THE BALDWIN EFFECT IN THE NARROW EMISSION LINES OF ACTIVE GALACTIC NUCLEI

KAI ZHANG1,2, TING-GUI WANG1, C. MARTIN GASKELL3, and XIAO-BO DONG1

1 Key Laboratory for Research in Galaxies and Cosmology, The University of Sciences and Technology of China, Chinese Academy of Sciences, Hefei, Anhui 230026, China; zkdc@mail.ustc.edu.cn, twang@ustc.edu.cn, xbdong@ustc.edu.cn
2 Key Laboratory for Research in Galaxies and Cosmology, Shanghai Astronomical Observatory, Chinese Academy of Sciences, 80 Nandan Road, Shanghai 200030, China
3 Centro de Astrofísica de Valparaíso y Departamento de Física y Astronomía, Facultad de Ciencias, Universidad de Valparaíso, Av. Gran Bretaña 1111, Valparaíso, Chile; martin.gaskell@uv.cl

Received 2012 August 30; accepted 2012 November 5; published 2012 December 14

ABSTRACT

The anti-correlations between the equivalent widths of emission lines and the continuum luminosity in active galactic nuclei (AGNs), known as the Baldwin effect, are well established for broad lines, but are less well studied for narrow lines. In this paper we explore the Baldwin effect of narrow emission lines over a wide range of ionization levels and critical densities using a large sample of broad-line, radio-quiet AGNs taken from Sloan Digital Sky Survey Data Release 4. These type 1 AGNs span three orders of magnitude in continuum luminosity. We show that most narrow lines show a similar Baldwin effect slope of about $-0.2$, while the significant deviations of the slopes for [N ii] $\lambda 6583$, [O ii] $\lambda 3727$, [Ne v] $\lambda 3425$, and the narrow component of Hα can be explained by the influence of metallicity, star formation contamination, and possibly by the difference in the shape of the UV–optical continuum. The slopes do not show any correlation with either the ionization potential or the critical density. We show that a combination of 50% variations in continuum near 5100 Å and a lognormal distribution of observed luminosity can naturally reproduce a constant Baldwin effect slope of $-0.2$ for all narrow lines. The variations of the continuum could be due to variability, intrinsic anisotropic emission, or an inclination effect.

Key words: galaxies: active – galaxies: Seyfert – quasars: emission lines

Online-only material: color figures

1. INTRODUCTION

The anti-correlation between the equivalent widths (EWs) of broad emission lines and the luminosity of active galactic nuclei (AGNs; the “Baldwin effect,” hereafter BE) was first discovered by Baldwin (1977) for the C iv $\lambda 1549$ broad emission line in high-redshift AGNs. It was initially hoped that this effect could be used to calibrate AGN luminosity so that AGNs could be used as cosmological standard candles (Baldwin et al. 1978), but the large dispersion of this relationship rendered this impossible (Baldwin et al. 1989; Zamorani et al. 1992). The BE is now well established for nearly all broad emission lines and the slope of the BE steepens with increasing ionization potential (Zheng & Malkan 1993; Dietrich et al. 2002). Several mechanisms have been proposed to explain this effect (see Shields 2007 for a review). Among these, perhaps the most widely accepted one is that the ionizing continuum softens with increasing luminosity, so there are relatively fewer ionizing photons for broad emission line formation in high-luminosity AGNs. This model can reproduce the ionization energy–BE slope relationship fairly well (Korista et al. 1998) and the assumption has observational support (Binette et al. 1989; Zheng & Malkan 1993; Wang et al. 1998; Korista et al. 1998). Some theoretical models (Netzer 1985, 1987; Netzer et al. 1992; Wandel 1999a, 1999b) could produce a softer ionizing spectrum in high-luminosity sources, but the standard thin disk model (Shakura & Sunyaev 1973) they adopt suffers from many problems (Antonucci 2002, 2013; Gaskell & Klimik 2003; Gaskell 2008; Lawrence 2012). In this sense, the explanation of the broad line BE is still elusive and controversial. An outlier of the ionization energy–BE slope relationship is N v $\lambda 1240$, which has an ionization energy of 97.7 eV but shows no BE at all (Dietrich et al. 2002). It has been proposed that this can be explained by a dependence of metallicity on AGN luminosity, which in turn is a combination of Eddington ratio ($L/L_{edd}$) and black hole mass (Korista et al. 1998; Hamann & Ferland 1993, 1999; Dietrich et al. 1999; Dietrich & Wilhelm-Erken 2000). In local galaxies, gas metallicity correlates well with the mass of galaxies (Tremonti et al. 2004), and more massive galaxies have larger black hole masses ($M_{BH}$) according to the $M_{BH}$–$M_*$ relationship. For the type 1 AGN population, the brighter AGNs would have a higher $M_{BH}$ on average (Kollmeier et al. 2006; Steinhardt & Elvis 2010a, 2010b; Lusso et al. 2012) and thus higher metallicities. Recent studies of the BE for C iv, Mg ii, and Fe ii, however, challenge this picture by showing that the BE might instead be driven by the EW’s correlation with the Eddington ratio (Baskin & Laor 2004; Bachev et al. 2004; Warner et al. 2004; Zhou et al. 2006; Dong et al. 2009a, 2009b) or with $M_{BH}$ (Netzer et al. 1992; Wandel et al. 1999; Shields 2007; Kovacević et al. 2010).

