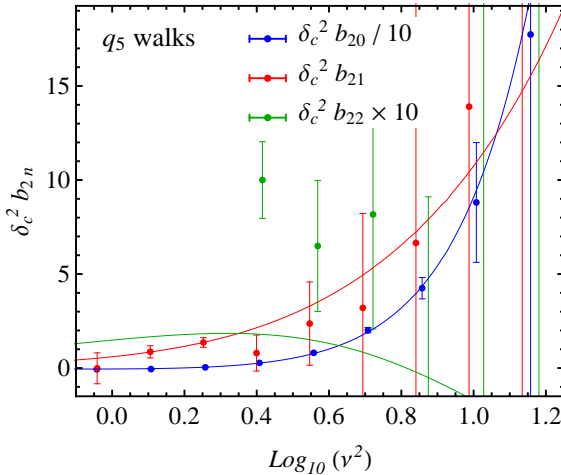


Figure 14. Same as Figure 8, but now for the case in which $q_c = 6.5$. I.e., estimates of the scale-independent bias coefficients b_{20} , b_{21} and b_{22} come from inserting the various estimates of b_2 shown in the previous Figure into equation (26). Curves show the predicted values. The estimates are noisier than those in Figure 8 because of the stochasticity coming from the variable q .

estimate of b_2 , the scale-independent bias coefficients b_{20} , b_{21} and b_{22} from inserting the various estimates of b_2 shown in the previous Figure into equation (26), and the consistency relation (equation 27). In all cases, the methodology works as well here as it did for the constant deterministic barrier of the previous section – although some of the estimates (e.g. b_{22}) are slightly noisier because of the stochasticity.

4.2 Signature of stochasticity

In stochastic barrier models, the value of δ at first crossing on scale s is not deterministic, but has a distribution $p(\delta_{1\times}|s)$. Equation (37) with $n = 1$ estimates the mean of this distribution, $\langle\delta_{1\times}|s\rangle$. Figure 12 shows that we were able to correctly reconstruct the fact that it depends on s , even if this was not known a priori. When $n = 2$, then equation (37)

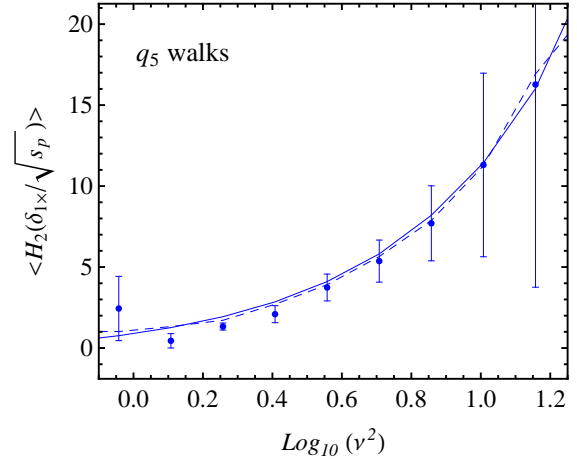


Figure 15. Test of consistency relation (equation 27): Symbols show the result of inserting the b_{20} , b_{21} and b_{22} estimates shown in the previous Figure into the right hand side of equation (27). Solid curve shows the predicted value. Dashed curve shows $H_2(\langle\delta_{1\times}|s\rangle/\sqrt{s_p})$, with $\langle\delta_{1\times}|s\rangle$ given by Figure 12.

can be combined with $\langle\delta_{1\times}|s\rangle$ to yield an expression for the variance around the mean in terms of the first and second order bias factors. We will write the variance in two ways.

First, we note that

$$\text{Var}(\delta_{1\times})/s = \langle H_2(\nu_{1\times}) \rangle - H_2(\langle\nu_{1\times}\rangle). \quad (38)$$

Since non-zero variance is a signature of stochasticity, the expression above shows that the difference between $\langle H_2(\nu_{1\times}) \rangle$ and $H_2(\langle\nu_{1\times}\rangle)$ is a direct measure of stochasticity. The solid and dashed curves in Figure 15 show that these two quantities are actually rather similar – evidently, the stochasticity in this particular model is small, for reasons we quantify in the next subsection.

Using the consistency relations to write the Hermite polynomials above in terms of the b_{ij} leads to our second suggestive expression for the variance:

$$\frac{\text{Var}(\delta_{1\times}|s)}{s} = 1 + s (b_{20} - b_{10}^2 + 2b_{20} - 2b_{10}b_{11} + b_{22} - b_{11}^2). \quad (39)$$

This shows explicitly how measurements of the first and second order bias factors yield information about the rms stochasticity. Of course, this argument can be generalized to the statement that measurements of b_n constrain the n th order moments of $p(\delta_{1\times}|s)$. We exploit this fact elsewhere.