The BE of narrow emission lines is much less well studied, and some results are still controversial. Steiner (1981) discovered a strong BE for [O iii] $\lambda 5007$ (specifically in AGNs with strong optical Fe ii) and Wills et al. (1993) found that the narrow lines in high-luminosity AGNs are very weak. Meanwhile, Croom et al. (2002) found a significant BE for [Ne v] $\lambda 3425$ and [O ii] $\lambda 3727$ but obtained a null result for [Ne iii] $\lambda 3870$ and [O iii] $\lambda 5007$ using the Two-Degree Field sample. Subsequent work shows, however, that [O iii] $\lambda 5007$ does show a BE (Dietrich et al. 2002; Netzer et al. 2004). Hönig et al. (2008) and Keremdziev et al. (2009) used Spitzer data to find that the BEs of different mid-IR narrow emission lines have a nearly constant slope. Several mechanisms have been proposed to explain the narrow-line region (NLR) BE. NLRs follow a size–luminosity relationship (Schmitt et al. 2003b; Bennert et al. 2002; Greene et al. 2011). The size of the NLR may grow.
beyond the size of the host galaxy in high-luminosity AGNs, thus turning from ionization-bounded to matter-bounded, and so producing the BE (Croom et al. 2002). As the EW is proportional to the covering factor (CF) and it is found that the CF contributes to much of the variance of the EW (Baskin & Laor 2005), a luminosity-dependent CF is also a possible cause of the BE (Shields et al. 1995; Stern & Laor 2012b, 2012c). It has recently been proposed that the EW of narrow emission lines is dependent on the inclination to the accretion disk (Risaliti et al. 2011), at least in the highest EW sources. In principle, this is a potential cause of the BE, too.

To make progress in understanding the NLR BE, we need to measure narrow emission lines of a wide range of ionization potentials and critical densities, and determine the nature of their BEs more accurately. In this paper, we use a well-defined sample drawn from Sloan Digital Sky Survey (SDSS) Data Release 4 (DR4) and employ the technique of composite spectra to investigate the BE for prominent narrow emission lines in the optical band and to study the origin of the NLR BE. We use a cosmology with \( H_0 = 70 \, \text{km s}^{-1} \, \text{Mpc}^{-1} \), \( \Omega_m = 0.3 \), and \( \Omega_\Lambda = 0.7 \) throughout this paper.

2. SAMPLE AND MEASUREMENTS

2.1. Sample

We need accurate emission line and continuum measurements to ensure reliable determinations of EWs and continuum luminosities. From the spectral data set of the SDSS DR4 (Adelman-McCarthy et al. 2006), we have selected 4178 Seyfert 1 galaxies and quasars (i.e., type 1 AGNs) as described in Dong et al. (2011). We apply a redshift cutoff of 0.8 so that the redshift of the spectrum can be accurately determined using [O III] \( \lambda \lambda 5007 \). To ensure high-quality spectra, we require a median signal-to-noise ratio (S/N) of \( \geq 10 \) per pixel in the optical. To minimize the host-galaxy contamination (see the Appendix of Dong et al. 2011), we restrict the weak stellar absorption features, such that the rest-frame EWs of CaK (3934 Å), CaH + He (3970 Å), and H\( \delta \) (4102 Å) absorption features are undetected at \( < 2 \sigma \) significance. This criterion ensures that the host contamination is less than 10% around 4200 Å (Dong et al. 2011). One may note that Dong et al. (2011) do not consider a very young stellar population (i.e., emission line galaxies). But previous analysis has shown that in massive galaxies, the optical continuum is not dominated by a very young stellar population although the UV continuum may be dominated (Schawinski et al. 2007). After removing duplications and sources with too many bad pixels in the H\( \beta \) + [O III] region, we obtain 4178 type 1 AGNs. Including radio-loud AGNs may influence the measurement of EW of emission, thus producing a false effect for two reasons: first, the jet may interact with the ISM to enhance the narrow line emission (Labiano et al. 2008), and second, it might also enhance the continuum through beaming if the jet points close to our line of sight. By matching with the FIRST catalog (Becker et al. 1995) using the method of Lu et al. (2010), we reject 499 radio-loud AGNs so that our final sample consists of 3677 sources.

2.2. Spectral Fitting and Measurements

We give a brief description of our spectrum fitting process here; the details can be found in Dong et al. (2011). To model the spectrum, we used a code based on the MPFIT package (Markwardt 2009) and fit the AGN featureless continuum, the Fe II multiplets, and other emission lines simultaneously. The AGN continuum is represented locally by a power law, for the region of 4200–5600 Å and for the H\( \alpha \) region (if present). The Fe II template by Véron-Cetty et al. (2004) that we use is constructed using the identification and measurement of Fe II lines in I Zw 1. It has two separate sets of templates in analytical forms, one for the broad-line system and the other for the narrow-line system. Within each system, the relative velocity shifts and relative strength are assumed to be the same as those in I Zw 1. Broad Fe II lines share the same profile as broad H\( \beta \), while each narrow Fe II line is modeled with a Gaussian. During the fitting, the normalization and redshift of each system are taken as free parameters. The broad Balmer lines are fitted with as many Gaussians as are statistically justified. All narrow emission lines, except for the [O III] \( \lambda \lambda 4959, 5007 \) doublet lines, are fitted with a single Gaussian. Each line of the [O III] doublet is modeled with two Gaussians, one accounting for the line core and the other for a possible blue wing, as seen in many objects. Since the sources in our sample do not suffer from significant host galaxy contamination, we do not apply starlight corrections to individual spectra. For each source, we use \( \lambda L_\lambda (5100) \) and the FWHM of the broad H\( \beta \) line to obtain the \( M_{BH} \) using the virial mass estimates from the Dibai method (Dibai 1977) using the formalism of Wang et al. (2009). The typical statistical scatter about the \( M_{BH} \) obtained by reverberation-mapping is about 0.4 dex, and it may also be subjected to more systematic errors (Krolik 2001; Collin et al. 2006; Shen et al. 2008; Fine et al. 2008; Marconi et al. 2008; Denney et al. 2009; Rafiee & Hall 2011a; Steinhardt 2011). The error of \( L/L_{Edd} \) is of similar magnitude to that of \( M_{BH} \).

2.3. Composite Spectra Generating and Fitting

A convenient way to explore the correlations between EWs and other parameters like luminosity is to make composite spectra for different parameter bins. We normalize individual spectra to the mean flux around 4200 Å and then construct the geometric composite as in Vanden Berk et al. (2001). To get an accurate continuum measurement, we need to subtract the broad emission lines, specifically H\( \alpha \) and H\( \beta \), and Fe II emission from the spectrum. For the broad component of H\( \beta \) and Fe II, we subtract them from the original spectrum, but for the broad component of H\( \alpha \), we leave it unsubtracted when making composite spectra. This is because the broad H\( \alpha \) is highly blended with [N II] \( \lambda \lambda 6583 \) and [N II] \( \lambda 6548 \), so deblending in individual spectrum is not reliable while deconvolving the blends in the high S/N composite spectrum is easier. The fitting algorithm used to model the composite spectrum is the same as described in Section 2.2. More specifically, we fit the blends using one Gaussian for the narrow component of H\( \alpha \), [N II] \( \lambda \lambda 6583 \), and [N II] \( \lambda 6548 \), two Gaussians for [S II] \( \lambda \lambda 6717, 6731 \), and three for H\( \beta \). The broad-line-subtracted spectra are shown in Figure 1. The embedded panels show the fitting result for the H\( \alpha \) blend.

3. RESULTS

3.1. The Baldwin Effect for Different Lines

First, we want to explore whether the BE exists in the prominent narrow lines [Ne v] \( \lambda \lambda 3425, [O III] \lambda 3727, [Ne III] \lambda 3870, H\( \beta \), [O III] \( \lambda 5007 \), [O II] \( \lambda 6300 \), H\( \alpha \), [N II] \( \lambda 6583 \), and [S III] \( \lambda \lambda 6717, 6731 \). Measuring narrow lines in individual spectra is subject to the S/N limit and the EW–\( L_{EB} \) relation of all our lines shows a large dispersion (typically 0.2 dex; this can be seen clearly in the middle panel of Figure 4), so we turn
to composite spectra for reliable measurements. We divide our sample into intervals of 0.3 dex in 5100 Å luminosity starting from the top with \( \log L_{\lambda}(5100) = 43.35 \) [erg s\(^{-1}\)]. The spectra are normalized to the [4200Å, 4300Å] window and shifted vertically to show the weakening of lines with luminosity more clearly. The lines we are concerned with are marked with dashed lines labeled at the top. The embedded plots show the fitting results for the narrow component of \( \lambda L\lambda \) slope.

1. With increasing luminosity, the narrow lines vanish.

2. The slope of the observed continuum becomes bluer with increasing luminosity.

To see the dependencies of the EWs on the luminosity more clearly, in Figure 2 we plot the log EW derived from composite spectra for each luminosity bin for each of the lines listed above against \( \log L_{\lambda}(5100) \) in order to get more qualitative results. For each line, we show a weighted linear regression of the logarithm of the EW on the logarithm of the luminosity and make a composite spectra in each luminosity bin as described in Section 2.3. From Figure 1 we can see two clear and remarkable results.

- With increasing luminosity, the narrow lines vanish.
- The slope of the observed continuum becomes bluer with increasing luminosity.

To see the dependencies of the EWs on the luminosity more clearly, in Figure 2 we plot the log EW derived from composite spectra for each luminosity bin for each of the lines listed above against \( \log L_{\lambda}(5100) \) in order to get more qualitative results. For each line, we show a weighted linear regression of the logarithm of the EW on the logarithm of the luminosity and make a composite spectra in each luminosity bin as described in Section 2.3. From Figure 1 we can see two clear and remarkable results.

1. With increasing luminosity, the narrow lines vanish.
2. The slope of the observed continuum becomes bluer with increasing luminosity.

To see the dependencies of the EWs on the luminosity more clearly, in Figure 2 we plot the log EW derived from composite spectra for each luminosity bin for each of the lines listed above against \( \log L_{\lambda}(5100) \) in order to get more qualitative results. For each line, we show a weighted linear regression of the logarithm of the EW on the logarithm of the luminosity and make a composite spectra in each luminosity bin as described in Section 2.3. From Figure 1 we can see two clear and remarkable results.

1. With increasing luminosity, the narrow lines vanish.
2. The slope of the observed continuum becomes bluer with increasing luminosity.

To see the dependencies of the EWs on the luminosity more clearly, in Figure 2 we plot the log EW derived from composite spectra for each luminosity bin for each of the lines listed above against \( \log L_{\lambda}(5100) \) in order to get more qualitative results. For each line, we show a weighted linear regression of the logarithm of the EW on the logarithm of the luminosity and make a composite spectra in each luminosity bin as described in Section 2.3. From Figure 1 we can see two clear and remarkable results.

1. With increasing luminosity, the narrow lines vanish.
2. The slope of the observed continuum becomes bluer with increasing luminosity.

To see the dependencies of the EWs on the luminosity more clearly, in Figure 2 we plot the log EW derived from composite spectra for each luminosity bin for each of the lines listed above against \( \log L_{\lambda}(5100) \) in order to get more qualitative results. For each line, we show a weighted linear regression of the logarithm of the EW on the logarithm of the luminosity and make a composite spectra in each luminosity bin as described in Section 2.3. From Figure 1 we can see two clear and remarkable results.

1. With increasing luminosity, the narrow lines vanish.
2. The slope of the observed continuum becomes bluer with increasing luminosity.

To see the dependencies of the EWs on the luminosity more clearly, in Figure 2 we plot the log EW derived from composite spectra for each luminosity bin for each of the lines listed above against \( \log L_{\lambda}(5100) \) in order to get more qualitative results. For each line, we show a weighted linear regression of the logarithm of the EW on the logarithm of the luminosity and make a composite spectra in each luminosity bin as described in Section 2.3. From Figure 1 we can see two clear and remarkable results.