4.3 Two types of consistency relations for shear bias

We noted above that, although the stochasticity provided by q does affect the amplitude of scale dependent bias (by modifying the values of b_{10} and b_{11} ; one can see this by comparing the predicted bias at the highest ν shown in Figures 5 and 11), it does not contribute any new scale dependence. However, we do expect to get entirely new bias factors, c_n , say, which arise because \mathcal{C}_p depends on q_p and its derivative (c.f. equation 34). Such terms are usually referred to as ‘non-local’ bias coefficients (Chan et al. 2012; Sheth et al. 2013; Baldauf et al. 2012). We expect these to be scale dependent and to satisfy their own consistency relations in terms of

scale independent coefficients c_{nj} (Castorina & Sheth 2013; Castorina et al. 2016). Since q is fundamentally quadratic with respect to Gaussian variables (equation 35), we expect these coefficients to only occur for even n . E.g., when $n = 2$, we expect to be able to estimate c_{20} , c_{21} and c_{22} from cross-correlations with the shear field, without prior knowledge of q_c and/or stochasticity; and we expect cross-correlations of all quantities which correlate with q to furnish estimates of c_2 . We also expect all this to generalize to higher order nonlocal bias coefficients c_n .

We consider the issue of consistency relations first. The orthogonal polynomials associated with χ^2 -distributed variables are modified Laguerre polynomials (e.g. $L_1^{(\alpha)}(x) = 1 + \alpha - x$). In the Appendix, we show that consistency relations between the shear bias factors c_{nj} are similar to those for the density provided one works with these orthogonal polynomials. For example, in close analogy with equation (37) for $n = 2$,

$$c_{20} + 2c_{21} + c_{22} = - \left\langle L_1^{(3/2)}(5q_{1\times}^2/2) \right\rangle \quad (40)$$

(also see equations A15-A14). Similarly, the coefficients c_{nj} will sum to $L_k^{(3/2)}(5q_{1\times}^2/2s)$, where $n = 2k$.

The similarity in spirit between density and shear bias factors goes deeper. In the previous section, we made the point that the density bias factors b_{nj} can be rearranged to describe the moments of $p(\delta_{1\times}|s)$ (e.g. equation 38). For similar reasons, the shear bias factors c_{nj} can be rearranged to describe the moments of $p(q_{1\times}|s)$. While it is natural to expect this deeper connection, there is another type of consistency which is, perhaps, more surprising. These arise because

$$\langle \delta_{1\times}^n \rangle = \langle \delta_c^n (1 + \sqrt{s} q_{1\times}/q_c)^n \rangle, \quad (41)$$

so moments of $\delta_{1\times}$ can be written as linear combinations of the moments of $q_{1\times}$ and vice versa. Hence, for example,

$$\begin{aligned} \langle H_2(\nu_{1\times}) \rangle &= (\delta_c^2/s) \langle 1 + s(q_{1\times}/q_c)^2 + 2\sqrt{s} q_{1\times}/q_c \rangle - 1 \\ &= H_2(\langle \nu_{1\times} \rangle) + (\delta_c/q_c)^2 \text{Var}(q_{1\times}), \end{aligned} \quad (42)$$

where $\text{Var}(q_{1\times}) = \langle q_{1\times}^2 \rangle - \langle q_{1\times} \rangle^2$. Comparison with equation (38) shows that

$$\frac{\text{Var}(\nu_{1\times})}{\delta_c^2} = \frac{\text{Var}(q_{1\times})}{q_c^2}. \quad (43)$$

Note that the difference between $\langle H_2(\nu_{1\times}) \rangle$ and $H_2(\langle \nu_{1\times} \rangle)$ equals $(\delta_c/q_c)^2$ times the variance of $q_{1\times}$. In our model, both these factors are substantially smaller than 1, which is why the solid and dashed curves in Figure 15 are so similar.

In addition, from equation (41) and the first line in equation (42) we have

$$\begin{aligned} \left\langle L_1^{(3/2)}(5q_{1\times}^2/2s) \right\rangle &= \frac{5q_c^2}{2\delta_c^2} \left[1 + \langle H_2(\nu_{1\times}) \rangle - 2\nu \langle H_1(\nu_{1\times}) \rangle \right. \\ &\quad \left. + \nu^2 - (\delta_c/q_c)^2 \right]. \end{aligned} \quad (44)$$

This is a remarkable expression; it shows that the density bias factors b_{nj} and those for the shear c_{nj} are related, even though the density δ and the shear q are independent. As a result of these relations, one can get information about the nonlocal bias factors c_{nj} simply by taking appropriate combinations of cross correlations with the density rather than

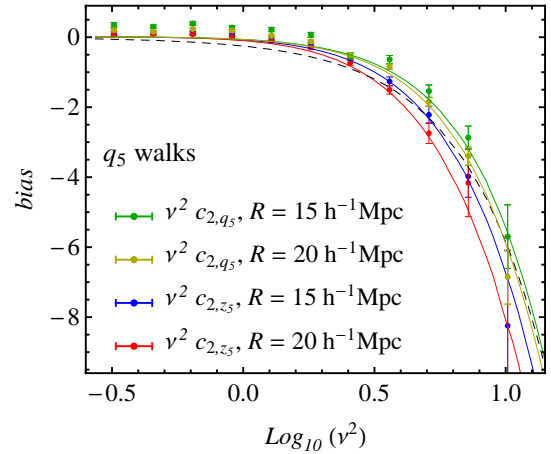


Figure 16. Same as Figure 2, but now for the case in which $q_c = 6.25$ and we use cross-correlations (equation 45) to estimate c_2 rather than b_1 . Two sets of symbols are for cross-correlating with the shear on two scales, and another two are for the shape (equation 47) on two scales. All the measurements are indistinguishable, and are in excellent agreement with the solid curve which shows equation (46).

the shear field. Since the former are much easier to measure, this represents a substantial simplification. Perhaps more importantly, the expression above shows that measurements of b_{nj} constrain the parameter space of allowed c_{nj} values.