1. With increasing luminosity, the narrow lines vanish.
2. The slope of the observed continuum becomes bluer with increasing luminosity.

To see the dependencies of the EWs on the luminosity more clearly, in Figure 2 we plot the log EW derived from composite spectra for each luminosity bin for each of the lines listed above against \( \log L_{\lambda}(5100) \) in order to get more qualitative results. For each line, we show a weighted linear regression of the logarithm of the EW on the logarithm of the luminosity and make a composite spectra in each luminosity bin as described in Section 2.3. From Figure 1 we can see two clear and remarkable results.

1. With increasing luminosity, the narrow lines vanish.
2. The slope of the observed continuum becomes bluer with increasing luminosity.

To see the dependencies of the EWs on the luminosity more clearly, in Figure 2 we plot the log EW derived from composite spectra for each luminosity bin for each of the lines listed above against \( \log L_{\lambda}(5100) \) in order to get more qualitative results. For each line, we show a weighted linear regression of the logarithm of the EW on the logarithm of the luminosity and make a composite spectra in each luminosity bin as described in Section 2.3. From Figure 1 we can see two clear and remarkable results.

1. With increasing luminosity, the narrow lines vanish.
2. The slope of the observed continuum becomes bluer with increasing luminosity.

To see the dependencies of the EWs on the luminosity more clearly, in Figure 2 we plot the log EW derived from composite spectra for each luminosity bin for each of the lines listed above against \( \log L_{\lambda}(5100) \) in order to get more qualitative results. For each line, we show a weighted linear regression of the logarithm of the EW on the logarithm of the luminosity and make a composite spectra in each luminosity bin as described in Section 2.3. From Figure 1 we can see two clear and remarkable results.

1. With increasing luminosity, the narrow lines vanish.
2. The slope of the observed continuum becomes bluer with increasing luminosity.

To see the dependencies of the EWs on the luminosity more clearly, in Figure 2 we plot the log EW derived from composite spectra for each luminosity bin for each of the lines listed above against \( \log L_{\lambda}(5100) \) in order to get more qualitative results. For each line, we show a weighted linear regression of the logarithm of the EW on the logarithm of the luminosity and make a composite spectra in each luminosity bin as described in Section 2.3. From Figure 1 we can see two clear and remarkable results.

1. With increasing luminosity, the narrow lines vanish.
2. The slope of the observed continuum becomes bluer with increasing luminosity.
Figure 2. Line equivalent widths: $W_\lambda$, against 5100 Å continuum luminosity: $\lambda L_\lambda$ (5100). We show weighted linear regressions (green lines) for each line and give the BE slopes as well as their errors in the upper right corners in each panel. In the narrow H$\beta$ panel, we also include the narrow component of H$\alpha$ (dashed line) for comparison. (A color version of this figure is available in the online journal.)

increases with luminosity. These indicate that the behaviors of broad lines and narrow lines are different.

The correlation coefficients between EW([O iii]) and $\lambda L_\lambda$ (5100), $z$, $M_{BH}$, and $L/L_{Edd}$ are $-0.12$, $-0.05$, $-0.01$, and $-0.21$, respectively. These are similar to the results in Zhang et al. (2011). For our flux-limited sample, the redshift, luminosity, black hole mass, and Eddington ratio are correlated with each other. The correlation coefficients between $\lambda L_\lambda$ (5100) and $z$, $M_{BH}$, and $L/L_{Edd}$ are 0.81, 0.67, and 0.15, respectively. After controlling for $z$ or $L/L_{Edd}$ in a partial correlation analysis, EW([O iii]) correlates with $\lambda L_\lambda$ (5100) with $r_s = -0.14$ and $-0.16$. This indicates that the BE of narrow lines is not a secondary effect of correlations between the EW and redshift or $L/L_{Edd}$ but an independent phenomenon. We note that as the correlation strength is not strong, one possibility is that there are other factors that regulate the EW as discussed in Zhang et al. (2011). Also, the measurement error in EW([O iii]) may act to smear the correlation. The measurement error could be introduced by limited S/N and the fitting process. In spite of these uncertainties, the BE of narrow lines does exist and may shed light on the physical process in AGNs as explored in detail below.

### 3.2. The Ionization Energy–BE Slope Relationship: No Correlation

A dependence of the BE slope on ionization energy is found for broad emission lines in AGNs (see the Introduction) and this is the most compelling evidence for a luminosity-dependent ionizing spectrum. The NLR lies far from the nucleus, has a complex geometry, and may contain dust (Netzer & Laor 1993; Tomono et al. 2001; Radomski et al. 2003; Schweitzer et al. 2008). This makes the response of the narrow-line flux to changes in the spectral energy distribution (SED) complicated. We plot the slope of BE against the ionization energies of the narrow lines in the left panel of Figure 3. We also plot the Keremedjiev et al. (2009) data for [S iv] $10.51 \mu m$: $-0.29 \pm 0.05$, [Ne ii] 12.81 $\mu m$: $-0.25 \pm 0.06$, [Ne iii] 15.56 $\mu m$: $-0.22 \pm 0.06$, and [Ne v] 14.32 $\mu m$: $-0.19 \pm 0.06$ with purple crosses and the broad-line BE from Dietrich et al. (2002) with blue rectangles for comparison. We can see that the narrow-line BE slopes do not correlate with the ionization energy ($P_{null} = 0.78$, meaning the two-sided probability that a correlation is not present is 78%) but cluster around $-0.2$ with a dispersion of $\pm 0.1$. [N ii] $\lambda 6583$, H$\alpha^n$, and [O ii] $\lambda 3727$ have
slopes of \(-0.10 \pm 0.014\), \(-0.29 \pm 0.033\), and \(-0.37 \pm 0.011\), so they deviate from \(-0.2\) significantly. Possible reasons for this are discussed in Section 4.1.