4.4 Estimators and tests of consistency relations

We turn now to direct estimates of the shear bias factors. By analogy with the deterministic barrier case, we set

$$\widehat{c}_{2,q_5}(s) = -\frac{1}{r^2} \frac{1}{N} \sum_{\alpha=1}^N L_1^{(3/2)}(5Q_{5,\alpha}^2/2) \quad (45)$$

where the sum runs over all the N walks that cross in a given narrow bin of width Δs centered on s , Q_5 denotes the value of the shear field on some smoothing scale that is larger than the one associated with first crossing and $r \equiv S_{0\times}/\sqrt{sS}^1$. Also by analogy with δ , we expect $\widehat{c}_{2,q_5}(s)$ to depend on a piece, c_{20} , which dominates on large scales, and two others, c_{21} and c_{22} , which yield scale dependence.

Figure 16 shows results for two different smoothing scales; for now, consider only the top two sets of (green and yellow) symbols. In both cases, we show $\nu^2 \widehat{c}_{2,q_5}(s)$, for ease of comparison with Sheth et al. (2013): $\delta_c^2 c_2^{\text{SCS}} = 2\nu^2 c_{20}$. First, notice that the signal is always negative. As discussed by Sheth et al. (2013), this is a result of the model assumption that the shear, q^2 , inhibits the formation of objects (because it adds to δ_c in equation 34). Second, it is hard to distinguish between the green and yellow symbols: evidently, c_{2,q_5} depends less strongly on scale than does b_1 . In the $R \rightarrow \infty$ limit, we expect the dominant contribution to

¹ We defined $S_{j\times} \equiv (2\pi)^{-3} \int d^3k k^{2j} P_{\delta\delta} W_{R_p} W_{R_q}$ and $S_j \equiv (2\pi)^{-3} \int d^3k k^{2j} P_{\delta\delta} W_{R_q}^2$.

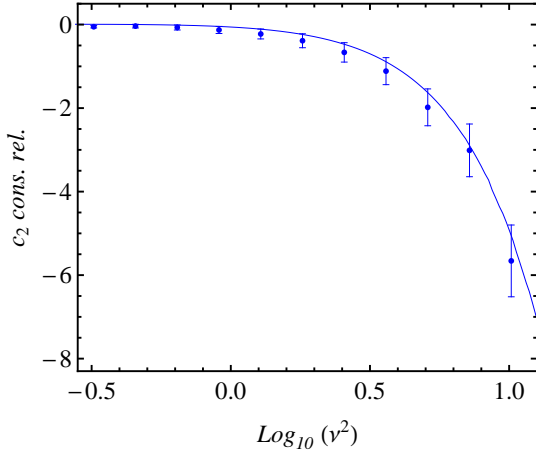


Figure 17. Comparison between the analytic prediction of non-local bias parameters c_2 (smooth curve) and the combination of moments of $\delta_{1\times}$ which appears on the rhs of equation (44) (symbols).

be (see Appendix A and B for details)

$$c_{20} = - \int dq p_5(q) \int d\dot{q} p(\dot{q}) \frac{f_{q\dot{q}}(s, q, \dot{q})}{f(s)} \times \left(L_1^{(3/2)}(5q^2/2) + 5\Gamma^2 \sqrt{s} q \dot{q} + \Gamma^2 L_1^{(-1/2)}(5\Gamma^2 s \dot{q}^2/2) \right), \quad (46)$$

where $\dot{q} \equiv dq/d\sqrt{s}$, $f_{q\dot{q}}(s, q, \dot{q})$ is given by equation (B2), and $f(s)$ is the first crossing distribution obtained by integrating $f_{q\dot{q}}(s, q, \dot{q})$ over q and \dot{q} (equation B1).

Notice that, since Q_5 correlates with \dot{q} , large scale non-local bias also receives contributions from \dot{q} and \dot{q}^2 . In Figure 16, the full analytic c_2 at two different scales (equation A12) is shown as a continuous line, and we see that it provides a good description of our Monte Carlo estimates of equation (45), shown as the green and yellow points. Although the measurements at the two scales are too close to each other to allow us an estimate of the scale dependent piece, they are still on sufficiently small smoothing scales that we can appreciate the difference between the points and c_{20} (shown as the dashed line).

We have already checked that moments of the PDF of the density at first crossing are well reproduced by the consistency relation for the b_n 's (Figures 12 and 15). To close the loop, Figure 17 tests the consistency relation for shear bias in the form of equation (44). The agreement between the analytical prediction from the c_2 terms (smooth curve) and the particular combination of moments of $\delta_{1\times}$ given on the rhs of equation (44) (symbols) is impressive; it supports our understanding of why density and shear bias coefficients are related, and how they can be used to study halo formation. Presumably, good approximations to the higher order moments of $p(q_{1\times}|s)$ can also be written entirely in terms of the density bias factors b_n , but we have not pursued this further.

4.5 Relation to previous work

We now turn to the other two sets of symbols, the blue and red points, in Figure 16. These show the result of replacing

$$Q^2 \rightarrow \zeta_5^2 \quad \text{and} \quad r \rightarrow r_\zeta = \frac{S_{1\times}}{\sqrt{s} S_2} \quad (47)$$

in equation (45), with ζ_5^2 defined as follows. Since Q^2 is χ^2 with five degrees of freedom, we can think of it as $\sum_{i=1}^5 G_i^2/S$ where each of the G_i are independent Gaussian random numbers. We then set $Y^2 \equiv \sum_{i=1}^5 Y_i^2$, where $Y_i \equiv dG_i/dS$, and finally define $\zeta_5^2 \equiv Y^2/\langle Y^2 \rangle$. In this sense, ζ^2 is to Q^2 as $d\delta/ds$ was to δ in the previous section. For this reason, we might expect the result of averaging $L_1^{(3/2)}(5\zeta_5^2/2)$ around first crossing positions to be similar to, or (in the limit of infinite smoothing scale) the same as averaging $L_1^{(3/2)}(5Q_5^2/2)$, provided we use the correct cross-correlation coefficient, r_ζ , of equation (47). The agreement between \hat{c}_{2,q_5} and \hat{c}_{2,ζ_5} , blue and red points, in Figure 16 shows that this is indeed the case. As before, we are actually able to understand the small differences between the two set of measurements, since they arise by replacing $\epsilon_{\Delta\delta}$ in equation A12 with $\epsilon_{\chi\delta}$.

This agreement is particularly relevant to recent work on the role of shape in determining protohalo formation. Biagetti et al. (2014) measured cross-correlations between protohaloes and a quantity which is like our ζ_5^2 . Although they differ in detail – their quantity is the same as ours only for a Gaussian smoothing filter – the essential point is that the quantity they consider is correlated with q_5^2 . They define

$$s_2 \chi_{01} \equiv - \frac{1}{\bar{r}^2} \frac{1}{N} \sum_{\alpha=1}^N L_1^{(3/2)}(5\zeta_{5,\alpha}^2/2), \quad (48)$$

which they expected would return the correlation between the small and large scale values of ζ , for which the cross correlation coefficient is $\bar{r}^2 \equiv S_{2\times}^2/s_2 S_2$. Their prediction – which is shown as a blue dashed curve in their Figure 5² – increases weakly as protohalo mass increases. In contrast, their measurements – shown as triangles in their Figure – decrease strongly. This discrepancy led them to question their assumption about the role halo shape plays in determining where haloes form.

However, shape and shear are closely related; crudely speaking, shear is to shape as δ is to $d\delta/ds$. Therefore, if shear plays a role in halo formation, then cross-correlating the large scale shape field with protohalo positions may actually return the bias factor associated with shear c_{20} rather than shape. The agreement between \hat{c}_{2,q_5} and \hat{c}_{2,ζ_5} in Figure 16 shows that this is indeed what happens in our idealized toy model. Since our ζ_5 is very close to ζ_5 in Biagetti et al. (2014), their

$$s_2 \chi_{01} \approx (r_\zeta/\bar{r})^2 c_{20} = (S_{1\times}/S_{2\times})^2 (s_2/s) c_{20}. \quad (49)$$

The difference in correlation coefficient factors r is a consequence of their assuming that shape is fundamental, whereas our model has shear as the fundamental quantity. The prefactor on the right hand side decreases as ν increases, so it

² Beware, the notation χ_{10} in Biagetti et al. (2014) is for the bias parameter obtained from cross-correlating with a quantity which does not correlate with q_5 , so it has nothing to do with c_{20} !

flattens the trend we see in Figure 16, making it more like that in Figure 5 of Biagetti et al. (2014). And, vice-versa, we have checked that this factor transforms their dashed curve into our solid one here. This strongly suggests that the procedure used by Biagetti et al. (2014) actually returns the bias associated with shear rather than shape. (This argument does not explain why their measured signal crosses zero at small masses. In our walks, this happens if our sample size is not sufficiently large, but it is not clear if a similar effect is to blame for them as well. For instance, it may be that the model we are using to illustrate the effect of shear on halo formation is too simplistic. But building a more sophisticated model is beyond the scope of the present work.)

5 CONCLUSIONS

We argued that it was useful to generalize the notion of bias with respect to the *overdensity* field of the dark matter, to the difference between averages around special positions in space and those over all space. This makes it clear that averages centered on protohalo patches – i.e. cross-correlations between biased tracers and the dark matter – pick up correlations with any quantity which matters for halo formation. Since bias is typically measured using averages on scales larger than those of a halo, the cross-correlation signal is typically dominated by the terms with the fewest orders of k . As a result, cross-correlating protohaloes with the large scale density and curvature fields should return the same answer. We showed this was true using the excursion set approach, in which, for many cross-correlations, the answer is known analytically (Figure 2). We then used this fact to make two points.

First, we showed that this allows one to estimate bias factors with no prior knowledge of the physics of collapse (Figure 5). In effect, this enables the cross-correlation methods pioneered by Musso et al. (2012) to be performed with fewer assumptions than before. We illustrated this explicitly for linear and quadratic density bias factors (Figures 2 and 8).