3.3. Critical Density–Slope Relationship: No Correlation

It is well known that the NLR is stratified, that the high ionization lines rise from the inner part of the NLR while low ionization lines rise further away (see, for example, Veilleux et al. 1991c; Robinson et al. 1994; Bennert et al. 2006a, 2006b; Kraemer et al. 2009). The central electron temperature, density, and ionization parameter are, in general, higher in Seyfert 1s than in Seyfert 2s (Gaskell 1984; Schmitt 1998; Bennert et al. 2006b). The lines of lower critical density are not as strong relative to carbon, oxygen, etc., because nitrogen is a secondary element whose abundance scales as \(Z^2\). This trend could compensate for the decrease of EW in higher luminosity sources, thus making the BE slope flatter. This explanation is supported by a number of recent studies. The metallicities of the broad-line region (BLR; Hamann & Ferland 1993; Nagao et al. 2006a; Juarez et al. 2009) and NLR (e.g., Nagao et al. 2006b; Matsuoka et al. 2009) are both found to correlate with the luminosity of the AGN. So, in principle, the \(Z_{\text{NLR}} - L\) relationship could produce the flattening of the BE of \([\text{N}\, \text{II}]\, \lambda 6583\) we observe here. However, this explanation has a major problem because \([\text{N}\, \text{IV}]\) and \([\text{N}\, \text{III}]\), which would be expected to deviate from the ionization energy–slope relationship for the same reason, actually lie on it (Dietrich et al. 2002). Despite this serious problem that needs to be resolved, Occam’s razor suggests that an enhanced abundance in high-luminosity AGNs is the simplest and most plausible explanation of the flatter BE slope of \([\text{N}\, \text{II}]\, \lambda 6583\).

The second significant deviation from a BE slope of \(-0.2\) is the steep slopes of \([\text{H}\, \alpha]\) and \([\text{O}\, \text{II}]\, \lambda 3727\). Both of these are star formation (SF) indicators in star-forming galaxies (Kennicutt 1998; Ho 2005) and their EWs may reach several hundred Å in starburst galaxies. Thus, even though the AGNs we selected show no absorption feature in the continuum, the emission lines are still possibly subject to SF contamination because of their large EWs. In low-luminosity sources whose continuum is low, the emission line from SF may contribute significantly to the total line flux, and thus enhance the EW. In type 2 AGNs, the SF contribution of \([\text{H}\, \alpha]\) is estimated to be more than 60% (Brinchmann et al. 2004). The \([\text{H}\, \alpha]\) emitting region of type 1 AGNs, however, may have a higher fraction of the emission originating with the AGN (Zhang et al. 2008), but still be heavily influenced by SF. Furthermore, a variety of studies have established a correlation between the strength of AGN activity and SF in the local universe (e.g., Rowan-Robinson 1995; Croom et al. 2002; Netzer et al. 2007; Shao et al. 2010). The steepest correlation is \(L_{\text{SF}} \propto L_{\text{AGN}}^{0.8}\) (Netzer 2007), meaning \(L_{\text{SF}} / L_{\text{AGN}} \propto L_{\text{AGN}}^{-0.2}\), a BE slope steeper than \(-0.2\). So it makes sense that the SF in host galaxies would steepen the BE slopes of \([\text{H}\, \alpha]\) and \([\text{O}\, \text{II}]\, \lambda 3727\). However, the \([\text{O}\, \text{II}]\, \lambda 3727\) and \([\text{H}\, \alpha]\) BEs cannot be attributed entirely to SF–AGN relationship. For a plausible range of ionization parameters, densities, and ionizing spectra, the intensity of \([\text{O}\, \text{II}]\, \lambda 3727\) is proportional to that of \([\text{O}\, \text{II}]\, \lambda 5007\) (10%–30%; Ferland & Osterbrock 1986; Ho et al. 1993a, 1993b). Since \([\text{O}\, \text{II}]\, \lambda 5007\) shows a significant BE, \([\text{O}\, \text{II}]\, \lambda 3727\) is
unlikely to show a radically different trend. Thus it is safe to conclude that the [O\textsc{ii}] $\lambda$3727 and He$^+$ BE could be partly (but not totally) produced by SF contamination.

A third effect we can see in Figure 1 is that the optical continuum becomes bluer toward higher luminosity. This would leverage the continuum and lower the EW if the line flux remains unchanged. This could arise if the NLR has a filtered SED (Kraemer et al. 1998; Groves et al. 2004a, 2004b) so that a change in the ionizing spectrum of an AGN would not change the NLRs SED much. The [Ne\textsc{v}] $\lambda$3425, [O\textsc{ii}] $\lambda$3727, and [Ne\textsc{iii}] $\lambda$3870 lines are most likely influenced by a continuum shape difference effect because of their shorter wavelength. A possible explanation of the bluer color of the UV–optical SED is the anisotropy of the continuum emission. Because the accretion disk is optically thick, we will see a dimmer continuum when viewing it edge-on. (This is the combined result of the $\cos i$ projection effect and the disk equivalent of “limb darkening.”) We could also preferentially be seeing the inner, high-temperature part of the disk when viewed face-on. These effects will combine to give a higher observed luminosity with a face-on viewing angle. The reddening of spectra also depends on the viewing angle (Keel 1980; de Zotti & Gaskell 1985; Zhang et al. 2008). Gaskell et al. (2004) used radio orientations to obtain AGN reddening curves and found that the continuum shape is profoundly affected by reddening for all but the bluest AGNs. Because of these effects, sources with small inclination would have both a large luminosity and low reddening. An alternative explanation of the continuum shape difference is host galaxy contamination. Shen et al. (2011) made composite spectra of different $\lambda L_\odot$ (5100) luminosity bins and found that the UV parts of all the composite spectra are similar while the optical parts flatten with decreasing luminosity. They interpreted this trend as due to host galaxy contamination in the optical region of the spectrum in low-luminosity AGNs. A similar argument is given by Stern & Laor (2012a). Even though we have rejected objects with significant stellar light contributions, weak absorption lines can be spotted on the final composite spectrum of the lowest luminosity bin because of its extremely high S/N. Our data cannot distinguish between different mechanisms that give rise to the continuum shape difference; this is beyond the scope of this paper. After correcting for the deviations listed above, our conclusion that the slopes of BEs for different narrow lines are nearly constant is further strengthened.

4.2. Possible Causes of the NLR Baldwin Effect

4.2.1. Softening of the Ionizing Continuum

It has been argued on both observational and theoretical grounds that there is a softening of the ionizing continuum with increasing luminosity (see, for example, Binette et al. 1989; Netzer et al. 1992; Zheng & Malkan 1993; Wang et al. 1998; Korista et al. 1998). If this is indeed the case, an important prediction of this is the ionization-potential–BE-slope relationship (Korista et al. 1998). It is successful in explaining the broad line BE, but a major failing is that it cannot explain the constant BE slope of narrow lines. It is already known that the NLRs of different AGNs are similar in the sense that the line ratios show less than a 0.5 dex difference from object to object (Koski 1978; Veilleux & Osterbrock 1987; Veilleux 1991a, 1991b, 1991c; Véron-Cetty & Véron 2000; Dopita et al. 2002; Gorjian et al. 2007). Kraemer et al. (2000), Dopita et al. (2002), and Groves et al. (2004a, 2004b) proposed an NLR model where dust regulates the incident ionizing spectrum so as to keep the ionization parameter in the NLR constant. In this model, variation of the SED is filtered by dust in the NLR and thus the continuum shape seen by the NLR is dominated by the effect of dust rather than by intrinsic changes in the SED. This leads to more stable conditions in the NLR. This could be the cause of the lack of a dependence of the slope of the BE on the ionization energy.

4.2.2. Luminosity-dependent Covering Factor

A luminosity-dependent CF is a natural explanation of both a BLR and an NLR BE. However, to explain the ionization dependence, the variation in CF must have a different dependence for lines of different ionizations. This would not be a surprise for the BLR since there is strong radial ionization stratification and the highest ionization lines are an order of magnitude closer to the black hole than the lowest ionization lines (see Gaskell 2009 for a review). Shields et al. (1995) suggested that the broad-line BE could be caused by a luminosity-dependent CF for clouds that are optically thin to photons with wavelengths less than 912 Å. The optically thin clouds, which have small column densities, preferentially emit high ionization lines. However, a comparison of line profiles shows that the BLR BE is due to changes in the low-velocity BLR gas (Francis et al. 1992) rather than changes in the very broad component. Furthermore, Snedden & Gaskell (2007) argued that optically thin gas does not make a substantial contribution to the BLR. Nevertheless, given the strong radial ionization stratification of the BLR, and the decreasing CF with ionization (e.g., Francis et al. 1992), an additional luminosity dependence of the CF could explain the BE and why the high ionization lines show a steeper slope than low ionization lines. While this could be consistent with the BLR, it is not obvious how such an explanation can be reconciled with the lack of an ionization dependence we find here for the BE.

4.2.3. A Disappearing NLR?

The NLR size is correlated with the luminosity of an AGN as $R_{\text{NLR}} \propto L^{0.5}$ (Bennert et al. 2002, 2006a, 2006b; Schmitt et al. 2003b; Greene et al. 2011). So in high-luminosity AGNs, the NLR may possibly turn from ionization-bounded to matter-bounded, and the luminosity of narrow lines would cease increasing with AGN luminosity. This model could, in principle, explain part of the BE of the NLR (Croom et al. 2002), but it would be too rash a conclusion that it is the origin of the BE because the bulk of the NLR emission originates from within the central few tens or hundreds of parsecs (Schmitt et al. 2003a), so it is unlikely to exceed the scale of the galactic bulge. Besides, the NLR has no natural size because it has no definite edge, so some cut in surface brightness or line ratio is needed to define the size (Schmitt et al. 2003b; Bennert et al. 2006a, 2006b; Greene et al. 2011). This makes the interpretation of size complicated. Netzer et al. (2004) argue that $R_{\text{NLR}} \propto L^{0.5}$ is theoretically sound yet this relationship must break down for $R_{\text{NLR}}$ exceeding a few kiloparsecs. They found that high-$z$ AGNs have NLR sizes no larger than 10 kpc. It is also well known that the NLR is stratified so that the high ionization lines rise from the inner part of the NLR while low ionization lines come from farther away (Veilleux et al. 1991c; Bennert et al. 2006a, 2006b). Thus, the “disappearing NLR effect,” if it exists, could preferentially influence low ionization and low critical density lines. This again stands against the constant slope we find.
4.2.4. Continuum Variation