We then showed that this same effect matters for other variables which are not correlated with overdensity. One such pair of current interest is the quadrupolar pair associated with the tidal shear and shape. We argued that cross-correlating protohaloes with the large scale shape field will produce a signal which is very similar to that obtained by cross-correlating with the shear field (Figure 16). This may explain some puzzling results in previous work (Biagetti et al. 2014); in particular, the puzzling cross-correlations there may actually be consistent with the nonlocal bias signals reported in Chan et al. (2012) and Sheth et al. (2013).

To date, it has been thought that one must know something about the physics of halo formation – such as the critical density $\delta_c(s)$ a protohalo patch must have if it is to become a halo of mass $m(s)$ – to correctly predict halo abundances and bias. We have shown that one can turn the argument around: One can obtain interesting constraints on this physics from measurements of halo bias. Our methodology correctly reproduces $\delta_c(s)$ with no prior knowledge of either the stochasticity associated with halo formation, its dependence on halo mass, or the mass and scale dependence of halo bias (Figures 6 and 12). The same methodology can

be used to quantify the stochasticity in halo formation (Figures 9 and 15), and to relate the nonlocal bias factors associated with this stochasticity to nonlinear density bias (equation 44). This has the potential to vastly simplify analyses in the next generation of datasets which will be sensitive to nonlocal bias.

In a companion paper (Castorina et al. 2016), we show how to extend some of the results discussed in this paper for the stochastic barrier which depends on the shear to the more realistic case where haloes are also peaks of the density field (Bardeen et al. 1986; Paranjape & Sheth 2012; Paranjape et al. 2013). A comparison of the predictions of this new model to measurements of the halo mass function and the density and shear bias parameters in N -body simulations is also the subject of work in progress.

Finally, although our results relate to the Lagrangian bias of protohalo patches, in the peaks model, the lowest order mapping to the Eulerian bias of the evolved halo field preserves the consistency relations we have exploited (Desjacques et al. 2010). Work in progress studies if this is generic.

ACKNOWLEDGEMENTS

EC, AP and RKS thank E. Sefusatti and L. Guzzo for their kind hospitality at INAF-OAB Merate during the summer of 2014. The research of AP is supported by the Associateship Scheme of ICTP, Trieste and the Ramanujan Fellowship awarded by the Department of Science and Technology, Government of India.

REFERENCES

- Baldauf T., Seljak U., Desjacques V., McDonald P., 2012, *Phys. Rev. D*, **86**, 083540
- Bardeen J. M., Bond J. R., Kaiser N., Szalay A. S., 1986, *ApJ*, **304**, 15
- Biagetti M., Chan K. C., Desjacques V., Paranjape A., 2014, *MNRAS*, **441**, 1457
- Bond J. R., Cole S., Efstathiou G., Kaiser N., 1991, *ApJ*, **379**, 440
- Castorina E., Sheth R. K., 2013, *MNRAS*, **433**, 1529
- Castorina E., Paranjape A., Hahn O., Sheth R. K., 2016, Arxiv eprints
- Chan K. C., Scoccimarro R., Sheth R. K., 2012, *Phys. Rev. D*, **85**, 083509
- Chan K. C., Sheth R. K., Scoccimarro R., 2015, preprint, ([arXiv:1511.01909](https://arxiv.org/abs/1511.01909))
- Cooray A., Sheth R., 2002, *Phys. Rep.*, **372**, 1
- Desjacques V., Crocce M., Scoccimarro R., Sheth R. K., 2010, *Phys. Rev. D*, **82**, 103529
- Frusciante N., Sheth R. K., 2012, *J. Cosmology Astropart. Phys.*, **11**, 016
- Musso M., Sheth R. K., 2012, *MNRAS*, **423**, L102
- Musso M., Sheth R. K., 2014, *MNRAS*, **442**, 401
- Musso M., Paranjape A., Sheth R. K., 2012, *MNRAS*, **427**, 3145
- Paranjape A., Sheth R. K., 2012, *MNRAS*, **426**, 2789
- Paranjape A., Sheth R. K., Desjacques V., 2013, *MNRAS*, **431**, 1503
- Sheth R. K., Tormen G., 2002, *MNRAS*, **329**, 61
- Sheth R. K., Chan K. C., Scoccimarro R., 2013, *Phys. Rev. D*, **87**, 083002
- Szalay A. S., 1988, *ApJ*, **333**, 21

APPENDIX A: CROSS-CORRELATING WITH LAGUERRE POLYNOMIALS IN THE STOCHASTIC MODEL

The main text described the motivation for computing cross correlations between special positions (protohaloes) and the large scale shear field: since the shear field is χ_5^2 -distributed, the generalized Laguerre polynomials $L_n^{(3/2)}(x)$ are particularly interesting. Here we show what this cross-correlation is expected to yield in the context of the stochastic model for halo formation described in the main text.