Because different lines all seem to share a similar BE slope, it is natural to think that it might be the continuum, rather than the emission lines, that causes the NLR Baldwin Effect. Jiang et al. (2006) proposed that continuum variation can produce a weak BE ($\beta = -0.05 \pm 0.05$) if the light crossing time for the region emitting narrow FeK$\alpha$ exceeds the variability timescale for the X-ray continuum and the amplitude of variability anticorrelates with the luminosity. Shu et al. (2012) found a strong anti-correlation between the EW of the narrow FeK$\alpha$ line and $L_X$ (EW/[$\text{EW}$] $\propto (L/\langle L \rangle)^{-0.82 \pm 0.10}$, where $\langle \rangle$ means time-averaged values) consistent with the X-ray BE expected in an individual AGN if the narrow-line flux remains constant while the continuum varies. For an NLR whose light crossing time is about $10^3$ yr, the narrow-line luminosity can safely be assumed to be constant. The amplitude of variability of the ionizing continuum can be obtained from the monitoring of AGNs after correction for the constant host galaxy light contribution. As is well known, the amplitude of variability of an AGN increases with time. This can be most readily seen from the structure functions of AGNs (the variance as a function of time interval between observations). We are interested in variability on the longest timescales (timescales similar to the NLR light crossing time). We do not, of course, have monitoring on such long timescales, but we can estimate the amplitude from structure functions. Cid Fernandes et al. (2000) have presented structure functions for bright AGNs. These are all consistent with the amplitude of variability rising as the time interval increases to a characteristic time of a year or a few years and then remaining constant. Collier & Peterson (2001) obtained a similar result for lower-luminosity AGNs (but with a shorter characteristic time) and also found that the forms of the UV structure functions are similar to the optical ones. There are indications that the structure function increases gradually on much longer timescales, but data on this part of the structure function are limited. The observed variability thus provides a lower limit for the variability on the light crossing timescale of the NLR.

In order to minimize the host galaxy light contribution, one needs to go to as short a wavelength as possible. This means using $U$-band observations or space UV observations. For example, for NGC 4151, Merkulova (2006) found an observed peak-to-peak $U$-band amplitude of a factor of 7.5. UV variability can be a lot larger. For example, the peak-to-peak amplitude at 1300 Å for Fairall 9 is a factor of 25 (Koratkar & Gaskell 1989). Inspection of the results of long-term UV monitoring of a number of AGNs with the IUE satellite (Koratkar & Gaskell 1989, 1991a, 1991b; Clavel et al. 1991; O’Brien et al. 1998) gives a median UV peak-to-peak variability of a factor of 7, but the upper quartile is a factor of 13. After correction for host-galaxy light, the amplitudes of optical variability are similarly large. For example, the NGC 4151 photometry of Lyuty & Doroshenko (1999) gives a peak-to-peak amplitude of a factor of about 25–30.

Continuum anisotropy (Wang et al. 1998) can produce an effect similar to that caused by the actual variability of the continuum. Risaliti et al. (2011) recently found that the optically thick disk emission responsible for the continuum and isotropic [O III] emission will produce the EW([O III]) distribution very well. Variability and continuum anisotropy are, in practice, indistinguishable, so we can consider them together.

We made Monte Carlo simulations to explore whether continuum variability can explain the BE slope. In a flux-limited survey, the luminosity will show an approximately lognormal distribution. We generated an artificial sample of sources with a similar distribution of $\lambda L_{\lambda}(5100)$ as the observed AGN sample (0.35 dex here). We set the EW of the artificial sources to be the mean EW of the whole sample, and we add a Gaussian of 0.22 dex to the EW to account for the intrinsic dispersion. UV and optical AGN variability is approximately lognormal—i.e., it looks normal when plotted in magnitudes (see Gaskell 2004 for a discussion of lognormal variation of AGNs). The continuum variation was simulated by adding a Gaussian with $\sigma = 50\%$ (0.18 dex) to the continuum while the emission line flux was kept unchanged. This is an oversimplification but it can help us to gain some insight into the effect of continuum variability. We generated 4000 sources in each round and measured the slope of the BE using the 4000 sources. An example of the simulation is shown in Figure 4. With a peak-to-peak continuum variation of a factor of three, the simulation can produce a BE slope of $-0.2 \pm 0.01$ while a factor of six variation produces a BE slope of $-0.3 \pm 0.01$. The simulated distribution of EW–$\lambda L_{\lambda}(5100)$ can be seen to be very similar to what is observed. Obviously, this model will produce a similar slope for every narrow line if the continuum changes with similar amplitude in the wavelength range we consider. It should be noted that due to variation in the continuum slope, not all lines show the same BE slope, but depend on wavelength. We assume a constant narrow line flux during the continuum variability. This is an approximation for the continuum variations on timescales of much shorter than the light travel time over the NLR because at such a short timescale, the NLR has little response to the continuum variations. For variations on longer timescales, one needs to properly convolve the continuum variations with the response of emission lines (transfer function). By considering the latter response, the variations of EW will be somewhat smaller, but this is equivalent to requiring a somewhat larger continuum variability amplitude.

Conversely, since continuum variability or the equivalent of continuum variability will inevitably produce a BE, the BE slope we observe could provide an upper limit for the few $10^2$–$10^3$ yr variability of the AGN optical continuum. If the variability amplitudes exceed a factor of three, a steeper BE will emerge. At first, this might seem to be at odds with the amplitudes of the UV variability and structure functions observed, but it must be remembered that the EW is the ratio of the line flux to the observed optical continuum and the observed continuum has a substantial starlight contamination. An example of how the apparent continuum variability is substantially less than the real variability is shown in Figure 4 of Gaskell et al. (2008). Host galaxy light limits the apparent peak-to-peak continuum variability to about a factor of three (i.e., an rms variability of a few tenths of a magnitude). This is just what is observed for the PG AGNs studied by Cid Fernandes et al. Since the structure functions are relatively flat after a year or so (Cid Fernandes et al. 2000), the variability effect on the BE should be apparent in a few years. An obvious test of this is to re-observe AGNs after a few years.

4.3. Drawbacks

In summary, the combination of a 50% variation of $\lambda L_{\lambda}(5100)$ and a lognormal distribution of luminosity will naturally produce a $-0.20$ slope of BE for every narrow line, as we observe. The model we employ is obviously an oversimplification. To make a more realistic model, we need a dedicated treatment of sample selection effects and to make a more realistic assumption of the variability including amplitude and variation form,
and their dependence on wavelength about the light curve. Despite these simplifications, our results do show that continuum variation will inevitably produce a similar BE for each narrow line. Meanwhile, while the model predicts the same BE slope for each narrow line, the differences in slope are still significant due to the small error bars. There must therefore be some additional factor at work. It has been shown in Figure 3 that the differences do not correlate with the ionization potential or critical density, so other factors must have an effect. A deeper exploration of other factors such as ionization slope and NLR geometry and a more realistic model are needed to make progress, but this is beyond the scope of this paper.