Our procedure will be to show the result of averaging the generating function of generalized Laguerre polynomials,

$$\mathcal{L}(t|1 + \alpha, x) \equiv \sum_m^\infty t^m L_m^{(\alpha)}(x) = \frac{e^{-tx/(1-t)}}{(1-t)^{1+\alpha}}, \quad (\text{A1})$$

over the special positions, from which the individual averages can be obtained in the usual way (by taking derivatives with respect to t). This follows [Castorina & Sheth \(2013\)](#) who studied the case when the stochastic variable is Gaussian rather than χ^2 . Before proceeding to the derivation we need to define a few quantities. From equation (35) we define the derivative of q with respect to \sqrt{s}

$$\begin{aligned} \frac{dq}{d\sqrt{s}} &\equiv \dot{q} = \frac{1}{5q} \sum_i \frac{g_i}{\sqrt{s}} \left(\frac{\dot{g}_i}{\sqrt{s}} - \frac{g_i}{s} \right) \\ &= \frac{\Gamma\sqrt{s}}{5q} \sum_i G_i X_i \end{aligned} \quad (\text{A2})$$

which is Gaussian distributed with zero mean and variance $5/(\Gamma^2 s)$ with $\Gamma^2 \equiv \gamma^2/(1-\gamma^2)$ ([Musso & Sheth 2014](#)). For ease of notation we have defined the normalized variables $G_i = g_i/\sqrt{s}$ and $X_i = \dot{G}_i \Gamma \sqrt{s}$, which have the convenient property of being independent from each other, i.e. $\langle G_i X_j \rangle = 0$ for all i, j . To obtain an expression for the bias coefficient with respect to traceless shear we have to compute the average Laguerre polynomial in the large scale shear Q^2 given some conditions on the small scale shear and its derivative, $\mathcal{C} = \mathcal{C}(q, \dot{q})$,

$$\begin{aligned} &\langle L_1^{(3/2)}(5Q^2/2) | \mathcal{C} \rangle = \\ &\prod_i \int p(G_i) dG_i \prod_i \int p(X_i) dX_i \prod_i \int p(G_{0,i} | G_i, X_i) dG_{0,i} \\ &\times \int dQ^2 \delta_D \left(Q^2 - \sum_i G_{0,i}^2/5 \right) L_1^{(3/2)}(5Q^2/2) \mathcal{C}(q, \dot{q}) \end{aligned} \quad (\text{A3})$$

where $Q^2 = \sum G_{0,i}^2/5$ is the large scale shear. Throughout this section, we will assume that the constraint is suitably normalised, such that $\int dq^2 p(q^2) \int d\dot{q} p(\dot{q}) \mathcal{C}(q, \dot{q}) = 1$. The more general excursion set constraint is discussed in Appendix B.

The conditional probability of the large scale $G_{0,i}$ given the small scale G_i and X_i is a Gaussian with mean $r_G G_i + r_X X_i$ and variance $1 - r_G^2 - r_X^2$, the cross-correlation coefficients being

$$r_G \equiv \langle G_0 G \rangle = S_\times / \sqrt{sS} \quad (\text{A4})$$

and

$$r_X \equiv \langle G_0 X \rangle = \frac{S_\times \Gamma}{\sqrt{sS}} (\epsilon_{\Delta\delta} - 1) = r_G \Gamma (\epsilon_{\Delta\delta} - 1), \quad (\text{A5})$$

where $\epsilon_{\Delta\delta}$ was defined in equation (17). The generating function in equation (A1) allows us to rewrite equation (A3) as

$$\begin{aligned} &\langle L_1^{(3/2)}(5Q^2/2) | \mathcal{C} \rangle \\ &= \prod_i \int p(G_i) dG_i \prod_i \int p(X_i) dX_i \prod_i \int p(G_{0,i} | G_i, X_i) dG_{0,i} \\ &\times \left[\frac{\partial}{\partial t} \mathcal{L} \left(t | 5/2, \sum_i G_{0,i}^2/2 \right) \Big|_{t=0} \right] \mathcal{C}(q, \dot{q}), \end{aligned} \quad (\text{A6})$$

which we then integrate over the large scale $G_{0,i}$ before taking the derivative. The net result is

$$\begin{aligned} &\langle L_1^{(3/2)}(5Q^2/2) | \mathcal{C} \rangle \\ &= \prod_i \int p(G_i) dG_i \prod_i \int p(X_i) dX_i \\ &\times \left[r_G^2 L_1^{(3/2)} \left(\sum_i G_i^2/2 \right) + r_X^2 L_1^{(3/2)} \left(\sum_i X_i^2/2 \right) \right. \\ &\quad \left. - r_G r_X \sum_i G_i X_i \right] \mathcal{C}(q, \dot{q}) \end{aligned} \quad (\text{A7})$$

which explicitly shows that the integral is zero if $\mathcal{C}(q, \dot{q}) = 1$. The first and last terms in square brackets in equation (A7) can be straightforwardly written in terms of q and \dot{q} . The second term instead looks more complicated.

To proceed, it is useful to note the following identity for χ^2 variables. Namely, if $\eta^2 \equiv \sum_{i=1}^n (\eta_i^2/s)/n$ and $\xi^2 \equiv \sum_{i=1}^n (\xi_i^2/s)/n$ are independent χ_n^2 variates, then the quantity $\sum_i (\eta_i \xi_i/s)/n$, being the dot product of the two underlying Gaussian vectors, can be written as $\eta \xi \cos \theta$. The joint distribution of g, ξ and θ is given by the product of three independent distributions:

$$\begin{aligned} &p(\eta) p(\xi) p(\theta) d\eta d\xi d\theta \\ &= 2 \frac{d\eta}{\eta} \frac{(n\eta^2/2)^{n/2}}{\Gamma(n/2)} e^{-n\eta^2/2} 2 \frac{d\xi}{\xi} \frac{(n\xi^2/2)^{n/2}}{\Gamma(n/2)} e^{-n\xi^2/2} \\ &\times d\theta \frac{\sin^{n-2} \theta}{B[1/2, (n-1)/2]} \end{aligned} \quad (\text{A8})$$

where $B(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$ is the Beta-function. It is useful to think of ξ as a radial variable made from the Cartesian variables $\xi_x = \xi \cos \theta$ and $\xi_y = \xi \sin \theta$, so that $d\xi_x d\xi_y = d\theta d\xi \xi$. In these variables,

$$\begin{aligned} &d\xi_x p(\xi_x) d\xi_y p(\xi_y) = p(\xi) p(\theta) d\xi d\theta \\ &= d\xi_x \frac{e^{-n\xi_x^2/2}}{\sqrt{2\pi/n}} 2 \frac{d\xi_y}{\xi_y} \frac{(n\xi_y^2/2)^{(n-1)/2}}{\Gamma[(n-1)/2]} e^{-n\xi_y^2/2}. \end{aligned} \quad (\text{A9})$$

This shows that ξ_x is a Gaussian with variance $1/n$, whereas ξ_y is χ_{n-1}^2 .

By matching notation, it is easy to see that q here is η , and since $q\dot{q}$ is like a dot product, \dot{q} is like $\xi_x/(\Gamma\sqrt{s})$. [Musso & Sheth \(2014\)](#) noted that $p(\dot{q}|q)$ is a Gaussian, independent of q ; our transformation from (ξ, θ) to (ξ_x, ξ_y) shows why. Since the constraint $\mathcal{C}(q, \dot{q})$ depends on q and \dot{q} but not on ξ_y , and the distribution of ξ_y is independent of q and \dot{q} , we may integrate it out. We therefore arrive at the final expression for the average Laguerre given the constraint,

$$\begin{aligned} &\langle L_1^{(3/2)}(5Q^2/2) | \mathcal{C} \rangle \\ &= r_G^2 \int dq^2 p(q^2) \int d\dot{q} p(\dot{q}) \mathcal{C}(q, \dot{q}) \tilde{c}_2(q, \dot{q}; s, S), \end{aligned} \quad (\text{A10})$$

where we defined

$$\tilde{c}_2(q, \dot{q}; s, S) \equiv L_1^{(3/2)}(5q^2/2) - 5\Gamma^2 \sqrt{s}(\epsilon_{\Delta\delta} - 1)q\dot{q} + \Gamma^2(\epsilon_{\Delta\delta} - 1)^2 L_1^{(-1/2)}(5\Gamma^2 s\dot{q}^2/2), \quad (\text{A11})$$

with the notation reminding us that this quantity depends on both the large and small scales. It is convenient to write the expression for non local bias collecting terms by powers of the derivative of q , i.e. by powers of $\epsilon_{\Delta\delta}$. We thus write, similarly to b_2 in equation (24),

$$c_2 = -r_G^{-2} \left\langle L_1^{(3/2)}(5Q^2/2) \middle| \mathcal{C} \right\rangle = c_{20} + 2\epsilon_{\Delta\delta} c_{21} + \epsilon_{\Delta\delta}^2 c_{22} \quad (\text{A12})$$

where the coefficients c_{2j} can be read off from equations (A10) and (A11):

$$c_{22} = -\left\langle \Gamma^2 L_1^{(-1/2)}(5\Gamma^2 s\dot{q}^2/2) \middle| \mathcal{C} \right\rangle \quad (\text{A13})$$

$$c_{21} = \left\langle 5\Gamma^2 \sqrt{s} q\dot{q}/2 \middle| \mathcal{C} \right\rangle - c_{22} \quad (\text{A14})$$

$$c_{20} = -\left\langle L_1^{(3/2)}(5q^2/2) \middle| \mathcal{C} \right\rangle - 2c_{21} - c_{22}. \quad (\text{A15})$$

This shows that the consistency relation for shear, equation (40) in the main text, holds.