5. CONCLUSION

We have constructed a sample of 3677 z < 0.8 radio-quiet AGNs from SDSS DR4 that spans three orders of magnitude in luminosity to explore the relationship between EW of narrow lines and λL_α(5100). We have computed composite spectra for each λL_α(5100) bin to enhance the S/N. Hα and Hβ as well as Fe II were subtracted to obtain an accurate measurement of narrow emission lines and continuum. We find that most narrow lines show a similar BE slope of about −0.2 while the large deviation of [N II] λ6583, [O II] λ3727, Hα, and [Ne V] λ3425 might be explained by a metallicity effect, SF contamination, or the UV–optical continuum difference. The slope does not show any correlation with ionization energy and critical density. We propose that the combination of a 50% variation of the observed luminosity distribution will naturally produce a −0.2 slope of BE for every narrow line.

We thank our referee, Robert Antonucci, for inspiring comments that helped to improve the paper significantly and Brent Groves for helpful discussion of the NLR model. This work is supported by Chinese NSF grants NSF-10703006, NSF-10973013, NSF-11073019, and NSF-11233002; the GEMINI-CONICYT Fund of Chile through project No.32070017; and FONDECYT of Chile through project No. 1120957. Funding for the Sloan Digital Sky Survey (SDSS) has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Aeronautics and Space Administration, the National Science Foundation, the US Department of Energy, the Japanese Monbukagakusho, and the Max Planck Society. The SDSS is managed by the Astrophysical Research Consortium (ARC) for the Participating Institutions. The SDSS Web site is http://www.sdss.org/.

REFERENCES

Adelman-McCarthy, J. K., Jennifer, K., Ageros, M. A., et al. 2006, ApJS, 162, 38
Antonucci, R. 2002, in Proc. XII Canary Islands Winter School of Astrophysics, Astrophysical Spectropolarimetry, ed. J. Trujillo-Bueno, F. Moreno-Insertis, & F. Sánchez (Cambridge: Cambridge Univ. Press), 151
Antonucci, R. 2012, arXiv:1210.2716
Bachev, R., Marziani, P., Sulentic, J. W., et al. 2004, ApJ, 617, 171
Baldwin, J. A. 1977, ApJ, 214, 679
Baldwin, J. A., Burke, W. F., Gaskell, C. M., & Wampler, E. J. 1978, Natur, 273, 431
Baldwin, J. A., Wampler, E. J., & Gaskell, C. M. 1989, ApJ, 338, 630
Baskin, A., & Laor, A. 2004, MNras, 350, L31
Baskin, A., & Laor, A. 2005, MNras, 358, 1043
Becker, R. H., White, R. L., & Helfand, D. J. 1995, ApJ, 450, 559
Bennert, N., Falcke, H., Schulz, H., Wilson, A. S., & Wills, B. J. 2002, ApJ, 574, L105
Bennert, N., Jungwirth, B., Komossa, S., Haas, M., & Chini, R. 2006a, A&A, 459, 55
Bennert, N., Jungwirth, B., Komossa, S., Haas, M., & Chini, R. 2006b, A&A, 456, 953
Binette, L., Prieto, A., Soszokiewicz, E., & Zheng, W. 1989, ApJ, 343, 135
Brinchmann, J., Charlot, S., White, S. D. M., et al. 2004, MNRAS, 351, 1151
Cid Fernandes, R., Sodr´e, L., Jr., & Vieira da Silva, L., Jr. 2000, ApJ, 544, 123
Clavel, J., Reichert, G. A., Alloin, D., et al. 1991, ApJ, 366, 64
Collier, S., & Peterson, B. M. 2001, ApJ, 555, 775
Collin, S., Kawaguchi, T., Peterson, B. M., & Vestergaard, M. 2006, A&A, 456, 75
Croom, S. M., Rhook, K., Corbett, E. A., et al. 2002, MNras, 337, 275
de Zotti, G., & Gaskell, C. M. 1985, A&A, 147, 1
Denney, K. D., Peterson, B. M., Dietrich, M., Vestergaard, M., & Bentz, M. C. 2009, ApJ, 692, 246
Dibai, E. A. 1977, SVAL, 3, 1
Dietrich, M., Appenzeller, I., Wagner, S. J., et al. 1999, A&A, 352, L1
Dietrich, M., Hamann, F., Shields, J. C., et al. 2002, ApJ, 581, 912
Dietrich, M., & Wilhelm-Erkens, U. 2000, A&A, 354, 17
Dong, X., Wang, J., Tang, T., et al. 2009a, in ASP Conf. Ser. 408, The Starburst-AGN Connection, ed. W. Zhang, Y. Zhang, Y. Luo, & Z. Chen (San Francisco, CA: ASP), 83
Dong, X.-B., Wang, J.-G., Ho, L. C., et al. 2011, ApJ, 736, 86
Dopita, M. A., Groves, B. A., Sutherland, R. S., Binette, L., & Cecil, G. 2002, ApJ, 572, 753
Ferland, G. J., & Osterbrock, D. E. 1986, ApJ, 300, 658

Figure 4. Variability-driven BE simulation result. (a) The simulated BE, the slope of BE is shown in the upper right corner. (b) The observed EW–log λL_α(5100) distribution of [O III] λ5007 from our SDSS DR4 sample. The slope of BE is shown in the upper right corner. (A color version of this figure is available in the online journal.)