Although we have worked at lowest order in shear bias so far, the results above can be generalized to arbitrary order by working directly with the Laguerre generating function. For completeness, we sketch this calculation next. The basic idea is to perform the constrained average of the generating function before taking the derivative in equation (A6). This leads to

$$\begin{aligned} & \left\langle \mathcal{L}(t|5/2, 5Q^2/2) \middle| \mathcal{C} \right\rangle \\ &= \prod_i \int p(G_i) dG_i \prod_i \int p(X_i) dX_i \prod_i \int p(G_{0,i}|G_i, X_i) dG_{0,i} \\ & \times \int dQ^2 \delta_D \left(Q^2 - \sum_i G_{0,i}^2/5 \right) \mathcal{L}(t|5/2, 5Q^2/2) \mathcal{C}(q, \dot{q}) \\ &= \prod_i \int p(G_i) dG_i \prod_i \int p(X_i) dX_i \prod_i \int p(G_{0,i}|G_i, X_i) dG_{0,i} \\ & \times \mathcal{L} \left(t \middle| 5/2, \sum_i G_{0,i}^2/2 \right) \mathcal{C}(q, \dot{q}). \end{aligned} \quad (\text{A16})$$

The integral over the $G_{0,i}$ is a straightforward convolution of Gaussians, leading to

$$\begin{aligned} & \prod_i \int p(G_{0,i}|G_i, X_i) dG_{0,i} \mathcal{L} \left(t \middle| 5/2, \sum_i G_{0,i}^2/2 \right) \\ &= \prod_i \frac{\exp(-t[r_G G_i + r_X X_i]^2/2t\Sigma^2)}{(t\Sigma^2)^{1/2}} \end{aligned} \quad (\text{A17})$$

where

$$t\Sigma^2 = 1 - t(r_G^2 + r_X^2). \quad (\text{A18})$$

The evaluation of $\left\langle \mathcal{L}(t|5/2, 5Q^2/2) \middle| \mathcal{C} \right\rangle$ is therefore similar to that of $\left\langle L_1^{(3/2)}(5Q^2/2) \middle| \mathcal{C} \right\rangle$ in equation (A7), with the expression in square brackets in that equation replaced by the r.h.s. of equation (A17). The discussion below equa-

tion (A7) carries through, with the result

$$\begin{aligned} & \left\langle \mathcal{L}(t|5/2, 5Q^2/2) \middle| \mathcal{C} \right\rangle \\ &= \int dq^2 p(q^2) \int d\dot{q} p(\dot{q}) \frac{e^{-t5Q^2/2(t\Sigma^2)} \mathcal{C}(q, \dot{q})}{(t\Sigma^2)^{1/2} (1 - tr_G^2)^{3/2}}, \end{aligned} \quad (\text{A19})$$

where we defined

$$Q^2 \equiv (r_G q + r_X \Gamma \sqrt{s} \dot{q})^2 / (r_G^2 + r_X^2). \quad (\text{A20})$$

Equation (A19) can now be differentiated j times with respect to t to obtain bias parameters c_{2j} of arbitrary order.

APPENDIX B: DETAILS OF CALCULATIONS IN STOCHASTIC MODELS

In this Appendix, we provide some formulae for the first crossing distribution $f(s)$. Following [Musso & Sheth \(2014\)](#), if q^2 is drawn from a χ_5 distribution obtained by summing 5 independent Gaussian variables with variance s , the up-crossing distribution of the process in equation (34) can be written as

$$f(s) = \int dq p_5(q) \int d\dot{q} p(\dot{q}) f_{q\dot{q}}(s, q, \dot{q}; q_c) \quad (\text{B1})$$

where $p_5(q)$ is the χ -distribution associated with q^2 and $p(\dot{q})$ is a Gaussian with zero mean and variance $\langle \dot{q}^2 \rangle = 5/\Gamma^2 s$. The function $f_{q\dot{q}}(s, q, \dot{q}; q_c)$ is given by

$$\begin{aligned} & s f_{q\dot{q}}(s, q, \dot{q}; q_c) \\ &= \frac{e^{-B^2/2s}}{\sqrt{2\pi}} \int \frac{dx e^{-(x-\gamma B/\sqrt{s})^2/2(1-\gamma^2)}}{\sqrt{2\pi(1-\gamma^2)}} \\ & \times \left(x/\gamma - \dot{B} \right) \Theta_{\text{H}} \left(x/\gamma - \dot{B} \right), \end{aligned} \quad (\text{B2})$$

where $B = \delta_c + (\delta_c/q_c) \sqrt{s} q$ and $\dot{B} = dB/d\sqrt{s} = (\delta_c/q_c)(q + \sqrt{s}\dot{q})$. The integral over \dot{q} in equation (B1) can then be performed analytically, leaving two integrals (over x and q) that must be performed numerically. For brevity, we omit the expression involving these two integrals.

The smooth curve in Figure 1 shows that this formula provides an excellent description of the first crossing distribution in Monte Carlo realizations of the process. The expressions for non local bias we derived in the previous section follow through for the more general excursion set constraint discussed here, simply by replacing the constraint $\mathcal{C}(q, \dot{q})$ with the excursion set one, equation (34), since the latter does not depend on ξ_y . This is equivalent to replacing $\mathcal{C}(q, \dot{q}) \rightarrow f_{q\dot{q}}(s, q, \dot{q}; q_c)/f(s)$ in equation (A10), which gives

$$\begin{aligned} & \left\langle L_1^{(3/2)}(5Q^2/2) \middle| \text{Exc. Set} \right\rangle \\ &= r_G^2 \int dq p_5(q) \int d\dot{q} p(\dot{q}) \frac{f_{q\dot{q}}(s, q, \dot{q}; q_c)}{f(s)} \tilde{c}_2(q, \dot{q}; s, S), \end{aligned} \quad (\text{B3})$$

where $\tilde{c}_2(q, \dot{q}; s, S)$ was defined in equation (A11).